Genotype-phenotype correlation in L1 associated diseases

Erik Fransen, Guy Van Camp, Rudi D’Hooge, Lieve Vits, Patrick J Willems

Abstract
The neural cell adhesion molecule L1 (L1CAM) plays a key role during embryonic development of the nervous system and is involved in memory and learning. Mutations in the L1 gene are responsible for four X linked neurological conditions: X linked hydrocephalus (HSAS), MASA syndrome, complicated spastic paraplegia type 1 (SP-1), and X linked agenesis of the corpus callosum. As the clinical picture of these four L1 associated diseases shows considerable overlap and is characterised by Corpus callosum hypoplasia, mental Retardation, Adducted thumbs, Spastic paraplegia, and Hydrocephalus, these conditions have recently been lumped together into the CRASH syndrome. We investigate here whether a genotype-phenotype correlation exists in CRASH syndrome since its clinical spectrum is highly variable and numerous L1 mutations have been described. We found that (1) mutations in the extracellular part of L1 leading to truncation or absence of L1 cause a severe phenotype, (2) mutations in the cytoplasmic domain of L1 give rise to a milder phenotype than extracellular mutations, and (3) extracellular missense mutations affecting amino acids situated on the surface of a domain cause a milder phenotype than those affecting amino acids buried in the core of the domain. (J Med Genet 1998;35:399–404)

Keywords: L1; cell adhesion molecule; genotype-phenotype correlation; mental retardation

L1 (L1CAM) is a cell adhesion molecule (CAM) belonging to the immunoglobulin superfamily (IgSF). This surface glycoprotein consists of six immunoglobulin- and five fibronectin III-like domains, a single pass transmembrane domain, and a short cytoplasmic domain. L1 has cell adhesive properties and is involved in cell signalling. The extracellular domains are involved in homo- and heterophilic interactions with various ligands, including L1 itself, axonin-1/TAG-1, F3/F11, D-MGRASP, αβ integrin, and phosphacan. In addition, cis interactions with NCAM and nectadrin have been reported. The cytoplasmic domain of L1 is multifunctional: it provides an anchor to the cytoskeleton via ankyrin, it contains ser phosphorylation sites, and it communicates with various intracellular second messenger pathways including cAMP dependent signalling cascades, non-receptor tyrosine kinases, and a pathway which passes through activation of the FGF receptor. L1 is involved in the embryological development of the brain, as well as in cognitive function and memory. The most prominent expression of L1 is found on the membrane of axons and growth cones, and on Schwann cells during development of the nervous system, where it plays a key role in cell-cell and cell-substrate adhesion, axon outgrowth, path-finding and fasciculation, neuronal migration, and myelination. In addition, L1 is involved in the establishment of long term potentiation (LTP) and long term memory in the hippocampus later on. L1 expression has also been detected outside the nervous system, in the male urogenital tract, in intestinal crypt cells, and in cells of haematopoetic and tumoral origin. The non-neural L1 isoform lacks four amino acids in its cytoplasmic domain owing to alternative splicing.

X linked hydrocephalus (HSAS), MASA syndrome, X linked complicated spastic paraparesis type 1 (SP-1), and X linked agenesis of the corpus callosum (ACC) are four X linked neurological diseases with variable though overlapping clinical spectra. We and others have shown that each of these conditions is caused by mutations in the gene encoding L1. Since the main features of L1 associated diseases include Corpus callosum hypoplasia, mental Retardation, Adducted thumbs, Spastic paraplegia, and Hydrocephalus, this condition is now referred to as CRASH syndrome. The clinical spectrum of CRASH syndrome is highly variable and may involve any combination of these symptoms. At the severe end, there might be patients with massive hydrocephalus resulting in pre- or perinatal death, whereas at the mild end there are patients with mild mental retardation as their only abnormality.

So far, 77 different L1 mutations have been reported in families with CRASH syndrome. The mutational spectrum of L1 includes two gross rearrangements, 19 mutations leading to a frameshift (15 small deletions, two small insertions, and two complex mutations), 14 mutations leading to aberrant splicing, nine nonsense mutations, three in frame deletions, and 30 missense mutations. Most mutations are private, occurring in only a single family. We have constructed the L1 Mutation Web Page, a www site containing information about L1, including a continuously updated table of all published and many unpublished L1 mutations (http://dnalab-www.uia.ac.be/dnalab/l1).
Patients and methods

COLLECTION OF CLINICAL DATA
In order to collect clinical data, we distributed standardized questionnaires among the clinicians who had studied CRASH families with a known L1 mutation. The questionnaire contained questions on clinical abnormalities. Questionnaires containing information on 39 patients from 13 CRASH families were received. Additionally, 108 patients from 23 families were included, based upon information derived from published reports.

