Am. J. Hum. Genet. 74:326–337, 2004

Mutations within the *MGC4607* Gene Cause Cerebral Cavernous Malformations

C. Denier,[1,2] S. Goutagny,[1] P. Labauge,[1,4] V. Krivosic,[1] M Arnoult,[1] A. Cousin,[1] A. L. Benabid,[5] J. Comoy,[6] P. Frerebeau,[7] B. Gilbert,[8] J. P. Houtteville,[10] M. Jan,[11] F. Lapierre,[12] H. Loiseau,[13] P. Menei,[14] P. Mercier,[14] J. J. Moreau,[9] A. Nivelon-Chevallier,[15] F. Parker,[6] A. M. Redondo,[15] J. M. Scarabin,[17] M. Tremoulet,[13] M. Zerah,[3] J. Maciazek,[1] E. Tournier-Lasserve,[1,2] and Société Française de Neurochirurgie

[1]INSERM E365, Faculté de Médecine Lariboisière,[2]Laboratoire de Cytogénétique, Hôpital Lariboisière, Assistance Publique-Hôpitaux de Paris, and [3]Service de Neurochirurgie, CHU de Necker-Enfants Malades, Paris; [4]Service de Neurologie, CHU Montpellier Nîmes, Nîmes, France; [5]Service de Neurochirurgie, CHU de Grenoble, Grenoble, France; [6]Service de Neurochirurgie, CHU de Kremlin-Bicêtre, Kremlin-Bicêtre, France; [7]Service de Neurochirurgie CHU de Montpellier, Montpellier, France; [8]Génétique, CHU de Limoges, and [9]Service de Neurochirurgie, CHU de Limoges, Limoges, France; [10]Service de Neurochirurgie, CHU de Caen, Caen, France; [11]Service de Neurochirurgie, CHU de Tours, Tours; [12]Service de Neurochirurgie, CHU de Poitiers, Poitiers, France; [13]Service de Neurochirurgie, CHU de Toulouse, Toulouse; [14]Service de Neurochirurgie, CHU de Angers, Angers, France; [15]Génétique, CHU de Dijon, Dijon; [16]Service de Neurochirurgie, Hôpital Beaujon, Clichy, France; and [17]Service de Neurochirurgie, CHU de Rennes, Rennes, France

Cerebral cavernous malformations (CCM) are hamartomatous vascular malformations characterized by abnormally enlarged capillary cavities without intervening brain parenchyma. They cause seizures and focal neurological deficits due to cerebral hemorrhages. *CCM* loci have already been assigned to chromosomes 7q (*CCM1*), 7p (*CCM2*), and 3q (*CCM3*) and have been identified in 40%, 20%, and 40%, respectively, of families with CCM. Loss-of-function mutations have been identified in *CCM1/KRIT1*, the sole CCM gene identified to date. We report here the identification of *MGC4607* as the *CCM2* gene. We first reduced the size of the *CCM2* interval from 22 cM to 7.5 cM by genetic linkage analysis. We then hypothesized that large deletions might be involved in the disorder, as already reported in other hamartomatous conditions, such as tuberous sclerosis or neurofibromatosis. We performed a high-density microsatellite genotyping of this 7.5-cM interval to search for putative null alleles in 30 unrelated families, and we identified, in 2 unrelated families, null alleles that were the result of deletions within a 350-kb interval flanked by markers D7S478 and D7S621. Additional microsatellite and single-nucleotide polymorphism genotyping showed that these two distinct deletions overlapped and that both of the two deleted the first exon of *MGC4607*, a known gene of unknown function. In both families, one of the two *MGC4607* transcripts was not detected. We then identified eight additional point mutations within *MGC4607* in eight of the remaining families. One of them led to the alteration of the initiation codon and five of them to a premature termination codon, including one nonsense, one frameshift, and three splice-site mutations. All these mutations cosegregated with the disease in the families and were not observed in 192 control chromosomes. *MGC4607* is so far unrelated to any known gene family. Its implication in CCMs strongly suggests that it is a new player in vascular morphogenesis.

Introduction

Cerebral cavernous malformations (CCM [MIM 116860]) are vascular malformations, mostly located within the CNS. They are characterized by abnormally enlarged capillary cavities without intervening brain parenchyma (Russel and Rubinstein 1989). The most common symptoms are seizures and neurological deficits due to focal hemorrhages (Rigamonti et al. 1988; Robinson et al. 1991; Labauge et al. 1998). The prevalence of the condition has been estimated to be ~0.1%–0.5%, on the basis of cerebral magnetic resonance imaging (MRI) and autopsy studies of large cohorts of patients (Otten et al. 1989). Cavernous angiomas can occur in a sporadic or autosomal dominantly inherited form. The proportion of the familial form has been estimated to be as high as 50% in Hispanic American patients with CCM (Rigamonti et al. 1988) and ~10%–20% in white patients (E. Tournier-Lasserve, unpublished data). Familial cases are characterized by the presence of multiple lesions, whereas sporadic cases usually harbor only one CCM lesion (Rigamonti et al. 1988; Labauge et al. 1998).

Three CCM loci have been mapped—on 7q (*CCM1*), 7p (*CCM2*), and 3q (*CCM3*)—with 40% of kindreds with CCM showing linkage to *CCM1*, 20% showing linkage to *CCM2*, and 40% showing linkage to *CCM3* (Dubovsky et al. 1995; Craig et al. 1998; Dupre et al.

Received October 17, 2003; accepted for publication November 25, 2003; electronically published January 22, 2004.

Address for correspondence and reprints: Dr. E. Tournier-Lasserve, INSERM E365, Faculté de médecine Lariboisière, 10, avenue de Verdun, 75010 Paris, France. E-mail: tournier-lasserve@paris7.jussieu.fr

© 2004 by The American Society of Human Genetics. All rights reserved.
0002-9297/2004/7402-0014$15.00

2003). To date, only one CCM gene, *CCM1/KRIT1,* has been identified (Laberge-le Couteulx et al. 1999; Sahoo et al. 1999). Nearly 100 different *KRIT1* germline mutations have been detected in families with *CCM*. All these mutations lead to premature stop codons through either nonsense, splice-site, frameshift, or false missense mutations, strongly suggesting that Krit1 loss of function is the most likely mechanism in patients with CCM (Laberge-le Couteulx et al. 1999; Sahoo et al. 1999; Eerola et al. 2000; Zhang et al. 2000; Davenport et al. 2001; Lucas et al. 2001; Sahoo et al. 2001; Cave-Riant et al. 2002; Chen et al. 2002; Couteulx et al. 2002; Verlaan et al. 2002; Laurans et al. 2003; Lucas et al. 2003; Marini et al. 2003; Musunuru et al. 2003; Xu et al. 2003).

In the present study, we identified *MGC4607* as the *CCM2* gene. In a first step, a genetic linkage analysis conducted on five families without linkage to *CCM1* excluded linkage to *CCM3* and showed consistent linkage to *CCM2,* allowing us to refine the 22-cM *CCM2* interval, published elsewhere (Craig et al. 1998), to a 7.5-cM interval. In a second step, we performed a high-density microsatellite genotyping of this 7.5-cM interval in the 5 initial families and in 25 additional families with CCM, to search for putative null alleles, on the basis of the hypothesis that large deletions may be involved in this condition, as reported elsewhere for other loci in hamartomatous conditions such as tuberous sclerosis or neurofibromatosis types 1 and 2 (Viskochil et al. 1990; The European Chromosome 16 Tuberous Sclerosis Consortium 1993; Trofatter et al. 1993). This search led us to identify two distinct overlapping deletions in two unrelated families. These two deletions were both shown to delete exon 1 of *MGC4607,* a known gene of unknown function. In addition to these two deletions, molecular analysis of the 10 exons and intron-exon boundaries of *MGC4607* revealed eight different small germline mutations cosegregating in eight families of this panel. Five of these mutations led to a premature stop codon, and one of them altered the ATG initiator codon, strongly suggesting that *CCM2* lesions are the consequence of *MGC4607* haploinsufficiency. These data establish *MGC4607,* a gene unrelated to any known gene family, as a new player in vessel development and/or maturation.

