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

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.

Genetically, the gene CLN2 underlying NCL2 has been mapped to chromosome 11p15. It consists of 13 exons spanning 6.65 kb of genomic DNA 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. Mutations in CLN2 abolish TPP1 enzymatic activity in NCL2 patients. 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). 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.

NCL with late infantile onset has been reported to be common in the population in Newfoundland, Canada, 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. 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. 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, 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, 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. In cases in which neither common mutations nor TPP1 deficiency were detected, analyses of common mutations in the CLN3 gene were carried out.

evidence that the G284V mutation may be a new mutation for 55% and 32.5%, respectively, in this study. To our family, was found in 11 families and 13 alleles, which accounts for 50% (10/22) of Canadian NCL2 cases with typical EM profile.

Identification of novel mutations and polymorphisms in the CLN2 gene

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.

Genetic heterogeneity

Table 1 Mutations* identified in LINCL

<table>
<thead>
<tr>
<th>Number</th>
<th>First symptoms</th>
<th>Age at onset (y)</th>
<th>EM findings</th>
<th>TPP1</th>
<th>CLN2 allele 1</th>
<th>CLN2 allele 2</th>
<th>CLN3 1.02 kb deletion</th>
<th>Genetic diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Seizures</td>
<td>3</td>
<td>CV</td>
<td></td>
<td>IVS5-1C</td>
<td>IVS5-1C</td>
<td>?</td>
<td>NCL2</td>
</tr>
<tr>
<td>2</td>
<td>Seizures</td>
<td>2.8</td>
<td>CV+GR</td>
<td></td>
<td>IVS5-1C</td>
<td>IVS5-1C</td>
<td>R208X</td>
<td>NCL2</td>
</tr>
<tr>
<td>3</td>
<td>Motor dysfunction</td>
<td>0.9</td>
<td>CV</td>
<td></td>
<td>IVS5-1C</td>
<td>IVS5-1C</td>
<td>K392N</td>
<td>NCL2</td>
</tr>
<tr>
<td>4</td>
<td>Seizures</td>
<td>3</td>
<td>CV</td>
<td></td>
<td>IVS5-1C</td>
<td>G284V</td>
<td>?</td>
<td>NCL2</td>
</tr>
<tr>
<td>5</td>
<td>Speech delay</td>
<td>2.5</td>
<td>?</td>
<td></td>
<td>IVS5-1C</td>
<td>R208X</td>
<td>?</td>
<td>NCL2</td>
</tr>
<tr>
<td>6</td>
<td>Developmental delay</td>
<td>3.5</td>
<td>CV</td>
<td></td>
<td>?</td>
<td>G284V</td>
<td>R208X</td>
<td>NCL2</td>
</tr>
<tr>
<td>7</td>
<td>Regression</td>
<td>2</td>
<td>CV</td>
<td></td>
<td>?</td>
<td>G284V</td>
<td>R208X</td>
<td>NCL2</td>
</tr>
<tr>
<td>8</td>
<td>Learning - slow</td>
<td>?</td>
<td>CV</td>
<td></td>
<td>?</td>
<td>G284V</td>
<td>?</td>
<td>NCL2</td>
</tr>
<tr>
<td>10</td>
<td>Seizures</td>
<td>3</td>
<td>CV</td>
<td></td>
<td>G284V</td>
<td>G284V</td>
<td>?</td>
<td>NCL2</td>
</tr>
<tr>
<td>11</td>
<td>Seizures</td>
<td>3</td>
<td>?</td>
<td></td>
<td>G284V</td>
<td>?</td>
<td>?</td>
<td>NCL2</td>
</tr>
<tr>
<td>12</td>
<td>Developmental delay</td>
<td>3</td>
<td>CV</td>
<td></td>
<td>G284V</td>
<td>?</td>
<td>?</td>
<td>NCL2</td>
</tr>
<tr>
<td>13</td>
<td>Speech delay</td>
<td>2.5</td>
<td>?</td>
<td></td>
<td>G284V</td>
<td>IVS5-1C</td>
<td>?</td>
<td>NCL2</td>
</tr>
<tr>
<td>14</td>
<td>Developmental delay</td>
<td>0.1</td>
<td>?</td>
<td></td>
<td>G284V</td>
<td>F481C</td>
<td>?</td>
<td>NCL2</td>
</tr>
<tr>
<td>15</td>
<td>Seizures</td>
<td>3.5</td>
<td>CV</td>
<td></td>
<td>G284V</td>
<td>1-bp DEL</td>
<td>?</td>
<td>NCL2</td>
</tr>
<tr>
<td>16</td>
<td>Seizures</td>
<td>3</td>
<td>CV</td>
<td></td>
<td>G284V</td>
<td>?</td>
<td>?</td>
<td>NCL2</td>
</tr>
<tr>
<td>17</td>
<td>Seizures</td>
<td>3</td>
<td>CV</td>
<td></td>
<td>G284V</td>
<td>?</td>
<td>?</td>
<td>NCL2</td>
</tr>
<tr>
<td>19</td>
<td>Seizures</td>
<td>2</td>
<td>CV</td>
<td></td>
<td>G2778P</td>
<td>?</td>
<td>?</td>
<td>NCL2</td>
</tr>
<tr>
<td>20</td>
<td>Progression</td>
<td>3</td>
<td>CV+FP</td>
<td></td>
<td>IVS12-1C</td>
<td>?</td>
<td>?</td>
<td>NCL2</td>
</tr>
<tr>
<td>21</td>
<td>Developmental delay</td>
<td>3.5</td>
<td>CV+FP</td>
<td></td>
<td>NL/NL</td>
<td>Non-NCL2</td>
<td>?</td>
<td>NCL3</td>
</tr>
<tr>
<td>22</td>
<td>Vision loss</td>
<td>4</td>
<td>CV</td>
<td></td>
<td>NL/NL</td>
<td>Non-NCL2</td>
<td>?</td>
<td>NCL2</td>
</tr>
<tr>
<td>23</td>
<td>Vision loss</td>
<td>3.5</td>
<td>FP</td>
<td></td>
<td>DEL/DEL</td>
<td>NCL3</td>
<td>?</td>
<td>NCL2</td>
</tr>
</tbody>
</table>

