

Identification of novel *CLN2* mutations shows Canadian specific NCL2 alleles

W Ju, R Zhong, S Moore, D Moroziewicz, J R Currie, P Parfrey, W T Brown, N Zhong

J Med Genet 2002;**39**:822–825

NCL2, the neuronal ceroid lipofuscinosis (NCL) with a classical late infantile onset (LINCL or Jansky-Bielschowsky disease, MIM 204500) is among the most common childhood neurodegenerative disorders. Usually, affected children show generalised tonic-clonic and/or myoclonic seizures starting by 2 to 3 years of age as the initial, and the most noticeable, clinical symptom. This is followed by regressive cognitive dysfunction, including speech delay, slow learning, and mental retardation; neuromotor dysfunction, including ataxia and inability to walk; vision problems; behavioural changes and eventual dementia. Pathological studies by electron microscopy (EM) examination of peripheral blood "buffy coats" and/or tissue biopsy (including skin, conjunctiva, rectal, muscle, brain, or peripheral nerve) have found lysosomal curvilinear (CV) inclusions. These CV inclusions reflect the typical ultrastructural profile of autofluorescent lipofuscin storage materials that consist predominantly of mitochondrial ATP synthase (ATPase) subunit C. Such EM observations are the common basis for the clinical diagnosis of LINCL.¹

Genetically, the gene *CLN2* underlying NCL2 has been mapped to chromosome 11p15.² It consists of 13 exons spanning 6.65 kb of genomic DNA³ and encodes a 46 kDa protein, lysosomal tripeptidyl peptidase 1 (TPP1), which cleaves tripeptides from amino termini of peptides that bear free α -amino groups.⁴ Mutations in *CLN2* abolish TPP1 enzymatic activity in NCL2 patients.⁵ A total of 42 mutations, which include 10 small deletions/insertions, 20 missense and four nonsense point mutations, and eight splicing errors have been identified (<http://www.ucl.ac.uk/ncl>).⁶ We have reported that the heterogeneity of NCL with late infantile onset results from missense mutations in *CLN2* that cause atypical onset and clinical symptoms.⁷

NCL with late infantile onset has been reported to be common in the population in Newfoundland, Canada,⁸ but detailed genetic studies are lacking. Here we report molecular genetic analyses of 23 Canadian families that were clinically diagnosed with late infantile onset NCL.

MATERIALS AND METHODS

NCL families

Families that were clinically diagnosed as affected by late infantile onset NCL were collected from Canada, 21 from the Newfoundland area and two from the Toronto area. Informed consent signed by the participants' parent(s) and guardian(s) and clinical information, including age at onset, initial neurological symptoms, and ultrastructural profiles, were obtained through health care professionals.

Molecular studies

Molecular analyses for mutations in genes *CLN1*, *CLN2*, and *CLN3* were conducted with the two phase approaches we described earlier.⁷⁻¹¹ Clinically suspected cases were first analysed for common mutations in the *CLN2* gene, which account for about two-thirds of late infantile onset NCL cases globally.

These common mutations are IVS5-1G→C and R208X (636C→T) in *CLN2*.¹⁰ Analyses were conducted by allele specific PCR. If no common mutation(s) were detected for those cases that showed typical lipofuscin storage or were TPP1 enzyme deficient, which was determined by a simple assay using a synthetic peptide as substrate,¹² then the second phase, the gene scan procedures to search for uncommon mutations in the entire *CLN2* gene, were performed. The genomic locus of *CLN2* was PCR amplified, as we have described elsewhere,¹³ by four fragments to cover all exons, intron-exon junctions, partial sequences of promoter, and 3'-UTR regions. The amplified PCR product was gel purified and subjected to automatic DNA sequencing analyses by CEQ2000XL (Beckman Coulter, Fullerton, CA) with the same primers used for the genomic PCR reaction. For cases in which neither common mutations nor TPP1 deficiency were detected, analyses of common mutations in the *CLN3* gene were carried out.¹¹⁻¹³

