Mutations in a Gene Encoding a Novel Protein Containing a Phosphotyrosine-Binding Domain Cause Type 2 Cerebral Cavernous Malformations

Christina L. Liquori,1 Michel J. Berg,2 Adrian M. Siegel,3 Elizabeth Huang,1 Jon S. Zawistowski,1 T’Prien Stoffer,4 Dominique Verlaan,5 Fiyinfolu Balogun,1 Lori Hughes,1 Tracey P. Leedom,1 Nicholas W. Plummer,1 Milena Cannella,6 Vittorio Maglione,6 Ferdinando Squitieri,6 Eric W. Johnson,4 Guy A. Rouleau,5 Louis Ptacek,7 and Douglas A. Marchuk1

1Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC; 2Strong Epilepsy Center, Department of Neurology, University of Rochester Medical Center, Rochester, NY; 3Department of Neurology, University Hospital Zurich, Zurich; 4Barrow Neurological Institute, Neurogenetics, Phoenix; 5Montreal General Hospital, Department of Neurology, Montreal; 6Neurogenetics Unit, Istituto di Ricovero e Cura a Carattere Scientifico, Neuromed, Pozzilli, Italy; and 7University of California, San Francisco, Department of Neurology, San Francisco

Cerebral cavernous malformations (CCMs) are congenital vascular anomalies of the central nervous system that can result in hemorrhagic stroke, seizures, recurrent headaches, and focal neurologic deficits. Mutations in the gene KRIT1 are responsible for type 1 CCM (CCM1). We report that a novel gene, MGC4607, exhibits eight different mutations in nine families with type 2 CCM (CCM2). MGC4607, similar to the KRIT1 binding partner ICAP1α, encodes a protein with a phosphotyrosine-binding domain. This protein may be part of the complex pathway of integrin signaling that, when perturbed, causes abnormal vascular morphogenesis in the brain, leading to CCM formation.
Table 1

Characteristics of Nine Families with CCM2 Mutations

<table>
<thead>
<tr>
<th>Family</th>
<th>No. of Affected Individuals</th>
<th>Exon</th>
<th>Nucleotide Change</th>
<th>Mutation Consequence</th>
<th>Predicted Amino Acid Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>2626</td>
<td>4</td>
<td>1</td>
<td>23delG</td>
<td>Frameshift</td>
<td>86fsX22</td>
</tr>
<tr>
<td>214</td>
<td>1</td>
<td>1</td>
<td>30G→A</td>
<td>Splicing alteration</td>
<td>K10K</td>
</tr>
<tr>
<td>CAV01b</td>
<td>4</td>
<td>2</td>
<td>56delG</td>
<td>Frameshift</td>
<td>19fsX22</td>
</tr>
<tr>
<td>70c</td>
<td>5</td>
<td>2</td>
<td>169_172delAGAC</td>
<td>Frameshift</td>
<td>57fsX58</td>
</tr>
<tr>
<td>IFCAS-1d</td>
<td>4</td>
<td>2</td>
<td>169_172delAGAC</td>
<td>Frameshift</td>
<td>57fsX58</td>
</tr>
<tr>
<td>229</td>
<td>4</td>
<td>4</td>
<td>314delT</td>
<td>Frameshift</td>
<td>105fsX105</td>
</tr>
<tr>
<td>IFCAS-14</td>
<td>2</td>
<td>4</td>
<td>319C→T</td>
<td>Nonsense</td>
<td>Q107X</td>
</tr>
<tr>
<td>2030</td>
<td>5</td>
<td>6</td>
<td>610-1G→A</td>
<td>Splicing alteration</td>
<td>NA</td>
</tr>
<tr>
<td>2812</td>
<td>1</td>
<td>6</td>
<td>653delG</td>
<td>Frameshift</td>
<td>218fsX291</td>
</tr>
</tbody>
</table>

a MRI positive and/or symptom positive.
b Data from Squitieri et al. 2000.
c Data from Steichen-Gersdorf et al. 1992.
d Data from Dupre et al. 2003.