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

Results 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 Acanthamoeba 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 Pseudomonas aeruginosa subtilisin A (1G1A, aa318-521), contains four regions that are highly similar to TPP1.

Heterogeneity and genetic counselling

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.

Analysis of mutations

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.


deletion. NCL2 presents a founder mutation in Canadian NCL2

In addition to the common mutations described above, a mutation, G284V, that we reported earlier 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 cases with typical EM profile. 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.

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).

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 is 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 \textit{CLN2V277M} and \textit{G278F}. 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 second-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.\textsuperscript{a} 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.

\textbf{DISCUSSION}

Late infantile onset NCL was reported as a common childhood neurodegenerative disorder in a Canadian population in the Newfoundland area.\textsuperscript{1} However, characterisation with genetic studies of the Newfoundland NCL cases has been lacking. In this study, we analysed 21 families from Newfoundland who had been diagnosed with clinical and pathological approaches\textsuperscript{a} as being affected by late infantile NCL, as well as two new families from Toronto. Our results confirmed that 20 families have \textit{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.\textsuperscript{1,2} 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 \textit{CLN2} should be the initial attempt to identify NCL2 patients/families and be applied for clinical molecular diagnosis.\textsuperscript{10,11}

G284V, a mutation localised in exon 7 of the \textit{CLN2} gene, was identified in a Canadian NCL2 patient.\textsuperscript{1} To our surprise, this mutation is present in 55\% (11/20) of NCL2 families and 32.5\% (13/40) of \textit{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 \textit{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 TPPI activity (non-NCL2) or to be homozygous for the 1.02 kb deletion in the \textit{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 neurological dysfunction or developmental delay before or during the infantile period, a pattern often misdiagnosed as NCL1 underlaid by the \textit{CLN1} gene. An ultrastructural profile of granular osmiophilic deposit inclusions (GR) is usually the underlaid by the \textit{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 TPPI downstream 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).\textsuperscript{1,2} To distinguish the clinically identified NCL from genetically identified NCL, a new nomenclature system using NCL1, 2, 3, etc. corresponding to the genetic loci \textit{CLN1}, 2, 3, etc., was suggested recently.\textsuperscript{12} 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

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|c|}
\hline
\textbf{Clinical onset} & \textbf{Clinically diagnosed NCLs*} & \textbf{Genetic loci/genes} & \textbf{Genetically diagnosed NCLs} \\
\hline
\textbf{Infantile} & LINCL & \textit{CLN1} & NCL1 \\
 & & \textit{CLN2} & NCL2 \\
\hline
\textbf{Late infantile} & LINCL & \textit{CLN1} & NCL1 \\
 & & \textit{CLN2} & NCL2 \\
 & & \textit{CLN3} & NCL3 \\
\hline
\textbf{Juvenile} & JNCL & \textit{CLN1} & NCL1 \\
 & & \textit{CLN2} & NCL2 \\
 & & \textit{CLN3} & NCL3 \\
\hline
\textbf{Adult} & \textit{ANCL} & \textit{CLN4} & NCL4 \\
\hline
\end{tabular}
\caption{A dual nomenclature system for NCLs}
\end{table}

\textsuperscript{*}Zhong.\textsuperscript{13} \textsuperscript{†}Locus/gene not yet identified.

\textsuperscript{1} \textsuperscript{10} \textsuperscript{11} \textsuperscript{12} \textsuperscript{13}
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

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

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; ibrcslmndl@aol.com

REFERENCES