Patients, Material, and Methods

Patients and Families

Thirty unrelated families were enrolled in this study, on the basis of the following three criteria: (1) each proband had at least one affected relative and/or had multiple cerebral cavernous angiomas, (2) the families of the probands were potentially informative for a study designed to identify large genomic deletions on the basis of the identification of microsatellite null alleles, and (3) results of molecular screening for a *KRIT1* mutation were negative. CCM diagnosis was based on cerebral MRI features and/or pathological analysis. We considered as "affected" all individuals showing cavernous angiomas on cerebral MRI, regardless of their clinical status. We considered as "healthy" all individuals with normal MRI results. Those who did not undergo MRI were classified as "unknown" (fig. 1).

This panel of 30 families included (1) 5 large families potentially informative for linkage analysis, on the basis of the number of their available meioses (families C004, C020, C039, C049, and C116) and (2) 25 families that included a smaller number of sampled individuals but which were potentially informative for the search of putative null alleles. In the five large families, previous haplotype analysis with *KRIT1* peri- and intragenic markers had ruled out *KRIT1/CCM1* linkage (Laberge et al. 1999 and data not shown).

DNA and RNA Extraction

Genomic DNA from probands and consenting relatives was extracted from peripheral blood by use of standard procedures. Genomic DNA from 96 unrelated healthy white French individuals was available as a control group. Total RNA was extracted from lymphoblastoid Epstein-Barr virus (EBV) cell lines for 26 of the 30 probands (lymphoblastoid cell lines were not available for probands C077, C096, C114, and C146), and cDNA was prepared according to standard procedures (Laberge-le Couteulx et al. 1999).

Genetic Markers

For the initial genetic linkage analysis, we used a panel of 32 microsatellite markers, including 15 markers spanning the *CCM2* interval published elsewhere (22-cM interval flanked by D7S516 and D7S1818) and 17 microsatellites spanning the *CCM3* interval (22-cM interval flanked by D3S3053 and D3S1262) (Craig et al. 1998). The distances between those markers ranged between 1 and 4 cM. The *CCM2* markers were D7S2564, D7S516, D7S2496, D7S2252, D7S2250, D7S2251, D7S2497, D7S510, D7S485, D7S521, D7S2548, D7S691, D7S478, D7S1818, and D7S2422 (in pter→cen orientation). The *CCM3* markers were D3S3530, D3S1580, D3S3686, D3S1262, D3S3592, D3S3609, D3S3730, D3S3715, D3S3676, D3S3520, D3S3725, D3S1574, D3S1282, D3S1614, D3S3622, D3S1264, and D3S3673 (in pter→cen orientation) (Centre d'Etude du Polymorphisme Humain; Cooperative Human Linkage Center; Entrez Genome).

Once the *CCM2* region was refined to a 7.5-cM interval, an additional series of 12 microsatellites spanning this reduced interval (average spacing ~600 kb) were used for high-density microsatellite genotyping. This

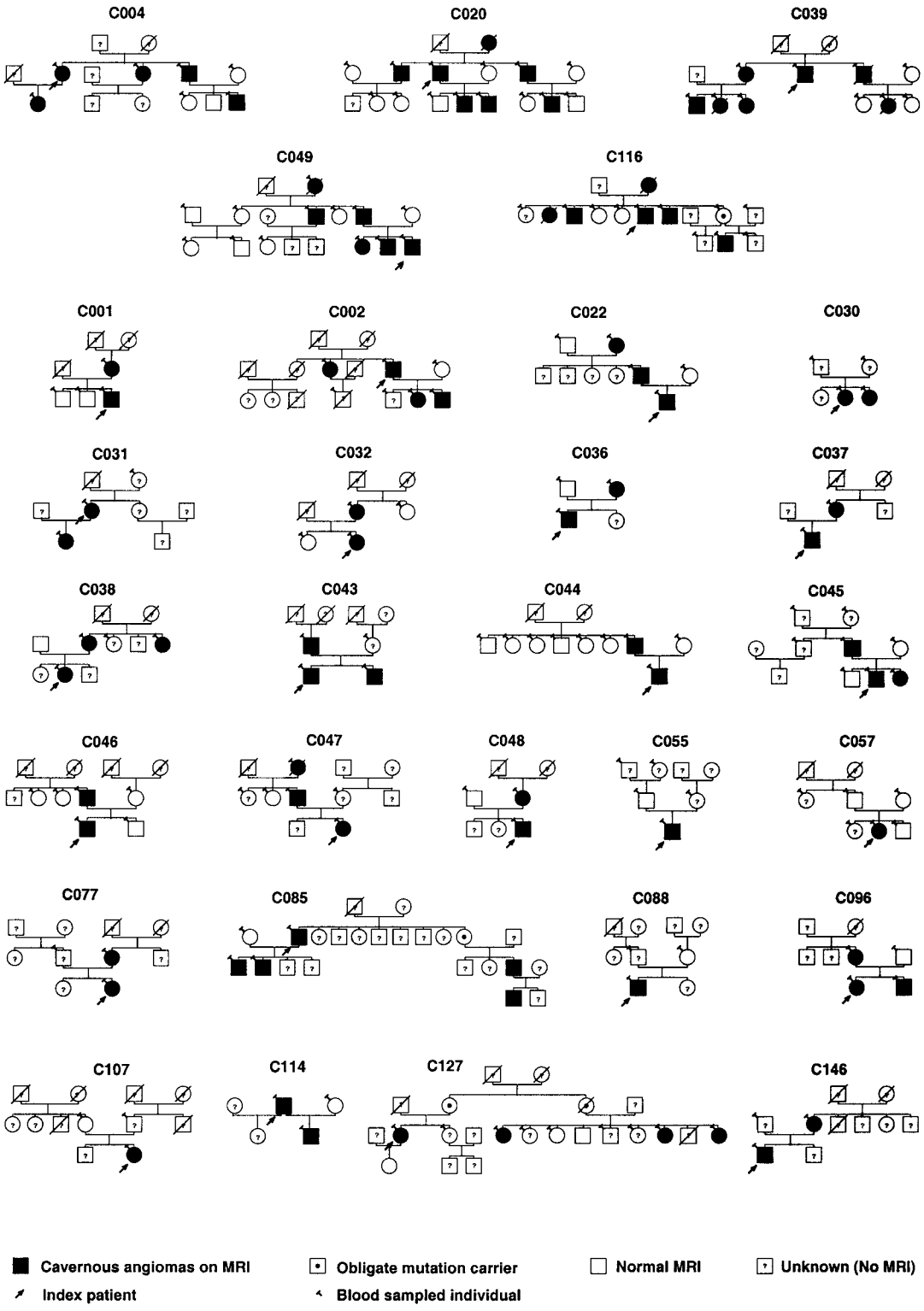

Figure 1 Genealogical trees of the 30 analyzed families with CCM. The family numbers are indicated above each pedigree. The five families used for genetic linkage analysis were families C004, C020, C039, C049, and C116.