CLASSIFICATION OF PATIENTS
In a search for a possible genotype-phenotype correlation, we defined clinical criteria allowing us to score the severity of the phenotype in every patient and every family. We did not use the previous subdivision into HSAS and MASA families, as families were often classified according to the index patient only. In view of the intrafamilial variation, this resulted in MASA patients belonging to HSAS families and vice versa.

A patient was scored as "severely" affected if he either died before the age of 2 years because of complications of CRASH syndrome, or if he was born macrocephalic (OFC >97th centile), or needed shunting. A patient was scored as "mildly" affected if he survived beyond the age of 2 without shunting or macrocephaly. We did not take into account the IQ of the patients, since it is impossible to estimate IQ in fetuses, stillbirths, or severely affected patients dying in early childhood. The presence of adducted thumbs, spastic paraplegia, or hypoplasia of the corpus callosum was not scored either, since very few case reports mention these abnormalities and, moreover, they are hard to quantify.

After we had scored every patient separately, we also scored each family in which data from at least two patients were available. A family was scored as "severely" or "mildly" affected, if all known patients of that particular family were respectively severely or mildly affected. A family containing both severely and mildly affected patients was scored as "variable". Only families from which information on two or more patients was available were included in our study. Scoring of the patients and families was done "blindly", without previous knowledge about the mutation in that family.

CLASSIFICATION OF L1 MUTATIONS
So far, 77 different L1 mutations in 81 unrelated families have been described and listed in the L1 Mutation Web Page. Of these, 36 L1 mutations were used in the genotype-phenotype correlation study.

The L1 mutations were subdivided into four classes, according to the effect they are expected to have on the L1 protein. The classification presented here differs slightly from the classification we used in previous reports, in that we now classify the extracellular splice site mutations as a distinct group. The four different classes are defined as follows.

(1) Class I includes mutations in the extracellular part of L1 expected to lead to a truncation or absence of the protein. These mutations include frameshift mutations (small deletions or insertions) or point mutations leading to a stop codon.
(2) Class II includes missense mutations resulting in an amino acid substitution in the extracellular part of L1. Within class II mutations, a subdivision can be made between those mutations affecting “key amino acid residues” and those affecting “surface residues”. The former group substitutes a highly conserved amino acid in the core of a domain, which is crucial for the structural build up of that L1 domain, or even the entire L1 protein. The latter group of mutations affects a less conserved residue at the surface of a domain, with presumably less influence on the structure of the L1 domain and the whole protein.

(3) Class III includes any mutation in the cytoplasmic domain. Mutations in the cytoplasmic domain lead to the presence of a mutated or truncated L1 that is still attached to the cell membrane, as the extracellular and transmembrane domains are intact. The mutations are expected to influence signal transduction, although cytoplasmic mutations might also alter cell adhesion by the extracellular part of L1.

(4) Class IV consists of extracellular mutations resulting in aberrant splicing of pre-mRNA. These splice site mutations are caused by mutations in the highly conserved sequences flanking the intron-exon boundary, in the branch point signal, or in a conserved XGGG sequence in the 5' end of an intron. We have classified these mutations as a separate group, as they might give rise to the presence of several mRNA splice forms besides the normally spliced mRNA, and the effect of these mutations on the protein is therefore unpredictable.

STATISTICAL ANALYSIS
Under the null hypothesis, we assumed no correlation between genotype and phenotype. The distribution of the three phenotypes (severe, mild, variable) among the four classes of L1 mutations was analysed using the $\chi^2$ test with Yates's correction for small expected frequencies. If the number of patients was too small, data were grouped into a 2 x 2 matrix and a Fisher exact test was carried out.

Results

CLINICAL DATA
A study of published reports on 25 different CRASH families with a known L1 mutation yielded enough clinical data to classify 108 patients into the different clinical categories “severe” or “mild”, and 23 families into “severe”, “mild,” or “variable”. In addition, the questionnaire sent to clinicians caring for CRASH patients with a known L1 mutation yielded clinical, pathological, or neuroradiological data from another 39 patients originating from 13 families. Taken together, we gathered detailed data from a total of 147 patients from 36 different families with a known L1 mutation. From the total of 36 families, 25 fitted the criteria for “severely affected”, seven for “mildly affected”, and four were classified as “variable”. We found five patients, originating from two families, which we were unable to fit into these criteria (patients 1 and 2 reported by Schrander-Stumpel et al, patients 1, 2, and 3 reported by Winter et al). Four of these patients survived into adulthood, although having an OFC > 97th centile. These two families were omitted from the study and not included in table 1. A family reported by Varadi et al. was also omitted since the phenotype of one of the family members was unclear. Table 1 lists the scores of the families with an identified L1 mutation, from which sufficient clinical data were available.

