

A founder mutation in the Ashkenazi Jewish population affecting messenger RNA splicing of the *CCM2* gene causes cerebral cavernous malformations

Carol J. Gallione, BA¹, Ann Solatycki, BS², Issam A. Awad, MD³, James L. Weber, PhD², and Douglas A. Marchuk, PhD¹

Purpose: Cerebral cavernous malformations can occur sporadically or are caused by mutations in one of three identified genes. Cerebral cavernous malformations often remain clinically silent until a mutation carrier suffers a stroke or seizure. Presymptomatic genetic testing has been valuable to follow and manage cerebral cavernous malformation mutation carriers. During routine diagnostic testing, we identified a two base pair change in seven unrelated people of Ashkenazi Jewish heritage. Because of the location of the variant beyond the invariant splice donor sequence, the change was reported as a variant of unknown significance. In this study, we determined whether this change was a disease-causing mutation and whether it represents a founder mutation in the Ashkenazi Jewish population. **Methods:** Transcripts arising from the normal and mutant alleles were examined by reverse transcription-polymerase chain reaction from affected and unaffected Ashkenazi Jewish cerebral cavernous malformation family members. A synthetic splicing system using a chimeric exon was used to visualize the effects of the change on splice donor site utilization. **Results:** The two base pair change in *CCM2*, c.30 + 5_6delinsTT, segregated with affected status in the study families. Reverse transcription-polymerase chain reaction revealed loss of the transcript allele that was in phase with the mutation. The two base pair change, when tested in an in vitro synthetic splicing system, altered splice donor site utilization. Resequencing of the genomic region proximal and distal to the *CCM2* gene mutation revealed a common single-nucleotide polymorphism haplotype in affected individuals. **Conclusions:** The two base pair change in *CCM2*, c.30 + 5_6delinsTT, disrupted proper splice donor utilization leading to a degraded transcript. Single nucleotide polymorphism haplotype analysis demonstrated that this mutation was due to a founder in the Ashkenazi Jewish population. These data have the potential to simplify genetic testing for cerebral cavernous malformation in the Ashkenazi Jewish population. *Genet Med* 2011;13(7):662–666.

Key Words: cerebral cavernous malformation, *CCM2*, mRNA splicing, founder mutation, Ashkenazi Jews

Cerebral cavernous malformations (CCMs, OMIM# 116860) are vascular malformations of the central nervous system consisting of clusters of dilated vessels lined by a single layer of

endothelium, lacking mature vessel wall elements, and filled with blood at various stages of thrombosis. CCMs can be clinically silent or can grow or bleed, resulting in stroke, seizures, and focal neurologic deficits (reviewed in Ref. 1). CCMs occur sporadically as solitary lesions and as an autosomal dominant trait in cases with multifocal lesions throughout the central nervous system. Three genes have been identified for inherited CCM, and mutations in one of these three genes have been identified in the majority of CCM families.^{2–6}

Genetic testing for CCM is now routinely available for CCM individuals and families and has value because of the potential for mutation carriers to appear clinically silent until a significant hemorrhage occurs. Identification of a mutation in a family member allows the rapid screening of the entire kindred for that specific mutation, reassures those who do not carry the mutation, and focuses clinical evaluation and follow-up on mutation carriers.

Although most mutations in each of the *CCM* genes are family specific, two common mutations have been identified. A 76 kb deletion of exons 2–10 in the *CCM2* gene has been identified in multiple samples from an as yet undefined subpopulation in the United States.⁷ Nearly all CCM families from the Mexican-American Hispanic population harbor the identical nonsense *CCM1* mutation³ because of a founder effect in this population.^{3,8,9} The knowledge of these common mutations has simplified genetic testing for CCM, especially in the Mexican-American Hispanic population.

The identical two base pair change in *CCM2*, c.30 + 5_6delinsTT, was identified in seven apparently unrelated probands from 10 different kindreds of Ashkenazi Jewish (AJ) descent. However, the clinical significance of this variant was unknown. Although it lies near the splice donor junction of the first exon, the change does not alter the canonical splice donor site. In this study, we examined genomic DNA and messenger RNA (mRNA) of AJ CCM families and individuals with this variant. We also constructed a novel chimeric exon harboring this mutation and used this chimera in a synthetic mRNA splicing system to determine whether this variant altered recognition of the splice donor by the cellular splicing machinery. Finally, as there are numerous founder mutations for various inherited disorders in the AJ population,^{10,11} we also examined single-nucleotide polymorphism (SNP) haplotypes across this variant to determine whether this change represents a founder mutation in this population.