Table 1

PCR Primers Used for Mutation Detection in the *MGC4607* Gene

Exon(s)	Oligonucleotide Sequence		Product Size (bp)
	Forward	Reverse	
1[a]	TCGAGCATGTAGCGGCTG	ACCACAGCAGTAGGACGC	139
2[a]	CCATAGGTACAACACAAAGC	TTTGGCAGGCACGTCCTG	274
3[a]	TAGTGATGGTGGTGTTGGC	AAGAGTACAGGGACCAAGG	216
4A[a]	GCCAGCACAATATTCTCAG	CCTGAGGATGATATCCTCC	194
4B[a]	TCAGCCTGTCTGCGTACAA	TCACACATGTACTCCTTAGG	188
5[a]	CCTTCCACTGTGCTAAACT	GGAGCCAGTTCTGAGTAAG	215
6[a]	TCCAGCCAGACTGACCGA	ACTGCAGTAGCCCCTCCT	235
7[a]	ACTCAAAATGCCTCCCCAC	CACCAGACAACGCATTTGG	180
8[a]	ACGTGTGTGGGATGGAGG	CCCCCTCCCTCATGCAG	215
9A[a]	ACTCTTGCCTACTGTGCCCA	TAGAGCTGCCGCAGGTTGAT	159
9B[a]	GCTGTCATCACAGGAGATCC	ACATTCCCAGGAGGGATCAG	199
10A[a]	TGAGAGAAGAGCTGAGTGG	AACTGGATGTGGTGCTCAG	203
10B[a]	TCATGACAGCTTTGGCAG	ATGACGACTGCGCGGAAG	240
1–10[b]	GCGGCGATATGGAAGAGG	CATCCACTGTCCATCATGC	1,356
1–6[b]	GCGGCGATATGGAAGAGG	ATCGCTGTGCAGGGACAG	755
2–8[b]	CCAGACAGACTGCTGAGC	ATGGTCTTGCTGTGGGG	750
6–10[b]	TCCACCATCGACTTTCTGG	CATCCACTGTCCATCATGC	679

[a] PCR primer used for genomic exon-by-exon amplification and mutation detection by SSCP and/or direct sequencing.

[b] RT-PCR primer used for cDNA studies.

panel included markers D7S1526, D7S678, D7S667, D7S478, D7S2427, D7S621, D7S2248, D7S519, D7S1800, D7S670, D7S665, and D7S2506 (fig. 3*A*).

Last, screening for di-, tri-, and tetranucleotide tracks of ad hoc genomic sequences spanning the D7S478–D7S621 interval (BACs AC004854, AC004847, AC013416, and AC004844) (fig. 3*A* and 3*B*) identified two new CA repeat markers designated MS004844 (BAC AC004844/nts 125763–125790) and MS004847 (BAC AC004847/nts 45874–45907).

In addition, 42 SNPs located between D7S478 and MS004844 were retrieved from the National Center for Biotechnology Information (NCBI [Entrez Genome]) and The Chromosome 7 Annotation Project databases; the SNP designations begin with "rs" (Single Nucleotide Polymorphism Web site).

Genotyping

Microsatellite genotyping and genetic linkage analysis were performed as described elsewhere (Laberge et al. 1999). Pairwise LOD scores were calculated with the MLINK program of the LINKAGE 5.1 package through use of the following parameters: autosomal dominant inheritance, 95% penetrance, disease gene frequency 1/1,000, and phenocopy rate 1/100. Haplotypes were constructed using the most parsimonious linkage phase. SNPs were screened by direct sequencing of potentially informative individuals within families.

MGC4607 *Mutation Screening*

In a first step, genomic DNA of probands of all families was screened using SSCP, except for probands from families C116 and C127. *MGC4607* (GenBank accession number NM_031443) contains 10 coding exons. We designed 13 sets of oligonucleotide primers from chromosome 7 genomic contig NT_007819 (GenBank GI number 29796755) to amplify the *MGC4607* exons and flanking splice sites (primer sequences shown in table 1). PCR and SSCP were conducted as described elsewhere (Cave-Riant et al. 2002). We tested all confirmed mutations for cosegregation with the affected phenotype in each family. Ninety-six unrelated spouses were systematically used as negative controls.

In a second step, all probands in whom we did not detect a *MGC4607* mutation were sequenced to search for mutations through use of cDNA, when available, or of genomic DNA. cDNA was amplified using two sets of primers amplifying fragments from exons 2–8 and 6–10 (table 1). Amplicons were checked on agarose gels and were sequenced on an ABI 3100 (PE Applied Biosystems). Because of the existence of an alternatively spliced transcript deleted from exon 2 in EBV cell lines, exons 1 and 2 were amplified and sequenced from genomic DNA from all probands.

In addition to this genomic and cDNA screening, the putative aberrant splicing effects of all mutations were systematically tested by sequencing RT-PCR products obtained with primers located $\geq$2 exons upstream and downstream of the mutation location.