Predictive structural analyses were performed by using the bioinformatic programs: Swiss-Prot (<http://www.expasy.ch>), PROSITE (<http://www.toulouse.inra.fr/prodom>), MultiAlign (<http://prodes.toulouse.inra.fr/multalin>), ClustalW (<http://www2.ebi.ac.uk>), BLINK (<http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/blink>), DART (Domain Architecture Retrieval Tool) (<http://www.ncbi.nlm.nih.gov/Structure>), Pfam alignments (<http://www.ncbi.nlm.nih.gov/Structure/cdd>), Protomap (<http://protomap.Stanford.edu>), GOR4 (http://npsa-pbil.ibcp.fr/cgi-bin/secpred_gor4), SOPMA (<http://npsa-pbil.ibcp.fr/cgi-bin/>

Key points

- Neuronal ceroid lipofuscinosis with late infantile onset is one of the most common childhood degenerative disorders. It has been reported in the Canadian population.
- Among 23 Canadian families studied using molecular methods, 20 were confirmed as having NCL2, the classical late infantile NCL, the underlying gene for which is *CLN2*.
- The mutation G284V, which is found only in Canadian NCL2 and accounts for 55% of the NCL2 families and 32% of the mutant alleles, indicates that G284V is a founder mutation.
- Along with other common *CLN2* mutations, –IVS5-1C, which accounts for 25% (10/40), and R208X, which accounts for 10% (4/40) of Canadian mutant NCL2 alleles, G284V should be studied in Canadian NCL2.
- Molecular study of these three mutations in Canadian late infantile onset NCL would detect 85% of NCL2 families and 67% of mutant NCL2 alleles. In addition, several novel mutations that are clustered in the distal coding sequence of the *CLN2* gene have been identified in Canadian NCL2 alleles.

Table 1 Mutations* identified in LINCL

Number	First symptoms	Age at onset (y)	EM findings	TPP1	CLN2 allele 1	CLN2 allele 2	CLN3 1.02 kb deletion	Genetic diagnosis
1	Seizures	3	CV		IVS5-1C	IVS5-1C		NCL2
2	Seizures	2.8	CV+GR		IVS5-1C	IVS5-1C		NCL2
3	Motor dysfunction	0.9	CV		IVS5-1C	IVS5-1C		NCL2
4	Seizures	3	CV		IVS5-1C	G284V		NCL2
5	Speech delay	2.5	?		IVS5-1C	K392N		NCL2
6	Developmental delay	3.5	CV		IVS5-1C	R208X		NCL2
7	Regression	2	CV		?	R208X		NCL2
8	Learning - slow	2	CV		G284V	R208X		NCL2
9	?	?	?		G284V	R208X		NCL2
10	Seizures	3	CV		G284V	G284V		NCL2
11	Seizures	3	?		G284V	G284V		NCL2
12	Developmental delay	3	CV		G284V	?		NCL2
13	Speech delay	2.5	?		G284V	IVS5-1C		NCL2
14	Developmental delay	0.1	?		G284V	F481C		NCL2
15	Seizures	3.5	CV		G284V	1-bp DEL		NCL2
16	Seizures	3	CV		G284V	?		NCL2
17	Seizures	3	CV		G284V	?		NCL2
18	?	?	?		V277M	?		NCL2
19	Seizures	2	CV		Q278P	?		NCL2
20	Progression	3	CV+FP		IVS12-1C	?		NCL2
21	Developmental delay	3.5	CV+FP	NL			NL/NL	Non-NCL2
22	Vision loss	4	CV	NL			NL/NL	Non-NCL2
23	Vision loss	3.5	FP				DEL/DEL	NCL3

* Novel mutations identified in this study have been deposited in the NCL mutation database (<http://www.ucl.ac.uk/ncl>).

?: information is not available.

EM: electron microscopy.

CV: curvilinear inclusions.

GR: granular osmiophilic deposit inclusions.

FP: fingerprint inclusions.

NL: normal.

DEL: deletion.

secpred_sopma), Jpred2 (<http://jura.ebi.ac.uk/8888/jpred>), Predator (http://www.embl-heidelberg.de/cgi/predator_serv.pl), Predictprotein (<http://cubic.bioc.columbia.edu/predictprotein/>), Tmpred (<http://searchlauncher.bcm.tmc.edu>).

RESULTS

Detection of CLN2 common mutations

Previously, we identified two common mutations, the splicing mutation IVS5-1G→C and the nonsense mutation R208X (636C→T), in the *CLN2* gene in the American population, which is of mixed ethnicity. Analysis of these two common mutations detect about 70-75% of clinically suspected late infantile NCL cases with typical EM profile.⁹ In the current study, these two mutations were detected initially in 50% (10/20) of Canadian NCL2 families and 35% (14/40) of *CLN2* alleles.