KRIT1 mutations, and 10 harbored a KRIT1 mutation. The 27 non-KRIT1 (CCM1) probands were included in this study, and, where possible, additional family members were collected. The boundaries of the CCM2 locus were previously established by 1-LOD support analysis as an 11-cM interval flanked by markers D7S2846 and D7S1818 (Craig et al. 1998). Markers mapping within the interval were used to haplotype several families with CCM2, but no crossovers were identified in affected individuals. Since we were unable to genetically reduce the size of the CCM2 interval, we began to sequence positional candidate genes.

Within the genetically defined interval, 55 known or putative genes have been identified (UCSC Human Genome Assembly Web site). We initially selected eight genes for sequence analysis on the basis of biological

Figure 1  Sequence traces of the eight CCM2 mutations. Each mutation is indicated by a black arrow. The extent of a 4-bp deletion is shown with conjoined arrows.
plausibility: MGC4607, RALA (v-ral simian leukemia viral oncogene), CAMK2B (calcium/calmodulin-dependent protein kinase II β), STK17A (serine/threonine kinase 17a), CDC10 (cell division cycle 10 homolog), CDC2L5 (cell division cycle 2-like 5), HIP-55 (src homolog 3 domain-containing protein), and MYLC2A (myosin light chain 2a). The predicted gene MGC4607 was chosen because its translated protein product encodes a putative PTB domain. This same domain is found in ICAP1α, a binding partner of the CCM1 product KRIT1.

Among the panel of 27 probands without a KRIT1 mutation, we detected eight different mutations in MGC4607. One mutation (169_172delAGAC) was found in two separate families that—as a further investigation into ethnic background suggested—may be distantly related. The mutations include five frameshift, one nonsense, and two splicing mutations (table 1; fig. 1), each of which, if translated, would result in a truncated protein. The mutations map throughout the gene, including two in exon 1, suggesting that at least some may lead to loss-of-function alleles.

One splice-junction mutation (610-1G→A, in family 2030) is located at the invariant G residue at the splice-acceptor site adjacent to exon 6. A second splice-site mutation was less obvious. Family 214 harbored what appeared to be a silent mutation (30G→A; K10K) in the last nucleotide of exon 1. Nonetheless, we thought that this sequence alteration might also lead to a splicing defect, because it results in a less favorable nucleotide at this position of the donor splice junction and this position is not infrequently a site of mutations affecting splicing. Sequence analysis confirmed that the 30A variant of MGC4607 was not present in 606 control chromosomes. To examine its effect on the MGC4607 mRNA, we performed expression analysis on leukocyte mRNA isolated from the proband. RT-PCR products from both the proband and control leukocyte mRNA exhibited a number of normal splice variants of the message but no additional splice variants arising from the patient mutation. However, sequence analysis of the RT-PCR products from the proband revealed only the normal allele. This suggests that the mutant 30A allele is not expressed or, more likely, that this mutation results in a grossly aberrant splice variant that is subjected to degradation.

Four of the families harboring MGC4607 mutations were previously reported as consistent with showing linkage to CCM2 (Craig et al. 1998; Squitieri et al. 2000; Dupre et al. 2003). The remaining five families with

---

**Table 2**

Comparison of MGC4607 Mutation Status with MRI Results

<table>
<thead>
<tr>
<th>FAMILY</th>
<th>NO. OF FAMILY MEMBERS WITH A MGC4607 MUTATION, WITH MRI STATUS</th>
<th>NO. OF FAMILY MEMBERS WITHOUT A MGC4607 MUTATION, WITH MRI STATUS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MGC4607 Mutation, with MRI Status</td>
<td>MGC4607 Mutation, with MRI Status</td>
</tr>
<tr>
<td>2626</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>214</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>CAV01</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>70a</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>IFCAS-1d</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>229</td>
<td>4</td>
<td>1*</td>
</tr>
<tr>
<td>IFCAS-14</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2030</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>2812</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

* One individual with stroke and seizures.
* Data from Squitieri et al. 2000.
* Data from Steichen-Gersdorf et al. 1992.
* Data from Dupre et al. 2003.
* One individual with stroke at age 20 years.
CCM2 were of insufficient power to establish linkage. Families not showing CCM2 mutations harbored either mutations in the yet undiscovered CCM3 gene or mutations undetectable by DNA sequence analysis of coding exons at the CCM1 and CCM2 loci.