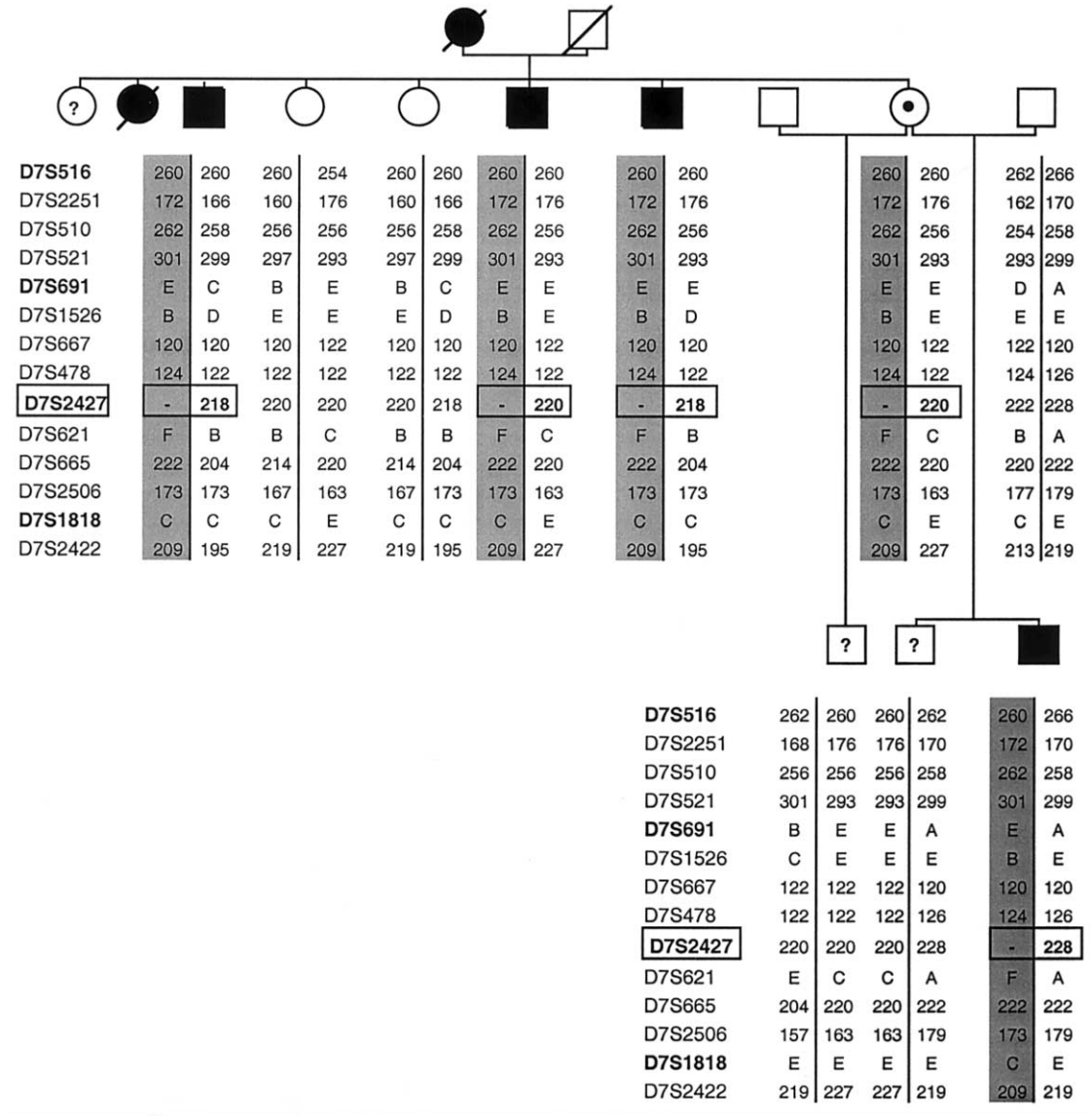

Figure 2 D7S2427 parental noncontribution within family C116. The family C116 genealogical tree and *CCM2* haplotypes are shown; markers D7S516 and D7S1818 flanking the published 22-cM *CCM2* interval are indicated, as well as D7S691, the new telomeric boundary. A null allele was detected at the D7S2427 marker (*boxed*) within the haplotype cosegregating with CCM (*gray*), strongly suggesting the existence of a deletion that was confirmed by a distinct D7S2427-specific primer set and by additional genotyping (fig. 3).

Multiple-Tissue cDNA Panels and RT-PCR

Two sets of primers were designed to amplify *MGC4607* cDNA in two overlapping fragments spanning exons 1–6 (755 bp) and exons 6–10 (679 bp). Human normalized multiple-tissue cDNA (MTC I [Clontech]) panel and cDNA from control EBV lymphoblasts were amplified with these two *MGC4607* primer pairs and with primers specific for the β-actin gene (forward, 5′-CCAGATCATGTTTGAGACCT-3′; and reverse, 5′-ACGTCACACTTCATGATGGA-3′).

Multiple-Tissue Northern Blot

A 1,332-bp cDNA probe of the full length *MGC4607* ORF was generated from a lymphoblastoid EBV cell line derived from a normal healthy control individual (fragments 1–10) (table 1) and cloned in pGEM-T Easy Vector System (Promega). Probe sequence was identical to the reported human sequence (GenBank accession number NM_031443). One hundred nanograms of the full-length cDNA fragment was radiolabeled with [^{32}P]-dCTP and was used to hybridize in ExpressHyb solution

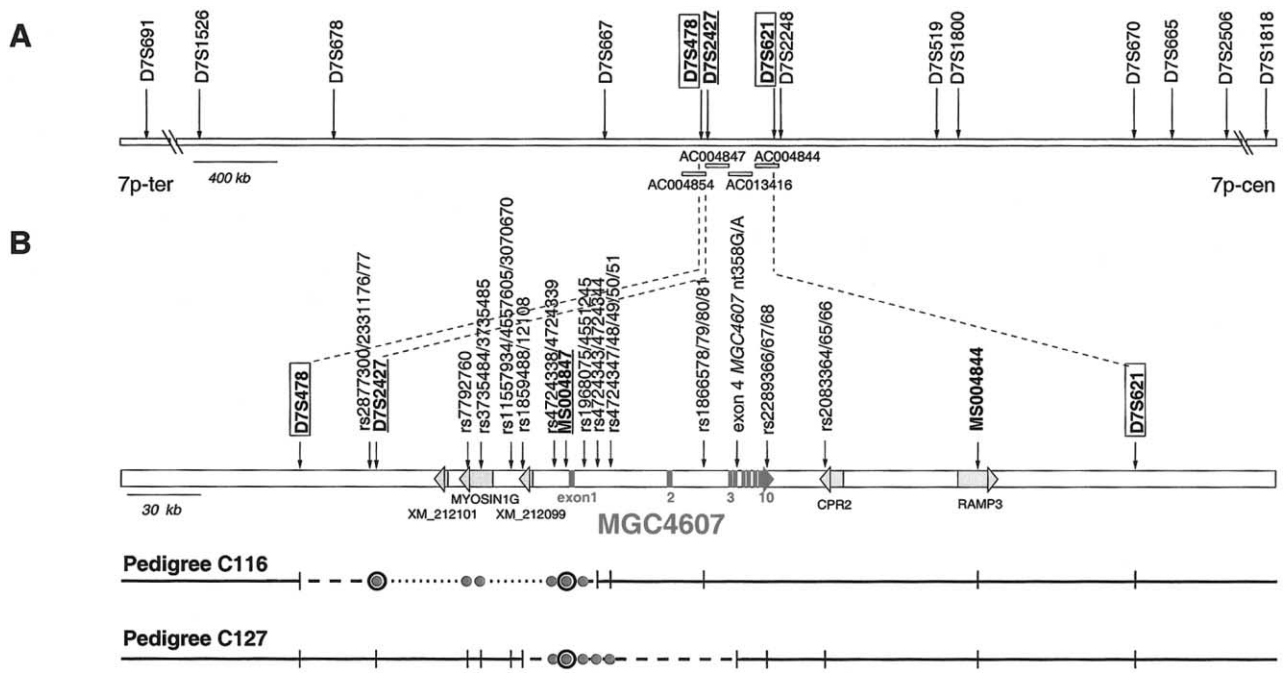

Figure 3 Large deletions within the *CCM2* interval. *A*, D7S2427 deletion in family C116. The 12 microsatellites located between D7S691 and D7S1818 are shown. Marker D7S2427, showing a null allele in family C116, is underlined. Markers D7S478 and D7S621 are boxed. Affected individual genotypes in family C116 were heterozygous for those two markers. Overlapping BACs containing the region of interest are represented with schematic bars. *B*, *CCM2* deletion refinement. MS004847 and MS004844 are indicated, as are the 42 SNP ("rs") markers retrieved from annotations of genomic sequence contigs. Dots denote null alleles detected with SNP markers; circled dots denote null alleles detected with markers D7S2427 and MS004847; vertical bars denote heterozygosities at given markers. Analysis of all 30 families with markers MS004844 and MS004847 (*underlined*) identified an additional null allele in family C127. Family C116 and C127 deletions were overlapping in a 12–29-kb region; the region of overlap contains the first exon of *MGC4607*.

(Clontech) at 1–2 × 10^6 cpm/ml, a human adult multiple-tissue northern blot (MTN I [Clontech]), according to manufacturer instructions. The same blot was subsequently stripped and hybridized with a control human β-actin probe (Clontech) to ensure that the RNA signal intensity was the same in each lane. Kodak film was exposed with the blot at −80°C for 96 h for *MGC4607* and for 6 h for the β-actin probe.

Results

CCM2 *and* CCM3 *Linkage Analysis*

The five larger families, C004, C020, C039, C049, and C116, were consistent with linkage to the *CCM2* locus. No recombinant was observed with several markers located in the centromeric part of the *CCM2* interval published elsewhere (Craig et al. 1998), and the combined maximum two-point LOD score at D7S478 was 5.43 at $\theta = 0.00$. It is interesting that *CCM2* haplotype analysis led us to identify two affected individuals within pedigrees C020 and C039 who were recombinant from D7S516 to D7S691. These recombinant events allowed us to reduce the size of the interval to 7.5 cM, the interval now bracketed by D7S691 and D7S1818, the *CCM2* centromeric boundary identified elsewhere (Craig et al. 1998).