Most families have a fairly uniform phenotype. Among the severely affected families are many unpublished families. Only four of the families show significant intrafamilial variability with both severely and mildly affected patients.

L1 MUTATIONS
The 36 mutations we have studied here include two gross rearrangements (a 2 kb deletion and a 1.3 kb duplication), nine mutations leading to a frameshift (eight small insertions, one complex mutation), eight mutations leading to aberrant splicing, two in frame nonsense mutations, and 15 missense mutations. Classifying the mutations into the four classes mentioned above subdivided the mutations into 10 extracellular truncation mutations (class I), 14 extracellular missense mutations (class II), five cytoplasmic mutations (class III), and seven extracellular mutations affecting pre-mRNA splicing (class IV).

GENOTYPE-PHENOTYPE CORRELATION
To study genotype-phenotype correlation, we investigated how the phenotypes of the families were distributed among the four classes of mutations. The overall distribution of the phenotypes is given in table 2. Chi-squared testing indicated that this distribution was non-random, showing that the distribution of the families among the three phenotypes and the type of mutation were significantly related ($\chi^2 = 13.12, df=6, p=0.041$). Two observations are noteworthy. Firstly, every family with an extracellular truncation (class I) mutation shows a severe phenotype, never a variable or mild one. Secondly, although a majority of the families in our study shows the severe phenotype, we did not find any severely affected patient having a cytoplasmic mutation (class III) and this latter group of mutations seems to lead to a variable or mild phenotype.

To investigate whether the observed differences were significant, we analysed the data using the Fisher exact test, which required regrouping of the data into 2 x 2 contingency

<table>
<thead>
<tr>
<th>Mutation class</th>
<th>Severe</th>
<th>Variable</th>
<th>Mild</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>I Extracellular truncation</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>II Extracellular missense</td>
<td>9</td>
<td>2</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>III Cytoplasmic</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>IV Splice site</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
<td>4</td>
<td>7</td>
<td>36</td>
</tr>
</tbody>
</table>

$\chi^2=13.12, df=6, p=0.041$
tables. First, we compared the relative proportion of the phenotypes within the group of 10 families affected by an extracellular truncation (class I) mutation with the proportion within the group of the 26 families affected by a mutation of a different class (table 3). A Fisher exact test, in which the phenotypes were scored as "severe" and "non-severe", showed significantly different phenotype proportions between the two mutation groups (p=0.016). Similarly, we compared the phenotype proportion of the five families with a class III intracellular mutation with the phenotype proportion of the 31 families with an extracellular mutation (table 4). A Fisher exact test, whereby the phenotypes were scored as "mild" and "non-mild", again indicated significant differences between the two mutation groups (p=0.027).

Class II extracellular missense mutations give rise to all three phenotypes, and the distribution of the 14 families was not significantly different from the expected distribution. Class II mutations can be subdivided into mutations affecting "key residues" and those affecting "surface residues", as reported by Bateman et al.17 "Key residues" are amino acid residues crucial for the structural build up of a domain. Typically, these residues are very conserved among species. A mutation in a key residue probably affects the entire domain if not the rest of the protein. A "surface residue" is generally less conserved. A mutation in a surface residue is expected to have little influence on the rest of the domain. In table 5, we have outlined the distribution of the families with an extracellular missense mutation. Mutation W9S was not taken into account since it occurs in the signal peptide. Eight of the 13 remaining families harboured a mutation affecting a key residue, whereas the five other families were affected by a mutation affecting a surface residue. We then investigated whether any difference in phenotype could be observed depending on whether the class II mutation affected a key or a surface residue using the Fisher exact test. In the first test (table 5A), phenotypes were subdivided into "severe" and "non-severe", and we found that mutations in key and surface residues did not have significantly different proportions of the two phenotypes (p=0.293). In a second Fisher exact test (table 5B), phenotypes were subdivided into "mild" and "non-mild". In this case, we found significantly different phenotype proportions between the patients with a mutation in a key residue and those with a mutation affecting a surface residue (p=0.035).

Seven families with a class IV extracellular splice site mutation were included in our study. The distribution of these families was not significantly different from the expected distribution.

**Discussion**

Given the wide array of functions and interactions attributed to L1, together with the various types of mutations and the broad clinical spectrum of malformations encountered in cases of L1 mutation, it is interesting to look for a genotype-phenotype correlation in L1 associated disease. We collected clinical data from 36 affected families with a known L1 mutation, and set criteria for the subdivision of the families according to severity of the clinical symptoms, and subdivision of the mutations according to their effect on the L1 protein.