The mutation in each of the nine families cosegregates with affected status for all confirmed (MRI- and/or symptom-positive) affected individuals (table 2). Except for one individual from family 2626, all symptomatic individuals had MRI data available that confirmed the presence of the CCM lesions. One mutation carrier in family 229 was apparently MRI negative at age 75 years but had suffered a stroke of unknown etiology at age 20 years and thus was classified as “presumed affected.” The families also contained a rather large number of mutation carriers that were, at the time of ascertainment, clinically silent (table 2). For nearly all of these, the lack of overt symptoms led to a presumed unaffected status, and, as a consequence, these individuals were not evaluated by MRI.

The large number of mutation-positive individuals that remain asymptomatic provides an explanation for the difficulty encountered in narrowing the candidate interval for this and other CCM loci. Asymptomatic individuals will receive a medical diagnosis of unaffected, but, without MRI examination, their status in terms of genetic linkage analysis remains uncertain. The paucity of confirmed affected individuals in most families precludes definitive linkage to a CCM locus and can even mask the autosomal dominant inheritance. Even when a family shows unequivocal linkage to the correct locus, there are generally fewer confirmed affected (MRI-positive) family members and thus fewer informative meioses that may harbor recombination events useful for refined mapping.

These data confirm the previously noted reduced penetrance within families with CCM with regard to expression of clinical symptoms. However, penetrance appears much higher in patients with MRI-diagnosed affected status. A more accurate determination of penetrance will require MRI examination of those individuals who now test positive for the mutation in their family. These data emphasize both the importance of MRI in the diagnosis of CCM and the potential for DNA-based diagnostics to identify those at risk.

To determine the expression pattern of MGC4607, we performed northern blot analysis of human tissues, using the entire cDNA as a probe (fig. 2). MGC4607 is most highly expressed in the skeletal muscle, heart, and liver, with minimal or no expression in the colon and lung. MGC4607 is also expressed in the brain. KRI1 and its binding partner, ICAP1α, have been also shown to be expressed in the brain (Faisst and Gruss 1998; Denier et al. 2002; Kehrer-Sawatzki et al. 2002). These expression data suggest that CCM lesion formation may not be due to an intrinsic vascular defect but, rather, that the cerebral vasculature is aberrantly responding to signals arising in the brain parenchyma. This is consistent with the histopathology of the CCM lesions, which are devoid of neuronal parenchyma within the lesions.

**Figure 3**  Schematic diagram of malcavernin. Each exon is shown as a box with the corresponding exon number. The PTB domain is shown in gray. The location and description of the eight CCM2 mutations are shown. Mutations affecting splice junctions are shown above, with the introns indicated by lines.
themselves. There are, however, other data that support vascular expression of the CCM1 gene product KRIT1 (Gunel et al. 2002).

We have named this novel protein “malcavernin” for its role in the development of cerebral cavernous malformations (fig. 3). The identification of ICAP1α as a KRIT1 binding partner (Zhang et al. 2001; Zawistowski et al. 2002) suggested a model in which KRIT1 and β1-integrin compete for ICAP1α binding, possibly regulating integrin signaling (Marchuk et al. 2003). The presence of a PTB domain in malcavernin predicts a possible interaction with the CCM1 protein, KRIT1, and/or with the β1-integrin cytoplasmic tail. Integrin signaling is critical for cell-cell and cell-extracellular matrix communication, which, in turn, modulates cellular migration and morphology. A biochemical pathway is emerging for CCM pathogenesis, involving regulation of integrin signaling through a complex set of interacting proteins that may compete for binding. The proteins involved in this pathway include the CCM1 and CCM2 gene products KRIT1 and malcavernin.

Acknowledgments

We are grateful to the patients and their families for participation in this study. This work was supported by National Institutes of Health grant NS43543 and American Heart Association grant (Bugher Foundation Award) 0070028N (both to D.A.M.).

Electronic-Database Information

The URLs for data presented herein are as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for CCM1, CCM2, and CCM3)
University of California, Santa Cruz (UCSC), Human Genome Assembly Web site, http://genome.ucsc.edu/ (for published assembly and genome browser)

References

Squitieri F, Maglione V, Buzzi MG, Nargi E, Novelletto A, Cannella M, Simonelli M, Colonnese C, Simonelli P, Inno-