CCM3 marker genotyping excluded linkage to this locus in all five families. In each of the five families, there was at least one affected individual who was recombinant on the whole *CCM3* interval. Multipoint LOD score values obtained with a panel of five markers covering the *CCM3* interval (D3S1614, D3S3676, D3S3730, D3S3592, and D3S3686) were ≤ -2 all along the interval.

D7S2427 Deletion in Family C116

We then hypothesized that large deletions might be involved in CCM, as already reported for other hamartomatous conditions, such as tuberous sclerosis or neurofibromatosis. An additional panel of 12 microsatellite markers located within the D7S691–D7S1818 interval was used to study segregation patterns within all 30 families, to identify putative null alleles (fig. 2; fig. 3*A*). The first 11 markers did not detect any abnormality, whereas apparent non-Mendelian inheritance was ob-

Table 2

***MGC4607* Mutations**

Pedigree	Location	Genomic DNA[a]	Mutation Effect on cDNA
C116	...	Deletion involving at least exon 1	Detection of only one transcribed allele
C127	...	Deletion involving at least exon 1	Detection of only one transcribed allele
C038	Exon 1	nt 1A→G	Mutation of the initiation codon
C039	Intron 3	nt 288+1G→A	Exon 3 deletion
C037	Exon 4	IVS3_exon 4 junction: delAGGTACAG insCTCAGGTGGCCCT	Exons 3 and 4 deletion, frameshift, and PTC
C096	Exon 5	nt 495_496insC	Direct frame-shift and PTC
C002	Exon 5	nt 593T→G	Missense mutation at codon 198
C004	Exon 5	nt 609G→A	Exon 5 deletion, frameshift, and PTC
C020	Exon 6	nt 652C→T	Nonsense mutation at codon 218
C047	Exon 10	nt 1248_1249delAG	Direct frame-shift and PTC

[a] Numbering of *MGC4607* nucleotides is according to the full-length cDNA (GenBank accession number NM_031443), beginning nucleotide numbering at A of the ATG initiator codon.

served for the last one, marker D7S2427, in family C116 (fig. 2). Patient III3 in that family looks homozygous and failed to inherit an allele from his homozygous obligate carrier mother (patient II9), suggesting the existence of a null allele; analysis of the other members of the family strongly supported the existence of a null allele cosegregating with the disease (fig. 2). The same family did not exhibit incompatible genotypes with other markers, thus eliminating inaccurate paternity information and sampling error as a source of incompatibility. Failure to amplify one allele can be caused by a polymorphic base in the sequence corresponding to the primers used to amplify the locus. Therefore, new primers were designed to amplify D7S2427, but the previously observed incompatibilities persisted, strongly suggesting the existence of a deletion at this locus in family C116. In contrast, genotypes for the next two flanking markers, D7S478 and D7S621, were heterozygous and were compatible with the family structure, defining a maximum deleted region of 350 kb flanked by those two markers. In each of the remaining 29 families, there was at least one affected individual who was heterozygous at D7S2427, which allowed us to exclude a deletion at that marker.

Detection of an Additional Deletion within Family C127

To refine the critical 350-kb deletion interval and to screen for putative deletions in the other families, two informative CA repeats located between D7S478 and D7S621—MS004844 and MS004847—were genotyped in all 30 families. In family C116, a null allele cosegregating with the affected phenotype was observed with MS004847, whereas, for MS004844, heterozygous affected members were identified (fig. 3*B*). Another MS004847 null allele cosegregating with the disease was detected in family C127 (fig. 3*B*). The MS004847 null alleles were confirmed with a new set of primers. MS004847 screening identified at least one heterozygous affected individual in each of the 28 remaining families, allowing us to exclude a deletion at that marker.

CCM2 *Deletions Refinement*

To further characterize the C116 and C127 deletions and their region of overlap, 42 SNPs located between D7S478 and MS004844 were genotyped in both families. In family C116, non-Mendelian inheritance compatible with a null allele was observed with rs1968075, as well as with five markers located between rs1968075 and D7S2427 (rs7792760, rs3735484, rs3735485, rs4724339, and MS004847), whereas genotypes for the most distant flanking markers, D7S478 and rs4724344, were heterozygous and compatible with the family structure (fig. 3*B*). These data were consistent with a deletion with an estimated size of 81–120-kb, encompassing part or all exons of two known genes, *MGC4607* and *Myosin1G,* and of two putative genes, XM_212099 and XM_212101. In family C127, null alleles cosegregating with the disease were observed with rs4724339, rs4724350, and six internal markers (MS004847, rs4551245, rs4724344, rs4724347, rs4724348, and rs4724349), whereas genotypes for the most distant markers, SNP rs12108 and *MGC4607* exon 4 nt358G/A, were heterozygous in some affected patients. That deletion had an estimated size of ~25.5–81 kb and encompassed part of one known gene, *MGC4607,* and one putative gene, XM_212099 (fig. 3*B*). In conclusion, the two deletions observed in families C116 and C127 overlapped over an interval, estimated at 12–29 kb, thus defining our new critical interval (fig. 3*B*).

MGC4607 *Point Mutations*

The first coding exon of two different genes maps within this new critical interval, including one known gene designated *MGC4607* (GenBank accession number NM_031443) and one putative gene, designated XM_212099. The putative gene XM_212099 includes two

exons, of which only the first is coding. We used SSCP to screen the coding exon in the 28 probands without any identified *CCM2* deletion, and we did not detect any abnormal conformer (data not shown). The second gene, *MGC4607,* comprises 10 coding exons (see next paragraph). *MGC4607* point mutations were identified in eight probands (table 2; fig. 4). Proband C038 harbored a mutation within the initiating ATG codon (nt 1A→G). Proband C039 had a splice-site mutation within intron 3 (nt288+1G→A), leading to a transcript deleted of exon 3. Proband C037 had a deletion/insertion at the boundary between IVS3 and exon 4, leading to a transcript deleted of exons 3 and 4 with a PTC. Two probands harbored mutations leading to a direct frameshift and a premature stop codon (C096/exon 5/nt 495_496insC; C047/exon10/nt 1248_1249delAG). Proband C002 har-