Mutation G284V presents a founder mutation in Canadian NCL2

In addition to the common mutations described above, a mutation, G284V, that we reported earlier⁷ in a Canadian NCL2 family, was found in 11 families and 13 alleles, which accounts for 55% and 32.5%, respectively, in this study. To our knowledge, this mutation has been found only in Canadian NCL2 families, not in any other ethnicities, which provides evidence that the G284V mutation may be a new mutation occurring in settler(s) who originated from the British Isles⁸ and presenting as a founder effect involved in the development of NCL2.

Identification of novel mutations and polymorphisms in the CLN2 gene

Among the total *CLN2* alleles studied, five novel mutations, V277M, Q278P, F481C, 1-bp deletion, and IVS12-1G→C, were identified (table 1). The V277M, Q278P, and 1-bp deletion are localised within exon 7, F481C is in exon 12, and IVS12-1G→C is involved in an abnormal splicing of exon 12. In addition to

novel mutations, three single nucleotide polymorphisms (SNPs), g3841C→T (IVS6-10C→T), g4066T→C (IVS7+17T→C), and g4090G→C (IVS7+41G→C), which flank exon 7 were identified from the *CLN2* genomic sequence (NCBI, nucleotide accession No AF039704).

Genetic heterogeneity

Among the total of 23 families with clinically diagnosed late infantile NCL, three families were determined not to have a genetic deficiency of *CLN2*. Normal TPP1 activity was observed in two families, and homozygosity for the 1.02 kb deletion in the *CLN3* gene, which was confirmed by testing other members in the families, was found in the proband of the third family. This result indicated that families with clinically diagnosed late infantile onset NCL were not affected by NCL2 but instead by NCL3 or NCL1 and/or other uncharacterised NCL variants such as NCL6.¹⁴

Analyses of the possible effects of the G284V mutation

Homology and structural searches using bioinformatic tools showed that TPP1 protein sequences are almost identical in mouse, rat, and dog. A more distantly related protein is found in *Amoeba proteus* pepstatin insensitive carboxyl proteinase 2 (PICP2). Two C-terminal domain active sites of TPP1 are in a somewhat conserved domain that belongs to that in TPP2, kexin, and subtilisin (the pfam00082, subtilase domain). TPP1 is related to 10 other proteins in cluster 3669 of Protomap. No BioSpace model is available yet for the family of tripeptidyl-peptidases, but the three dimensional model for Pseudomonaspepsin precursor, chain A (1GA1A, aa318-521), contains four regions that are highly similar to TPP1.

The G284V mutation, however, is closer to the N-terminal of the active protease. It is in a stretch of 27 amino acid residues that are completely conserved in human, rat, mouse, and dog TPP1. In all other more remotely related proteins, the G is absolutely conserved; no conservative substitutions are allowed. G284 is separated by five residues from two other

upstream sites in which mutations are linked to *CLN2*, V277M and Q278P. A potential N-glycosylation site is just downstream at residues 286-289. Two predicted N-myristoylation sites are nearby also (starting at residues 278 and 287). Every secondary structure prediction method indicated that G284 is located in a region of random coil (with beta turn in one prediction) between an alpha helical region to the N-terminal site and a beta sheet region C-terminally.¹⁵ Substituting a Val (V) residue at position 284 invariably results in the predicted lengthening of the preceding alpha helix and a change in the conformation of the immediately downstream sequence.

DISCUSSION

Late infantile onset NCL was reported as a common childhood neurodegenerative disorder in a Canadian population in the Newfoundland area.⁸ However, characterisation with genetic studies of the Newfoundland NCL cases has been lacking. In this study, we analysed 21 families from Newfoundland which had been diagnosed with clinical and pathological approaches⁸ as being affected by late infantile NCL, as well as two new families from Toronto. Our results confirmed that 20 families have *CLN2* deficiency (table 1) and can be genetically classified as having NCL2; among these families, 50% (10/20) were identified initially by studying two common mutations we reported earlier in mixed populations.^{5,9} Three families were homozygous for mutation of IVS5-1C (previously described as T523-1G→C), six families were heterozygous for either IVS5-1C or R208X, and one was doubly heterozygous for both IVS5-1C and R208X. Our data obtained from this particular population further support the hypothesis that these two common mutations may derive from recurrence and the conclusion that testing for these two common mutations in *CLN2* should be the initial attempt to identify NCL2 patients/families and be applied for clinical molecular diagnosis.^{10,11}