METHODS

Study subjects and nucleic acid extraction

Individuals were enrolled in the study after giving informed consent (approved by the Duke University Health System Institutional Review Board Committee). Diagnosis of CCM was based on a positive magnetic resonance imaging scan. AJ her-

From the ¹Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, North Carolina; ²PreventionGenetics, Marshfield, Wisconsin; and ³The Neurovascular Surgery Program and Section of Neurosurgery, University of Chicago Pritzker School of Medicine, Chicago, Illinois.

Douglas A. Marchuk, PhD, Box 3175, Duke University Medical Center, Durham, NC 27710. E-mail: douglas.marchuk@duke.edu.

Disclosure: The authors declare no conflict of interest.

Submitted for publication December 13, 2010.

Accepted for publication January 24, 2011.

Published online ahead of print May 3, 2011.

DOI: 10.1097/GIM.0b013e318211ff8b

itage was identified prospectively by the patients in a standardized questionnaire at the time of referral for genetic testing and was verified by telephone interview at the time of informed consent for this study. Genomic DNA and RNA were extracted from peripheral blood leukocytes from members of family 380 using the Gentra PureGene Blood Kit (Qiagen, Chatsworth, CA) and TRIzol Reagent (Invitrogen, San Diego, CA) extraction protocols, respectively, as per the manufacturer’s directions. Genomic DNA was extracted from saliva from members of family 377 using the Oragene-DNA kit as per manufacturer’s directions (DNAgenotek, Inc.). Nucleotides are numbered with the conventional nomenclature of the A in the first ATG as c.1.

Sequencing

Coding exons 1 and 2 of *CCM2* were polymerase chain reaction (PCR) amplified and sequenced. Any sequence variants were reamplified and resequenced to verify the changes. SNPs for the haplotype analysis were identified using the International HapMap Project database and were PCR amplified and sequenced. Allele frequencies were determined using the CEU and TSI population frequency figures on HapMap. Complementary DNA was prepared from RNA from peripheral blood and from 293T cells using M-MLV reverse transcriptase (Invitrogen) and oligo-dT as per manufacturer’s instructions. Reverse transcription-PCR products were separated on agarose gels, eluted, and sequenced. The full-length *CCM2* transcript is NM_031443, and the transcript initiated from the alternate exon 1 is NM_001029835.

Plasmid construction and transfection

A chimeric exon was constructed to resemble an internal exon with both splice acceptor and splice donor sites. The

chimeric exon contains approximately 340 bases of upstream intronic sequence, including the splice acceptor site, and approximately 100 bases of endoglin exon 2. This was ligated in frame to the 3’ portion of *CCM2* exon 1 with approximately 340 bases of contiguous intronic sequence, including the splice donor site. This entire insert was ligated into the MCS of pSPL3, a splicing vector.¹² The specific mutation, c.30 + 5,6delinsTT, was introduced using site-directed mutagenesis. Plasmid DNA was transfected into 293T cells using Fugene 6 Transfection Reagent (Roche), incubated for 24 hours, and then the RNA was extracted using TRIzol.

RESULTS

The *CCM2* c.30 + 5_6delinsTT was the only previously unreported change identified in any of the three *CCM* genes in seven unrelated individuals of AJ background. This variant is not present in the SNP databases. However, because the variant did not fall within the canonical splice site sequence, it was reported as a variant of unknown significance. We resequenced genomic DNA from four individuals from this original group of seven clinical samples and samples from other family members for two of the probands (families 377 and 380, Fig. 1, A and B). We reidentified the same change in all the affected probands, all other known affected family members, and in some family members of unknown disease status (Fig. 1, A and B). Thus, this variant met the minimal mutation criterion of segregating with affected status in two families. We also noted that all the individuals with the putative mutation also showed at least one copy of the minor (A) allele at rs2107732, a SNP in *CCM2* exon 2 (Fig. 1B), demonstrating that for these families, the A allele in