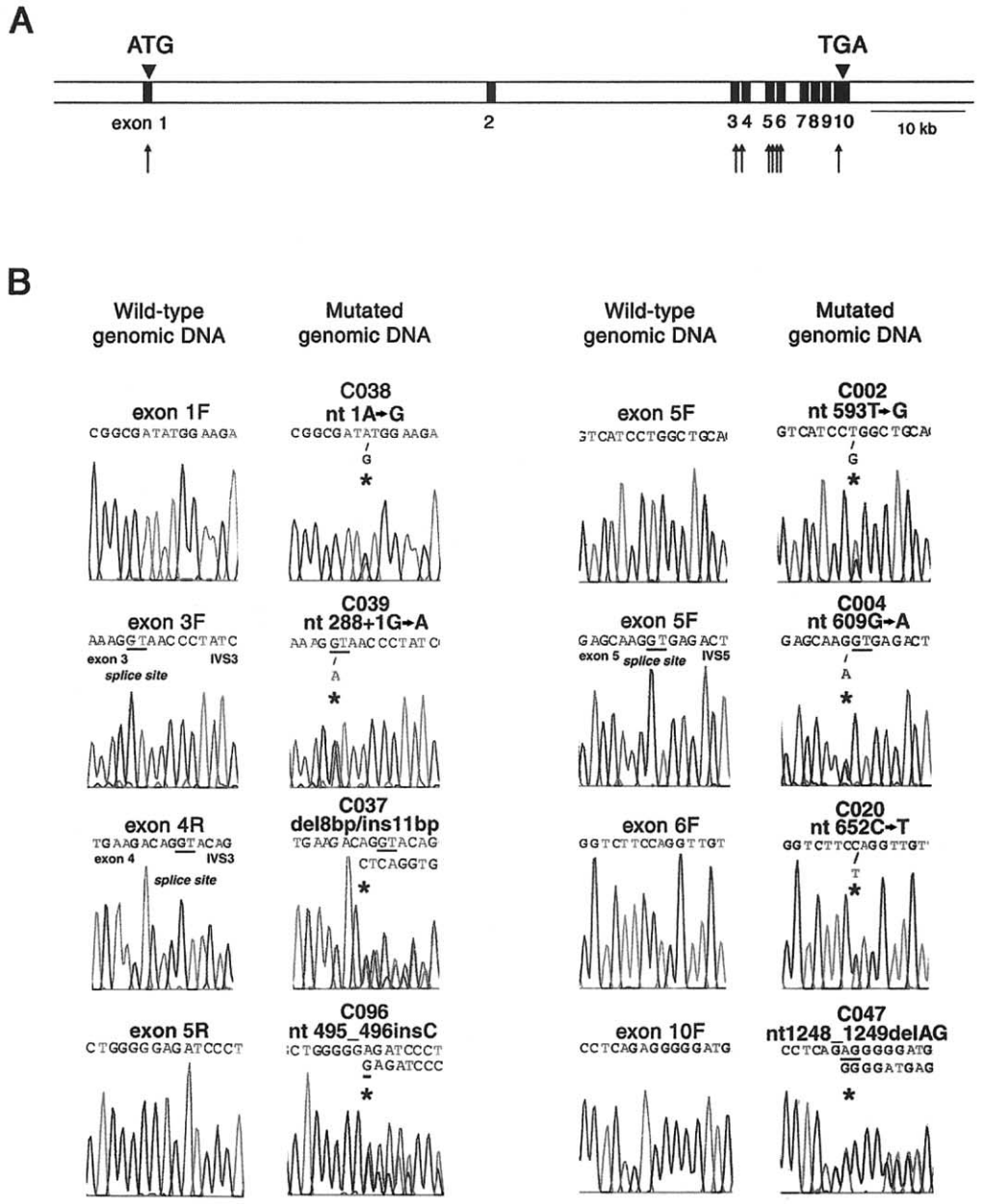

Figure 4 *MGC4607* point mutations. *A, MGC4607* genomic organization. The 10 *MGC4607* exons are numbered and are indicated by vertical hatches. The ATG initiator codon is included in the first exon, and the TGA stop codon in the 10th (last) exon. The locations of the eight different *MGC4607* point mutations are indicated by arrows. *B,* Genomic wild-type and mutated alleles in families harboring point mutations. Family numbers are given above the mutation. F = forward primers; R = reverse primers.

bored a nucleotide substitution leading to a missense mutation within exon 5 (nt 593T→G/aa 198Leu→Arg). Proband C020 harbored a nonsense mutation at codon 218 within exon 6 (nt 652C→T). Proband C004 had an exonic nucleotide substitution that replaces the last base of the exon 5 (nt 609G→A). That mutation is theoretically predicted to lead to a wobble at codon 203 (AAG→AAA/Lys→Lys), but it was quite surprising that this mutation led to an abnormal splicing of exon 5 and a PTC. For probands of families C116 and C127, sequencing of amplified cDNA showed that only one *MGC4607* allele was transcribed. Indeed, those two individuals, who were heterozygous at the genomic DNA level for a polymorphism located within exon 4, showed only one allele at the cDNA level.

In summary, a total of 10 different mutations, including 2 large deletions and 8 point mutations, have been observed in this panel of 30 families with CCM. All these mutations cosegregated with the CCM disease phenotype within the respective families and were not detected in 192 control chromosomes from subjects with a similar ethnic background.

In addition to these mutations, four different exonic variants were also identified and were considered as polymorphic variants, since they were detected at least once in one healthy control individual. One of those variants had been reported elsewhere in databases (rs2289367, substitution nt 915G→A in exon 8, ACG/ACA Thr305Thr) (Entrez Genome). The last three were novel substitutions (substitutions nt 126G→A (Leu42Leu), nt 157G→A (Val53Ile) in exon 2, and nt 358G→A (Val120Ile) in exon 4.

Structure of the MGC4607 *Gene, cDNA, and Predicted Protein*

MGC4607 cDNA and genomic structures have been reported in several genome databases with >100 reported ESTs (Entrez Genome; Center for Applied Genomics databases [ref C7orf22]; GenBank [accession number NM_031443]). It extends over 76 kb and includes 10 coding exons. The coding portion of the cDNA is 1,332 bp long and encodes for a 444-aa predicted protein. The ATG initiating codon is located in the first exon and is preceded 64 bp upstream by an inframe stop codon. Exon 1 is embedded in a 1,086-bp CpG island.

A search of databases did not result in identification of any paralog but identified several orthologs in various vertebrate species, including *Mus musculus* (*MGC37115*, 94.6% identity at the protein level with *MGC4607*) and *Rattus norvegicus* (GenBank accession number XP_223622). Partial homologous transcripts also have been reported for *Sus scrofa* (Ss.12675), *Bos taurus* (Bt. 7707), *Gallus gallus* (clone pgm2n.pk012h10), *Danio rerio* (ENSDARG00000013705), and *Fugu rubripes* (SINFRU00000133870) (Ensembl). We did not identify orthologs in invertebrates, including *Drosophila* and *Caenorhabditis elegans* (HomoloGene).

Searches of protein databases (ExPASy Molecular Biology Server) with *Homo sapiens MGC4607* amino acid sequence did not reveal a signal sequence, transmembrane domain, or any other functional or conserved domain, except for a phosphotyrosine interaction domain in the N-terminal half of the protein (aa 67–223), conserved in other homologous species (NCBI Conserved Domain database).

MGC4607 *Expression Analysis*

RT-PCR analysis with a human normalized multiple-tissue cDNA panel (MTC I [Clontech]), in addition to cDNA from control EBV lymphoblasts, showed that *MGC4607* was expressed in all tissues tested, in agreement with the expression pattern that can be deduced from the multiple *MGC4607* EST reported in the dBEST database. In all tested cDNAs, two populations of transcripts were found that differed by the absence or presence of exon 2, as reported in various ESTs (fig. 5) (e.g., ESTs BM912546, BX337466, and BX365022). This splicing event preserves the *MGC4607* ORF, since exon 2 is composed of 174 nucleotides. Northern blot analysis confirmed this ubiquitous expression. A band of ~1.8 kb was detected in all tissues tested, although at a different level (fig. 6), consistent with the 1,837-bp size of *MGC4607* cDNA reported in databases (GenBank accession number NM_031443).