G284V, a mutation localised in exon 7 of the *CLN2* gene, was identified in a Canadian NCL2 patient.⁷ To our surprise, this mutation is present in 55% (11/20) of NCL2 families and 32.5% (13/40) of *CLN2* mutant alleles in this study. So far, this mutation has not been found in any populations other than the Canadian, indicating that the mutation of G284V plays a founder effect in the development of NCL2 in this Canadian population and should be applied for molecular testing for Canadian NCL2. In fact, combining G284V with the two other common mutations, we can detect 85% (17/20) of Canadian NCL2 families and 67.5% (27/40) of mutant *CLN2* alleles. This information is valuable for clinical molecular screening for NCL2 in all families with clinically identified late infantile onset NCL.

Three families listed in table 1, which were found initially to have neither G284V nor two common mutations, turned out either to carry normal TPP1 activity (non-NCL2) or to be homozygous for the 1.02 kb deletion in the *CLN3* gene for NCL3, indicating the existence of genetic heterogeneity. In addition, heterogeneity was also noticed in the age at onset and EM profiles. The majority of NCL2 families have typical onset at 2 to 3 years of age and CV inclusions. However, two families showed that the probands started to have either neuromotor dysfunction or developmental delay before or during the infantile period, a pattern often misdiagnosed as NCL1 underlain by the *CLN1* gene. An ultrastructural profile of granular osmiophilic deposit inclusions (GR) is usually the typical EM finding for NCL1 (infantile onset NCL), and of fingerprint inclusions (FP) for NCL3 (juvenile onset NCL). In this study, two families showed mixed inclusions, CV+GR and CV+FP, that associate with mutations involved in intron-exon junctions (IVS5-1C and IVS12-1C), causing splicing errors; this finding provides evidence that NCL2 can result from mixed profiles. In addition, we have noticed that all seizures, the most common initial symptoms of NCL2 except one, are associated with homozygotes or heterozygotes for mutations IVS5-1C and/or G284V.

Although no mutation was identified in 17.5% (7/40) of the NCL2 alleles owing to the limited DNA material available, identification of novel mutations and polymorphisms around exon 7 suggests that DNA sequences in this region are highly mutable. Mutations resulted in substitutions of amino acids or frame shift, which would cause a change in protein folding.

The G284V mutation occurs in a highly conserved region and is absolutely conserved in many related proteases. The substitution of valine results in a predicted lengthening of the alpha helix. One transmembrane domain predicting algorithm indicated that this increased length might be enough to cause this domain to become membrane associated rather than freely soluble. The increased bulk of a substituted valine might also prevent the peptide chain from folding properly at that point. The more restricted rotation of valine compared to glycine would also add some constraint to the conformation adopted by the protein. A conformational alteration of the N-glycosylation or N-myristoylation sites nearby might also result from the G284V substitution. Critical glycine residues are often found in short loops between classical secondary structures. Many proteases contain a flap region in the vicinity of the active site that contains an invariable glycine residue. This flap region protects the protease from solvent and participates in substrate binding. The flap region can differ for enzymes with similar catalytic activity but different substrate specificities. Analysis of the change predicted in the G284V mutation TPP1 secondary structure suggests that a major local conformational change would accompany such a mutation, resulting in an alteration or elimination of the protease's activity.

In addition to late infantile onset NCL (LINCL), genetic heterogeneity has also been documented in infantile NCL (INCL) and juvenile NCL (JNCL).^{7,16,17} To distinguish the clinically identified NCL from genetically identified NCL, a new nomenclature system using NCL1, 2, 3, etc, corresponding to the genetic loci *CLN1*, 2, 3, etc, was suggested recently.¹⁶ Although this system works for genetic classification, use of NCL1, 2, 3, etc, would be confusing in clinical classification because the underlying genetic deficiency would be unknown. Therefore, we recommend that a more practical "dual nomenclature" system (table 2), both clinical and genetic, should be applied for NCL. We recommend that INCL, LINCL and JNCL, etc, continue to be used by physicians to identify clinically diagnosed infantile, late infantile, and juvenile onset NCLs. However, NCL1, NCL2, NCL3, etc., should be used only for genetically confirmed NCLs that correspond to the underlying