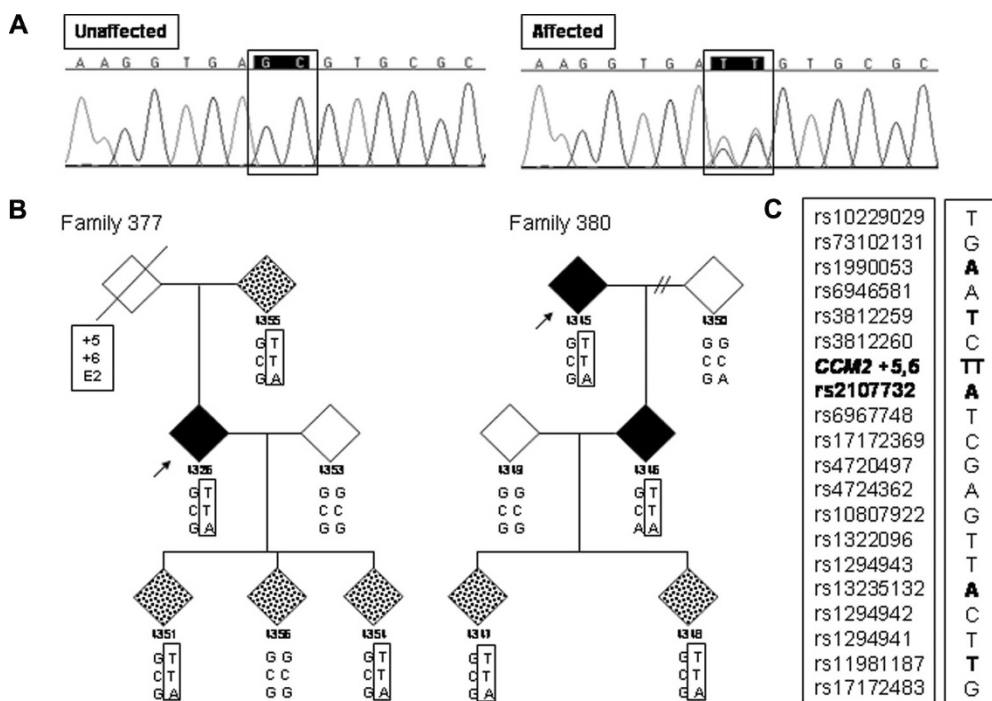


Fig. 1. A, DNA sequence traces from unaffected and affected individuals. The nucleotides at position c.30 + 5_6 are boxed. B, Pedigrees of two Ashkenazi Jewish families in this study. Genders are obscured to protect patient confidentiality. Clear diamonds indicate unaffected, dark diamonds indicate clinically affected, and speckled diamonds indicate uncertain disease status. Proband is indicated with arrows. The genotypes at c.30 + 5_6 and the exon 2 SNP (rs2107732) are noted. C, The *CCM2* AJ affected SNP haplotype is shown with the rare alleles shown in bold.

exon 2 was in phase (linkage disequilibrium) with the putative mutation flanking exon 1. As the putative mutation does not itself fall within the coding region of the *CCM2* transcript, we realized that rs2107732 could serve as a surrogate to follow the fate of the normal and mutant transcripts.

Because the putative mutation occurs near the splice donor site of *CCM2* exon 1 (Fig. 2A), we postulated that it might affect mRNA splicing. Different splice site recognition programs consistently predict that the putative mutation reduces the strength of the splice donor site.^{13–15} However, as this variant would alter the splice donor site of the first exon of the *CCM2* gene, an exon that by definition cannot be skipped, we did not expect to observe an aberrant splice variant that could validate the effect of the putative mutation on mRNA splicing. RT-PCR was performed on RNA isolated from peripheral blood leukocytes from unaffected and affected family members using primers located in exons 1 and 4 of *CCM2*. As expected, affected individuals with the putative mutation did not show an aberrant splice variant. However, for affected individuals who were heterozygous for the rs2107732 SNP in exon 2, where the A variant is in phase with the putative mutation, we could follow the fate of the transcripts generated from both the normal and mutant alleles. Normal individuals who are heterozygous at rs2107732 but lack the c.30 + 5_6delinsTT putative mutation showed approximately equal levels of both alleles in the RT-PCR product (Fig. 2B). Thus, transcription level and message stability are not affected by the allelic status at this SNP. By contrast, affected individuals with the c.30 + 5_6delinsTT putative mutation showed only one allele at rs2107732 in the RT-PCR product (Fig. 2B). Importantly, in all cases examined,

the transcript harboring the A allele—the allele in phase with the putative mutation—was absent. This suggested that the transcript derived from the mutant allele was either not generated or was lost in patients harboring the c.30 + 5_6delinsTT variant.