Discussion

We have identified *MGC4607* as the *CCM2* gene, on the basis of the identification, in 10 unrelated families with CCM, of 10 distinct mutations within this gene. The nature of these mutations strongly suggests that at least some

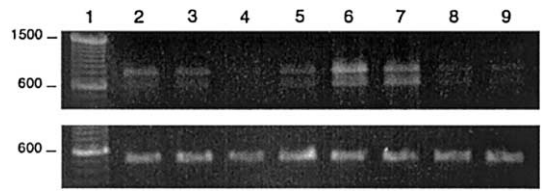

Figure 5 *MGC4607* expression analysis through use of RT-PCR. The expression of *MGC4607* transcripts was analyzed by RT-PCR through use of the cDNA MTC1 panel (Clontech). *Upper panel, MGC4607. Lower panel, β*-actin. *Lane 1,*100-bp ladder. *Lane 2,* Brain. *Lane 3,* Heart. *Lane 4,* Skeletal muscle. *Lane 5,* Kidney. *Lane 6,* Liver. *Lane 7,* Pancreas. *Lane 8,* Lung. *Lane 9,* Placenta. In all tested cDNAs, two populations of *MGC4607* transcripts existed, although at a different level, containing exon 2 (fragment from exons 1–6: 755 bp) or not containing exon 2 (fragment without exon 2: 581 bp).

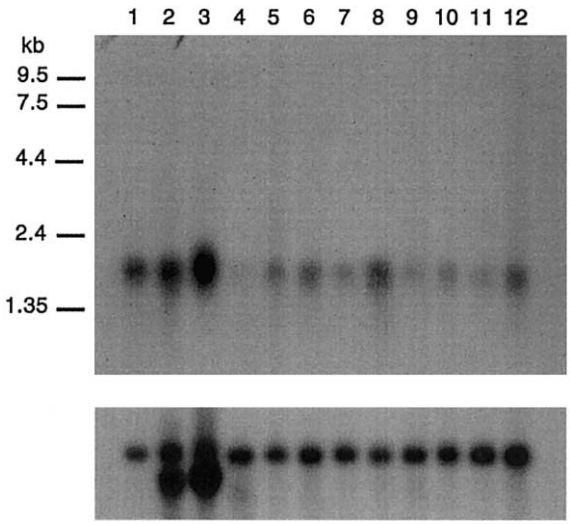

Figure 6 *MGC4607* northern blot analysis. Human adult MTN (human, 12 lanes [Clontech]) was hybridized with an *MGC4607* (*upper panel*) and β-actin (*lower panel*) human cDNA probe. *Lane 1*, Brain. *Lane 2*, Heart. *Lane 3*, Skeletal muscle. *Lane 4*, Colon. *Lane 5*, Thymus. *Lane 6*, Spleen. *Lane 7*, Kidney. *Lane 8*, Liver. *Lane 9*, Small intestine. *Lane 10*, Placenta. *Lane 11*, Lung. *Lane 12*, Peripheral blood leukocytes. Expression of ~1.8-kb *MGC4607* transcripts was detected in all tissues studied.

of these mutations led to *MGC4607* haploinsufficiency. Indeed, one of these mutations altered the initiation codon, and two other mutations deleted the first exon of this gene, leading to a nondetectable expression of one of the two alleles. Five mutations led to premature stop codons through direct nonsense, frameshifts or splice-site mutation events. One mutation led to the spicing of exon 3 without a frameshift. The last mutation, L198R, a missense mutation, cosegregated with the affected phenotype in family C002 and was not detected in the control panel; complete sequencing of *MGC4607* did not detect any other mutation or aberrant subsequent splice event within family C002. On the basis of these data, we considered that this mutation might be causing the disease in this family. However, that cannot yet be established definitely, since we cannot exclude a rare polymorphism, especially since the amino acid located at this position is not conserved in the *MGC4607* mouse ortholog.

We detected a mutation in 10 of 30 of our families. The absence of mutation in the 20 additional families is not unexpected, on the basis of previous linkage data. Indeed, CCM is genetically heterogeneous, and multilocus linkage data strongly suggested that only 20% of families with CCM have linkage to the *CCM2* locus (Craig et al. 1998). Since these 30 families were initially selected on the basis of a negative *KRIT1* mutation screening, one would have expected to detect ~10 families with mutations within this panel. Nevertheless, we cannot exclude the possibility that we may have missed a few mutations, including deletions and/or promotor mutations. This may be the case for family C049, in which we did not detect any mutation, although linkage data were consistent with linkage to the *CCM2* locus and excluded linkage to *CCM3*. A complete deletion of the *CCM2* gene has been excluded in this family, as well in the other 19 families without mutations, on the basis of heterozygosity data; however, additional work will be needed to exclude deletions of one or more exons.

Little is known yet about *MGC4607*. Its pattern of expression at the transcriptional level appears quite ubiquitous at present, and much has to be done to decipher its function. *Krit1* has been shown to interact with rap1A, a small Ras-like GTPase protein, and with ICAP1α, an integrin β1 modulator (Serebriiskii et al. 1997; Zhang et al. 2001; Zawistowski et al. 2002). Future studies will help to determine if *MGC4607* plays a role in these pathways.

Acknowledgments

We thank all families and all members of the Société Française de Neurochirurgie for their participation. We are also indebted to Dr. F. Riant (INSERM E365), for the patients' previous *KRIT1* screening; F. Bergametti (INSERM E365); C. Lescoat; and M. Cecillon (INSERM E365 and Laboratoire de Cytogénétique, Hôpital Lariboisière, Assistance Publique-Hôpitaux de Paris); as well as O. Gribouval (INSERM U574), for technical help in part of this work. We also thank A. Joutel for excellent critical review of this article. C.D. and S.G. are recipients of fellowships from Poste Accueil INSERM and Fondation pour la Recherche Médicale, respectively. This work was supported by INSERM and Association Française contre les Myopathies (Cohortes Maladies rares 2000), GIS Maladies rares (AO 2002–2004), and the Ministère Français de la Recherche.

Electronic-Database Information

Accession numbers and URLs for data presented herein are as follows:

Center for Applied Genomics, http://www.cag.icph.org/
Centre d'Etude du Polymorphisme Humain, http://www.cephb.fr/
Chromosome 7 Annotation Project, The, http://www.chr7.org/ (for sequence and SNP information about genomic contigs on chromosome 7)
Cooperative Human Linkage Center, http://gai.nci.nih.gov/CHLC/ (for *CCM2* and *CCM3* marker selection)
Ensembl, http://www.ensembl.org/Danio_rerio/ and http://www.ensembl.org/Fugu_rubripes/ (for the homology searches)
Entrez Genome, http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi? (for data on *CCM2* and *CCM3* marker selection and individual SNPs)
ExPASy Molecular Biology Server, http://us.expasy.org/ (for structural features prediction)
GenBank, http://www.ncbi.nlm.nih.gov/GenBank/ (for sequence

and SNP information for genomic contigs on 7p: *MGC4607* [accession number NM_031443] and contig NT_007819 [accession number GI:29796755])

HomoloGene, http://www.ncbi.nlm.nih.gov/HomoloGene/ (for the *MGC4607* homology search)

NCBI Conserved Domain Database, http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd/shtml

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for CCM)

Single Nucleotide Polymorphism Web site, http://www.ncbi.nlm.nih.gov/SNP/ (for data on individual SNPs)

References

Cave-Riant F, Denier C, Labauge P, Cécillon M, Maciazek J, Joutel A, Laberge-le Couteulx S, Tournier-Lasserve E (2002) Spectrum and expression analysis of *KRIT1* mutations in 121 consecutive and unrelated patients with cerebral cavernous malformations. Eur J Hum Genet 10:733–740

Chen DH, Lipe HP, Qin Z, Bird TD (2002) Cerebral cavernous malformation: novel mutation in a Chinese family and evidence for heterogeneity. J Neurol Sci 196:91–96

Couteulx SL, Brezin AP, Fontaine B, Tournier-Lasserve E, Labauge P (2002) A novel KRIT1/CCM1 truncating mutation in a patient with cerebral and retinal cavernous angiomas. Arch Ophthalmol 120:217–218