Table 2 A dual nomenclature system for NCLs

Clinical onset	Clinically diagnosed NCLs*	Genetic loci/genes	Genetically diagnosed NCLs
Infantile	INCL	<i>CLN1</i> <i>CLN2</i>	NCL1 NCL2
Late infantile	LINCL	<i>CLN1</i> <i>CLN2</i> <i>CLN3</i> <i>CLN5</i> <i>CLN6</i> <i>CLN7</i> <i>CLN8</i>	NCL1 NCL2 NCL3 NCL5 NCL6 NCL7 NCL8
Juvenile	JNCL	<i>CLN1</i> <i>CLN2</i> <i>CLN3</i>	NCL1 NCL2 NCL3
Adult	ANCL	<i>CLN1</i> <i>CLN4</i> †	NCL1 NCL4

*Zhong.¹⁸

†Locus/gene not yet identified.

genes *CLN1*, *CLN2*, and *CLN3*, etc. This dual system may solve the current confusion for both clinicians and geneticists.

ACKNOWLEDGEMENTS

This study was supported in part by grants from the New York State Office of Mental Retardation and Developmental Disabilities (OMRDD), the Batten Disease Support and Research Foundation (BDSRA), and the Children's Brain Diseases Foundation (CBDF). The authors would like to thank the families/subjects who participated in our research studies. Jennifer Shen provided valuable assistance in the bioinformatic studies.

Authors' affiliations

W Ju, R Zhong, D Moroziewicz, N Zhong, SCL-Molecular Neurogenetic Diagnostic Laboratory, NYS Institute for Basic Research, 1050 Forest Hill Road, Staten Island, NY, USA

W Ju, D Moroziewicz, J R Currie, W T Brown, N Zhong, Department of Human Genetics, NYS Institute for Basic Research, 1050 Forest Hill Road, Staten Island, NY, USA

N Zhong, Department of Neurology, SUNY Downstate Health Center, Brooklyn, NY, USA

R Zhong, Stuyvesant High School, New York, NY, USA

S Moore, P Parfrey, Clinical Epidemiology Unit, Health Sciences Center, Memorial University of Newfoundland, St John's, NF, Canada

Correspondence to: Dr N A Zhong, SCL-Molecular Neurogenetic Diagnostic Laboratory, NYS Institute for Basic Research in Developmental Disabilities, 1050 Forest Hill Road, Staten Island, NY 10314, USA; ibrsclmdl@aol.com