In addition to the consensus *CCM2* transcript, the *CCM2* gene harbors a rarely used downstream transcription initiation site leading to a transcript containing an alternate exon 1 (Fig. 2A). This alternate transcript is not sufficient to rescue the disease phenotype, however, as there are multiple reports of mutations in the consensus exon 1 of *CCM2*.^{4,5,16,17} Alternate exon 1 maps within the first intron of *CCM2*, distal to the more commonly used transcription start site of the consensus exon 1. As this alternate exon 1 lies downstream from the site of the putative mutation, we investigated whether the use of this alternate transcription start site would affect the presence of transcript from the mutant allele. We again examined RT-PCR products from an affected individual harboring the putative mutation who was also heterozygous for the rs2107732 SNP in exon 2. When using amplification primers located in alternate exon 1 and in exon 3, we observed both alleles of rs2107732 in the transcript (Fig. 2B). This indicates that when transcription begins from a point downstream of the putative mutation, the transcript from the mutant allele is not lost. This provided further support that c.30 + 5_6delinsTT affects the splicing of the consensus *CCM2* transcript and, therefore, is a bona fide mutation.

These combined observations are consistent with a model where the c.30 + 5_6delinsTT mutation destabilizes the splice donor site of the first exon, leading to an aberrant transcript of

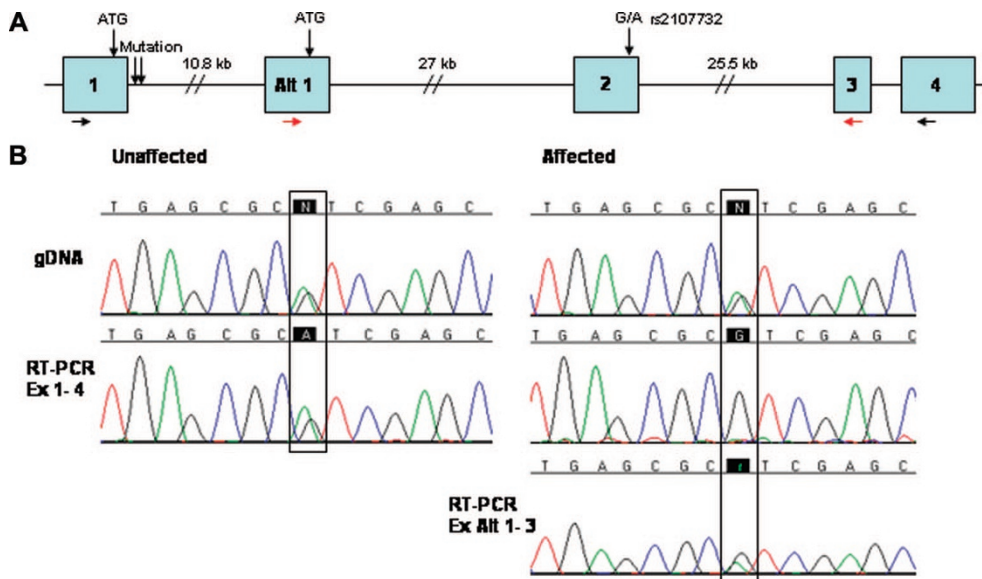


Fig. 2. A, Schematic diagram of *CCM2* exons 1–4 including the alternate exon 1. The boxes represent the exons, the primary transcription start codon in exon 1, secondary transcription start site in alternate exon 1, and the reporter SNP in exon 2 (rs2107732) are indicated, and the arrows below the exons indicate the primers used for RT-PCR. B, The top tracing for each column is the sequence of genomic DNA from an unaffected individual (left) and an affected individual (right) both also heterozygous at SNP rs2107732 in exon 2. The second set of tracings show RT-PCR products from the same unaffected and affected individuals using primers in exons 1 and 4 (black arrows). Transcripts from both alleles are present in the unaffected family member, but only one transcript is seen in the affected individual, with the transcript derived from the mutant allele being absent. The last tracing shows the RT-PCR product from the affected individual using primers in alternate exons 1 and 3 (red arrows). Transcripts from both alleles are present in the affected individual when transcription initiates downstream of the mutation.