Craig HD, Gunel M, Cepeda O, Johnson EW, Ptacek L, Steinberg GK, Ogilvy CS, Berg MJ, Crawford SC, Scott RM, Sabroe R, Kennedy CT, Mettler G, Beis MJ, Fryer A, Awad IA, Lifton RP (1998) Multilocus linkage identifies two new loci for a Mendelian form of stroke, cerebral cavernous malformation, at 7p15-13 and 3q25.2-27. Hum Mol Genet 7:1851–1858

Davenport WJ, Siegel AM, Dichgans J, Drigo P, Mammi I, Pereda P, Wood NW, Rouleau GA (2001) CCM1 gene mutations in families segregating cerebral cavernous malformations. Neurology 56:540–543

Dubovsky J, Zabramski JM, Kurth J, Spetzler RF, Rich SS, Orr HT, Weber JL (1995) A gene responsible for cavernous malformations of the brain maps to chromosome 7q. Hum Mol Genet 4:453–458

Dupre N, Verlaan DJ, Hand CK, Laurent SB, Turecki G, Davenport W, Acciarri N, Dichgans J, Ohkuma A, Siegel A, Rouleau GA (2003) Linkage to the CCM2 locus and genetic heterogeneity in familial cerebral cavernous malformation. Can J Neurol Sci 30:122–128

Eerola I, Plate KH, Spiegel R, Boon LM, Mulliken JB, Vikkula M (2000) *KRIT1* is mutated in hyperkeratotic cutaneous capillary-venous malformation associated with cerebral capillary malformation. Hum Mol Genet 22:1351–1355

European Chromosome 16 Tuberous Sclerosis Consortium, The (1993) Identification and characterization of the tuberous sclerosis gene on chromosome 16. Cell 75:1305–1315

Labauge P, Laberge S, Brunereau L, Levy C, Maciazek J, Tournier-Lasserve E (1998) Hereditary cerebral cavernous angiomas: clinical and genetic features in 57 French families. Lancet 352:1892–1897

Laberge S, Labauge P, Marechal E, Maciazek J, Tournier-Lasserve E (1999) Genetic heterogeneity and absence of founder effect in a series of 36 French cerebral cavernous angiomas families. Eur J Hum Genet 7:499–504

Laberge-le Couteulx S, Jung HH, Labauge P, Houtteville JP, Lescoat C, Cecillon M, Marechal E, Joutel A, Bach JF, Tournier-Lasserve E (1999) Truncating mutations in CCM1, encoding *KRIT1,* cause hereditary cavernous angiomas. Nat Genet 23:189–193

Laurans MS, DiLuna ML, Shin D, Niazi F, Voorhees JR, Nelson-Williams C, Johnson EW, Siegel AM, Steinberg GK, Berg MJ, Scott RM, Tedeschi G, Enevoldson TP, Anson J, Rouleau GA, Ogilvy C, Awad IA, Lifton RP, Gunel M (2003) Mutational analysis of 206 families with cavernous malformations. J Neurosurg 99:38–43

Lucas M, Costa AF, Garcia-Moreno JM, Solano F, Gamero MA, Izquierdo G (2003) Variable expression of cerebral cavernous malformations in carriers of a premature termination codon in exon 17 of the Krit1 gene. BMC Neurol 3:5

Lucas M, Costa AF, Montori M, Solano F, Zayas MD, Izquierdo G (2001) Germline mutations in the CCM1 gene, encoding *KRIT1,* cause cerebral cavernous malformations. Ann Neurol 49:529–532

Marini V, Ferrera L, Dorcaratto A, Viale G, Origone P, Mareni C, Garre C (2003) Identification of a novel KRIT1 mutation in an Italian family with cerebral cavernous malformation by the protein truncation test. J Neurol Sci 212:75–78

Musunuru K, Hillard VH, Murali R (2003) Widespread central nervous system cavernous malformations associated with cafe-au-lait skin lesions: case report. J Neurosurg 99:412–415

Otten P, Pizzolato GP, Rilliet B, Berney J (1989) 131 cases of cavernous angioma (cavernomas) of the CNS, discovered by retrospective analysis of 24,535 autopsies. Neurochirurgie 35: 82–83 and 128–131

Rigamonti D, Hadley MN, Drayer BP, Johnson PC, Hoenig-Rigamonti K, Knight JT, Spetzler RF (1988) Cerebral cavernous malformations: incidence and familial occurrence. N Engl J Med 319:343–347

Robinson JR, Awad IA, Little JR (1991) Natural history of the cavernous angiomas. J Neurosurg 75:709–714

Russel DS, Rubinstein LJ (1989) Pathology of tumors of the nervous system, 5th ed. Williams and Wilkins, Baltimore, pp 730–736

Sahoo T, Goenaga-Diaz E, Serebriiskii IG, Thomas JW, Kotova E, Cuellar JG, Peloquin JM, Golemis E, Beitinjaneh F, Green ED, Johnson EW, Marchuk DA (2001) Computational and experimental analyses reveal previously undetected coding exons of the *KRIT1* (CCM1) gene. Genomics 71:123–126

Sahoo T, Johnson EW, Thomas JW, Kuehl PM, Jones TL, Dokken CG, Touchman JW, Gallione CJ, Lee-Lin SQ, Kosofsky B, Kurth JH, Louis DN, Mettler G, Morrison L, Gil-Nagel A, Rich SS, Zabramski JM, Boguski MS, Green ED, Marchuk DA (1999) Mutations in the gene encoding *KRIT1,* a Krev-1/rap1a binding protein, cause cerebral cavernous malformations (CCM1). Hum Mol Genet 8:2325–2333

Serebriiskii I, Estojak J, Sonoda G, Testa JR, Golemis EA (1997) Association of Krev-1/rap1a with *KRIT1,* a novel ankyrin repeat-containing protein encoded by a gene mapping to 7q21-22. Oncogene 15:1043–1049

Trofatter JA, MacCollin MM, Rutter JL, Murrell JR, Duyao MP, Parry DM, Eldridge R, et al (1993) A novel moesin-,

ezrin-, radixin-like gene is a candidate for the neurofibromatosis 2 tumor suppressor. Cell 72:791–800

Verlaan DJ, Siegel AM, Rouleau GA (2002) *Krit1* missense mutations lead to splicing errors in cerebral cavernous malformation. Am J Hum Genet 70:1564–1567

Viskochil D, Buchberg AM, Xu G, Cawthon RM, Stevens J, Wolff RK, Culver M, et al (1990) Deletions and a translocation interrupt a cloned gene at the neurofibromatosis type 1 locus. Cell 62:187–192

Xu YL, Zhao JZ, Wu B, Zhong H, Wang S, Heng WJ (2003) A novel Krit-1 mutation in Han family with cerebral cavernous malformation. Zhonghua Bing Li Xue Za Zhi 32:220–225

Zawistowski JS, Serebriiski IG, Lee MF, Golemis EA, Marchuk DA (2002) KRIT1 association with the integrin-binding protein ICAP-1: a new direction in the elucidation of cerebral cavernous malformations (CCM1) pathogenesis. Hum Mol Genet 11:389–396

Zhang J, Clatterbuck RE, Rigamonti D, Chang DD, Dietz HC (2001) Interaction between Krit1 and Icap1 infers perturbation of integrin β1-mediated angiogenesis in the pathogenesis of cerebral cavernous malformation. Hum Mol Genet 10:2953–2960

Zhang J, Clatterbuck RE, Rigamonti D, Dietz HC (2000) Mutations in *KRIT1* in familial cerebral cavernous malformations. Neurosurgery 46:1272–1277