Class I mutations, leading to extracellular truncation of L1, were found to cause a severe phenotype in all families. Most likely, class I mutations lead to a total loss of function of L1 as the truncated peptide is secreted, loses all contact with the cell membrane, and probably degrades quickly.

Class III cytoplasmic mutations may only cause partial loss of function of L1 leading to a significantly milder phenotype. Possibly, only the cytoplasmic functions of L1 are affected, whereas extracellular functions remain intact. Cell aggregation studies have shown that L1 (and its Drosophila homologue Neuroglian) can still mediate homophilic adhesion without a cytoplasmic domain.

Class II missense mutations give rise to all three clinical phenotypes, and the distribution of the three phenotypes caused by class II mutations was not significantly different from that of the other mutations. It is possible that
the impact of a missense mutation on L1 functioning is dependent on the exact location of the mutation in L1 and on the nature of the amino acid substitution. Bateman et al.12 have constructed outline structures of all the extracellular domains of L1 indicating the position of each amino acid residue. They subdivided the extracellular missense mutations into mutations affecting key amino acid residues and those affecting surface residues. The former affect highly conserved residues being crucial for the structural build up of the Ig or Fn domain. A mutation in these residues probably affects the core structure of the domain in which they occur or even other domains as well. The latter are mutations affecting amino acid residues situated at the outer part of the domain. Bateman et al.12 have suggested that these surface residues are important for finely tuned interactions with extracellular ligands, whereas they probably have less influence upon the overall structure of the domain. We therefore analysed whether the two groups of class II mutations caused different clinical phenotypes. As Table 5A shows, a mutation in a key residue never leads to a mild phenotype, in contrast with mutations in a surface residue. The severe phenotype was caused both by mutations in a key residue and by mutations in a surface residue (Table 5B). Probably, mutations in a crucial amino acid have a devastating effect on the protein leading to a severe clinical picture, whereas mutations on the surface of a domain have a variable effect on the whole protein.

Class IV mutations affecting splicing of L1 do not cause a predictable phenotype, and no genotype-phenotype correlation is found. The effects of a splice mutation on the intron splicing are hard to predict and difficult to study. In three of the seven class IV mutations, the effect of the mutation on intron splicing has yet to be determined. Two of the class IV mutations give rise to multiple splice forms, including the normal one (mutations 2 and 28),13 14 whereas on two other occasions (4 and 19)15 (our own unpublished results), only one aberrantly spliced form was detected. The presence of the normally spliced mRNA species, probably leading to normal, functional L1 protein in some patients, is remarkable. However, this might be explained by either: (1) a reduction in the concentration of normal L1 protein or (2) aberrantly spliced L1 inhibiting the normal L1 protein ("dominant negative effect").

Previous reports have stressed the large intrafamilial variability as one of the main difficulties in establishing genotype-phenotype correlation.3 11 Indeed, some of the most extensively described CRASH families contain severely as well as mildly affected patients.25 27 28 However, also, families with a much more uniform phenotype, either in the mild15 32 33 or the severe end13 34 of the spectrum, have been reported and a variable phenotype is much less common among the unpublished families. Actually, most of the CRASH patients and families fit our criteria of "severely" affected. One possible explanation for this apparent paradox might be that there is a bias in published reports towards the variable and mildly affected families. In severely affected families, most, if not all, patients die at birth or in early childhood. The absence of living patients makes these families less suitable for writing up a detailed case report.

L1 associated diseases have always been described as showing considerable inter- and intrafamilial variability. The results presented here indicate intrafamilial variability is correlated with the dissimilar effects of the different types of mutations on the L1 protein. Mutations leading to a total loss of function of L1 lead to a significantly more severe phenotype than mutations causing a partial loss of function. However, since some families still show intrafamilial variability, secondary factors must be involved in the clinical picture of CRASH syndrome.

We are indebted to the following clinicians for referral of patients and sending of clinical information: M Sagi, C Schrander-Stumpel, S Tinisicht, B Minulla, F Douglas, D Chi-toya, J-P Fryns, D Quarell, M Bonduelle, Y Gillet, J Holiden, P Deier, R Rührborn, R Szibor, and E Wesby. This study was supported by a concerted action from the University of Antwerp (PFW and GVC, and by a grant from the French AFM (Association Française contre les Myopathies) to PFW.

References


Genotype-phenotype correlation in L1 associated diseases.

E Fransen, G Van Camp, R D'Hooge, L Vits and P J Willems

*J Med Genet* 1998 35: 399-404
doi: 10.1136/jmg.35.5.399

Updated information and services can be found at:
http://jmg.bmj.com/content/35/5/399

**Email alerting service**

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Notes**

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/