unknown composition, and this mutant transcript is rapidly lost due to degradation. The degradation of the mutant transcript leaves only negative data (loss of the mutant transcript) as evidence of the molecular effects of the variant and precludes the identification of an aberrant splice variant that would allow a formal test of its effect on splice donor site utilization. To investigate the effects of the mutation on donor site utilization, we moved the splice donor region of *CCM2* exon 1 into a synthetic sequence context of a chimeric, internal exon. This chimeric exon contains the splice acceptor region and approximately the 5' half of an internal (2nd) exon from a different gene (endoglin), fused to the 3' half of either the normal or mutant version of *CCM2* exon 1, including the region of the putative mutation and more than 300 bases beyond the mutation (Fig. 3B). The chimeric internal exon harboring the mutation was cloned into the *in vitro* splicing vector, pSPL3.¹² This vector contains a strong initial exon including flanking intronic sequence, a cloning site for insertion of exogenous genomic sequence, and a terminal 3' exon with a proper splice acceptor site (Fig. 3B). If the cloned genomic DNA contains an internal exon flanked by recognizable splice acceptor and donor sites, the exon will be included in the resulting spliced transcript. This system has been used to investigate the effects of sequence variants that might affect splicing but which map outside the canonical splice donor and acceptor sites.^{18–20}

Chimeric exons containing either the wild type or the putative mutant splice donor were transfected into 293T cells. mRNA was isolated after 24 hours and subjected to RT-PCR with vector-specific primers. As expected, the chimera containing the normal *CCM2* exon 1 splice donor produced a spliced product that included the intact chimeric exon (Fig. 3A) demonstrating that even in this artificial genomic context, the normal splice donor site is recognized, albeit not perfectly. By contrast, the chimeric exon harboring the putative splice donor mutation produced an additional, shorter product (Fig. 3A). Sequencing of this shorter product revealed the use of a cryptic splice donor site mapping within the chimeric exon. In this novel genomic context, the cryptic splice site that was chosen by the cellular splicing machinery maps within the portion of the chimeric exon derived from exon 2 of endoglin. This site would not be present in the mutation's normal genomic context

of the *CCM2* gene. Nonetheless, these data confirm that the putative mutation, c.30 + 5_6delinsTT, does indeed destabilize the splice donor site of *CCM2* exon 1, leading to faulty recognition and utilization of the mutant donor site.

Having shown that this sequence variant is indeed a bona fide *CCM2* gene splice donor site mutation, we determined whether the repeated identification of this same mutation in apparently unrelated individuals was due to a founder mutation in the AJ population. To do this, we determined the allele in phase with the mutation at selected SNPs mapping 177 kb proximal and 114 kb distal to *CCM2* in the two families and in the two other probands. We identified a unique SNP haplotype that was shared by all affected members of both families and was consistent in the two separate, unrelated individuals (Fig. 1C). This shared *CCM2* mutation SNP haplotype suggests a common founder for at least these four families and possibly for a majority of AJ CCM patients.

DISCUSSION

In this study, we have shown that the two base pair change in *CCM2*, c.30 + 5_6delinsTT, affects the splice donor site recognition of exon 1, leading to loss of the mutant transcript, and is thus a bona fide disease-causing mutation. Mutations located outside the invariant ag/gt intronic splice acceptor and donor sites can cause aberrant mRNA splicing and are common in human genetic disease.²¹ A pertinent example is the major, ancestral Jewish mutation causing familial dysautonomia. This mutation lies in the +6 splice donor position of intron 20 of the *IKBKAP* gene causing leaky, tissue-specific, skipping of exon 20 in the transcript.^{22,23}

Our discovery that this variant in *CCM2*, c.30 + 5_6delinsTT, is a disease-causing mutation will enable diagnostic genetic testing laboratories to simplify CCM testing for families of AJ descent. As CCM can remain clinically silent until presentation with a cerebrovascular event, presymptomatic genetic testing in individuals with a family history can aid patient management and clinical care.^{24,25} This point is demonstrated in the third generations of the two families in this study where four of the five family members who had not been diagnosed with CCM were found to be carrying the mutation. These four