REFERENCES

- 1 **Wisniewski KE**, Kida E, Golabek AA, Kaczmarek W, Connell F, Zhong N. Neuronal ceroid lipofuscinoses: classification and diagnosis. *Adv Genet* 2001;**45**:1-33.
- 2 **Sharp JD**, Wheeler RB, Lake BD, Savukoski M, Jarvela IE, Peltonen L, Gardiner RM, Williams RE. Loci for classical and a variant late infantile neuronal ceroid lipofuscinosis map to chromosome 11p15 and 15q21-23. *Hum Mol Genet* 1997;**6**:591-6.
- 3 **Liu CL**, Sleat DE, Donnelly RJ, Lobel P. Structural organization and sequence of *CLN2*, the defective gene in classical late infantile neuronal ceroid lipofuscinosis. *Genomics* 1998;**50**:206-12.
- 4 **Junaid MA**, Wu G, Pullarkat RK. Purification and characterization of bovine brain lysosomal pepstatin-insensitive proteinase, the gene product deficient in the human late-infantile neuronal ceroid lipofuscinosis. *J Neurochem* 2000;**74**:287-94.
- 5 **Sleat DE**, Gin RM, Sohar I, Wisniewski K, Sklower-Brooks S, Pullarkat RK, Palmer DN, Lerner TJ, Boustany RM, Uldall P, Siakotos AN, Donnelly RJ, Lobel P. Mutational analysis of the defective protease in classic late-infantile neuronal ceroid lipofuscinosis, a neurodegenerative lysosomal storage disorder. *Am J Hum Genet* 1999;**64**:1511-23.
- 6 **Mole S**, Zhong N, Sarpong A, Logan WP, Hofmann S, Yi W, Franken PF, van Diggelen OP, Breuning MH, Moroziewicz D, Ju W, Bauer A, Jung M, Stoltenberg-Didinger G, Reis A, von Mores A, Hubner C, Shuelke M, Salonen T, Holmberg V, Jarvela I, Taschner EM. New mutations in the neuronal ceroid lipofuscinosis genes. *Eur J Pediatr Neurol* 2001;**5**(suppl):7-10.
- 7 **Zhong N**, Moroziewicz DN, Jurkiewicz A, Ju W, Johnston L, Wisniewski KE, Brown WT. Genetic heterogeneity of late-infantile neuronal ceroid lipofuscinosis (LINCL). *Genet Med* 2000;**2**:312-18.
- 8 **Andermann E**, Jacob JC, Andermann F, Carpenter S, Wolfe L, Berkovic S. The Newfoundland aggregate of neuronal ceroid-lipofuscinosis. *Am J Med Genet* 1988;**5**(suppl):111-16.
- 9 **Zhong N**, Ju W, Moroziewicz DN, Kaczmarek AL, Hartikainen J, Suarez CT, McLendon L, Wisniewski KE, Jenkins EC, Brown WT. Molecular diagnosis of and carrier screening for the neuronal ceroid lipofuscinoses (NCLs). *Genet Testing* 2000;**4**:243-8.
- 10 **Zhong N**, Wisniewski KE, Hartikainen J, Ju W, Moroziewicz DN, McLendon L, Sklower Brooks SS, Brown WT. Two common mutations in the *CLN2* gene underlie late infantile neuronal ceroid lipofuscinosis. *Clin Genet* 1998;**54**:234-8.
- 11 **Zhong N**, Wisniewski KE, Kaczmarek AL, Ju W, Xu WM, Xu WW, McLendon L, Liu B, Kaczmarek W, Sklower-Brooks S, Brown WT. Identification of a novel missense mutation (E295K) underlying Batten disease. *Hum Genet* 1998;**102**:57-62.
- 12 **Junaid MA**, Sklower Brooks S, Wisniewski KE, Pullarkat RK. A novel assay for lysosomal pepstatin-insensitive proteinase and its application for the diagnosis of late-infantile neuronal ceroid lipofuscinosis. *Clin Chim Acta* 1999;**281**:169-76.
- 13 **Zhong N**. Molecular genetic testing for neuronal ceroid lipofuscinoses. *Adv Genet* 2001;**45**:217-24.
- 14 **Gao H**, Boustany RM, Espinola JA, Cotman SL, Srinidhi L, Antonellis AK, Gillis T, Qin X, Liu S, Donahue LR, Bronson RT, Faust JR, Stout D, Haines JL, Lerner TJ, MacDonald ME. Mutations in a novel *CLN6*-encoded transmembrane protein cause variant neuronal ceroid lipofuscinosis in man and mouse. *Am J Hum Genet* 2002;**70**:324-35.
- 15 **Andreeva NS**, Bochkarev A, Pechik I. A new way of looking at aspartic proteinase structure: a comparison of pepsin structure to other aspartic proteinases in the near active site region. In: Takahashi K, ed. *Aspartic proteinases: structure, function, biology and biomedical implications*. New York: Plenum Press, 1995:19-32.
- 16 **van Diggelen OP**, Thobois S, Tilikete C, Zabol MT, Keulemans JLM, van Bunderen PA, Taschner PEM, Losekoot M, Voznyi YY. Adult neuronal ceroid lipofuscinosis with palmitoyl-protein thioesterase deficiency: first adult-onset patients of a childhood disease. *Ann Neurol* 2001;**50**:269-72.
- 17 **Wisniewski KE**, Zhong N, Kida E, Kaczmarek W, Kaczmarek AL, Connell F, Sklower-Brooks S, Brown WT. Atypical late infantile and juvenile forms of neuronal ceroid lipofuscinosis and their diagnostic difficulties. *Folia Neuropathol* 1997;**35**:73-9.
- 18 **Zhong N**. Neuronal ceroid lipofuscinoses and possible pathogenic mechanism. *Mol Genet Metab* 2000;**71**:195-206.