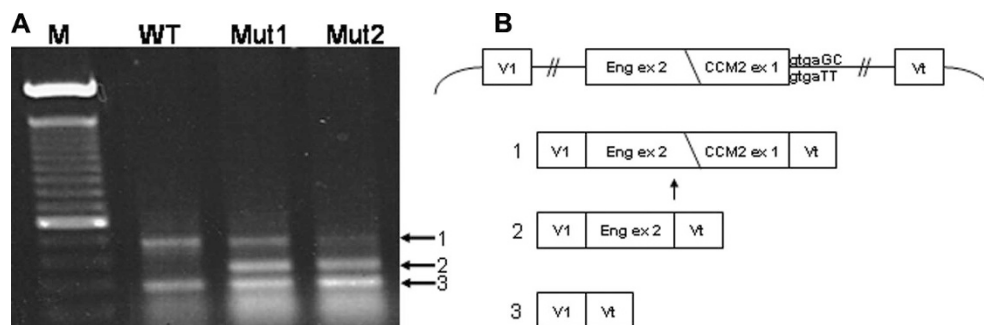


Fig. 3. A, RT-PCR products from the synthetic splicing system containing a chimeric exon. B, Schematic representations of the chimeric exon cloned into the pSPL3 vector (top) and the three different resulting RT-PCR products. The wild-type sequence is above the line, and the mutant sequence is below. 1, Full length transcript containing the vector-supplied exon 1 (V1), the vector-supplied terminal exon (Vt) of the splicing vector pSPL3, and the full sequence of the chimeric exon. 2, The product of the vector-supplied first and terminal exons spliced to a smaller fragment of the chimeric exon. The arrow indicates the position of the cryptic splice site used to generate the truncated transcript. 3, Transcript containing only V1 and Vt. M = 100 bp marker, WT = vector with chimeric exon containing wild-type splice donor sequence, Mut = vector with chimeric exon containing mutant splice donor sequence. Mut1 and Mut2 are from duplicate transfections of the mutant version of the construct.

individuals can now be followed up clinically in a proactive manner. Also important, those who do not harbor the mutation need not be repeatedly monitored for lesions by magnetic resonance imaging.

We have also demonstrated that at least for those individuals and families represented in this study, the *CCM2*, c.30 + 5_6delinsTT mutation is a result of a founder in the AJ population. There are many documented founder mutations in the AJ population for a large number of distinct clinical phenotypes.^{10,11} Because of the adult onset of CCM and the minimal effect on reproductive fitness, this *CCM2* mutation may have drifted to a moderate frequency in the AJ population, similar to the three common inherited AJ breast cancer mutations.²⁶ As is the case with founder effect mutations in other diseases, we do not expect that all Ashkenazi CCM patients will have this mutation. To date, 70% (7/10) of the families declaring AJ heritage who were tested for CCM at one large diagnostic testing center harbor this mutation. This mutation was also found in one additional proband who chose not to report ethnicity, so it is uncertain whether this mutation is present in other ethnic groups. Two of the other three self-reported AJ probands harbored different, unique mutations in the *CCM1* gene, and in the third, the only potential mutation identified was a missense variant of uncertain significance in the *CCM2* gene.

We do not know whether this mutation is only present in the US AJ population or whether it will also be found in the larger worldwide Ashkenazi population. The prevalence of this mutation in the United States and elsewhere depends on the date and place of origin of the founder mutation. The answer to these intriguing questions requires comprehensive testing for this variant in other AJ populations across the globe. In the interim, the discovery of this founder mutation will have immediate impact on genetic testing for CCM in the United States and potentially worldwide.

ACKNOWLEDGMENTS

Supported by NIH 1RO1NS060748 (to DAM and IAA). The authors thank Drs. Amy Akers and Cornelia Lee of Angioma Alliance for assistance in collecting patient samples and the family members for participating in this study. They acknowledge Drs. Wendy S. Rubinstein and Peter Hulick of the Center of Genetic Medicine at the NorthShore University HealthSystem in Evanston, Illinois, for counseling that facilitated the decision of several patients to pursue genetic testing.

REFERENCES

1. Yadla S, Jabbour PM, Shenkar R, Shi C, Campbell PG, Awad IA. Cerebral cavernous malformations as a disease of vascular permeability: from bench to bedside with caution. *Neurosurg Focus* 2010;29:E4.
2. Laberge-le Couteux S, Jung HH, Labauge P, et al. Truncating mutations in *CCM1*, encoding KRIT1, cause hereditary cavernous angiomas. *Nat Genet* 1999;23:189–193.
3. Sahoo T, Johnson EW, Thomas JW, et al. Mutations in the gene encoding KRIT1, a Krev-1/rap1a binding protein, cause cerebral cavernous malformations (CCM1). *Hum Mol Genet* 1999;8:2325–2333.
4. Liquori CL, Berg MJ, Siegel AM, et al. Mutations in a gene encoding a novel protein containing a phosphotyrosine-binding domain cause type 2 cerebral cavernous malformations. *Am J Hum Genet* 2003;73:1459–1464.
5. Denier C, Goutagny S, Labauge P, et al. Mutations within the *MGC4607* gene cause cerebral cavernous malformations. *Am J Hum Genet* 2004;74:326–337.
6. Bergametti F, Denier C, Labauge P, et al. Mutations within the programmed cell death 10 gene cause cerebral cavernous malformations. *Am J Hum Genet* 2005;76:42–51.
7. Liquori CL, Berg MJ, Squitieri F, et al. Deletions in *CCM2* are a common cause of cerebral cavernous malformations. *Am J Hum Genet* 2007;80:69–75.
8. Johnson EW, Iyer LM, Rich SS, et al. Refined localization of the cerebral cavernous malformation gene (*CCM1*) to a 4-cM interval of chromosome 7q contained in a well-defined YAC contig. *Genome Res* 1995;5:368–380.
9. Gunel M, Awad IA, Finberg K, et al. A founder mutation as a cause of cerebral cavernous malformation in Hispanic Americans. *N Engl J Med* 1996;334:946–951.
10. Motulsky AG. Jewish diseases and origins. *Nat Genet* 1995;9:99–101.
11. Charrow J. Ashkenazi Jewish genetic disorders. *Fam Cancer* 2004;3:201–206.
12. Church DM, Stotler CJ, Rutter JL, Murrell JR, Trofatter JA, Buckler AJ. Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. *Nat Genet* 1994;6:98–105.
13. Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S. Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information. *Nucleic Acids Res* 1996;24:3439–3452.
14. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol* 1997;4:311–323.
15. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004;11:377–394.
16. Verlaan DJ, Laurent SB, Rochefort DL, et al. *CCM2* mutations account for 13% of cases in a large collection of kindreds with hereditary cavernous malformations. *Ann Neurol* 2004;55:757–758.
17. Gianfrancesco F, Cannella M, Martino T, et al. Highly variable penetrance in subjects affected with cavernous cerebral angiomas (CCM) carrying novel *CCM1* and *CCM2* mutations. *Am J Med Genet B Neuropsychiatr Genet* 2007;144B:691–695.
18. Svenson IK, Ashley-Koch AE, Gaskell PC, et al. Identification and expression analysis of spastin gene mutations in hereditary spastic paraplegia. *Am J Hum Genet* 2001;68:1077–1085.
19. Svenson IK, Ashley-Koch AE, Pericak-Vance MA, Marchuk DA. A second leaky splice-site mutation in the spastin gene. *Am J Hum Genet* 2001;69:1407–1409.
20. Wheeler FC, Tang H, Marks OA, et al. Tnni3k modifies disease progression in murine models of cardiomyopathy. *PLoS Genet* 2009;5:e1000647.
21. Ward AJ, Cooper TA. The pathobiology of splicing. *J Pathol* 2010;220:152–163.
22. Anderson SL, Coli R, Daly IW, et al. Familial dysautonomia is caused by mutations of the *IKAP* gene. *Am J Hum Genet* 2001;68:753–758.
23. Slangenaupt SA, Blumenfeld A, Gill SP, et al. Tissue-specific expression of a splicing mutation in the *IKBKAP* gene causes familial dysautonomia. *Am J Hum Genet* 2001;68:598–605.
24. Felbor U, Sure U, Grimm T, Bertalanffy H. Genetics of cerebral cavernous angioma. *Zentralbl Neurochir* 2006;67:110–116.
25. Surucu O, Sure U, Gaetzner S, et al. Clinical impact of *CCM* mutation detection in familial cavernous angioma. *Childs Nerv Syst* 2006;22:1461–1464.
26. Rubinstein WS. Hereditary breast cancer in Jews. *Fam Cancer* 2004;3:249–257.