© 1999 Oxford University Press
Human Molecular Genetics, 1999, Vol. 8, No. 12 **2325–2333**

Mutations in the gene encoding KRIT1, a Krev-1/rap1a binding protein, cause cerebral cavernous malformations (*CCM1*)

Trilochan Sahoo[+], Eric W. Johnson[1,+], James W. Thomas[2], Peter M. Kuehl[3], Thomas L. Jones, Charles G. Dokken[4], Jeffrey W. Touchman[2], Carol J. Gallione, Shih-Queen Lee-Lin[2], Barry Kosofsky[5], Janice H. Kurth[1], David N. Louis[5], Gabrielle Mettler[6], Leslie Morrison[7], Antonio Gil-Nagel[8], Steven S. Rich[9], Joseph M. Zabramski[1], Mark S. Boguski[4], Eric D. Green[2] and Douglas A. Marchuk[§]

Department of Genetics, Duke University Medical Center, Durham, NC 27710, USA, [1]Department of Neurogenetics and Department of Neurosurgery, Barrow Neurological Institute, St Joseph's Hospital, Phoenix, AR 85013, USA, [2]Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA, [3]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA, [4]Center for Medical Genetics, Marshfield Clinic, Marshfield, WI 54449, USA, [5]Molecular Neuro-Oncology Laboratory, Massachusetts General Hospital, Charlestown, MA 02129, [6]Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, Ontario K1H 8L1, Canada, [7]Department of Neurology, University of New Mexico School of Medicine, Albuquerque, NM 87131, USA, [8]Programa de Epilepsia, Hospital Ruber Internacional, Madrid 28034, Spain and [9]Department of Public Health Sciences and Neurology, Bowman Gray School of Medicine, Winston-Salem, NC 27157, USA

Received September 2, 1999; Revised and Accepted September 8, 1999

Cerebral cavernous malformations (CCM) are congenital vascular anomalies of the brain that can cause significant neurological disabilities, including intractable seizures and hemorrhagic stroke. One locus for autosomal dominant CCM (*CCM1*) maps to chromosome 7q21–q22. Recombination events in linked family members define a critical region of ~2 Mb and a shared disease haplotype associated with a presumed founder effect in families of Mexican-American descent points to a potentially smaller region of interest. Using a genomic sequence-based positional cloning strategy, we have identified *KRIT1*, encoding a protein that interacts with the Krev-1/rap1a tumor suppressor, as the *CCM1* gene. Seven different *KRIT1* mutations have been identified in 23 distinct *CCM1* families. The identical mutation is present in 16 of 21 Mexican-American families analyzed, substantiating a founder effect in this population. Other Mexican-American and non-Hispanic Caucasian *CCM1* kindreds harbor other *KRIT1* mutations. Identification of a common Mexican-American mutation has potential clinical significance for presymptomatic diagnosis of CCM in this population. In addition, these data point to a key role for the Krev-1/rap1a signaling pathway in angiogenesis and cerebrovascular disease.

INTRODUCTION

Cerebral cavernous malformations (CCM) are congenital vascular anomalies of the brain, comprising 10–20% of all vascular malformations of the central nervous system, with an incidence in the general population of 0.1–0.5% (1,2). Histopathological studies show focal, well-circumscribed, thin-walled vascular spaces. The grossly dilated vascular spaces are lined by a single layer of endothelium, usually without any intervening neural parenchyma. They are also devoid of elastic tissue, smooth muscle and mature vessel wall elements (1–3).

Although cavernous malformations of the brain are congenital, they usually present clinically between the third and fifth decades of life (2,4,5). Therefore, many affected individuals are clinically quiescent in early life. Patients most often present with intractable seizures, intracerebral hemorrhage, recurrent headaches and focal neurological deficits (2,4,5). Hemorrhagic stroke, due to massive intracerebral hemorrhage of the vascular lesion, can be fatal and is often the first and only clinical presentation (6). CCM lesions therefore behave as intra-cerebral vascular tumors that present as space-occupying lesions, epileptogenic foci and sites for recurrent hemorrhage.

[+]These authors contributed equally to this work
[§]To whom correspondence should be addressed. Tel: +1 919 684 3290; Fax: +1 919 681 9193; Email: douglas.marchuk@duke.edu

Downloaded from https://academic.oup.com/hmg/article/8/12/2325/661862 by guest on 16 August 2022

	Family No.															
	10	11	13	14	15	16	200	300	500	600	800	900	1100	400	12	1200
Marker																
D7S630	218	202	220	218	218	218	218	216	218	218	218	216	218	218	202	198
D7S492	151	153	147	151	151	151	151	151	151	151	151	151	151	149	149	149
D7S2410	271	275	271	271	271	275	271	271	271	271	271	273	271	275	271	263
D7S627	248	249	256	249	249	249	249	249	248	248	249	249	249	241	245	244
D7S2408	178	178	178	178	178	178	178	178	178	178	178	178	178	178	178	186
D7S2409	217	217	223	217	217	217	217	217	217	217	217	217	217	221	225	227
D7S1813	137	137	128	137	137	137	137	137	137	137	137	137	137	131	140	128
D7S1789	137	137	137	137	137	137	137	137	137	137	137	137	137	137	131	137
Krit1 Mutation	A	A	A	A	A	A	A	A	A	A	A	A	A	B	C	D
085-411	137	137	137	137	137	137	137	137	137	137	137	137	137	137	135	137
085-222	170	170	170	170	170	170	170	170	170	170	170	170	170	170	170	174
D7S646	185	185	185	185	185	185	185	185	185	185	185	185	185	185	185	185
552-1166	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161
D7S689	129	129	125	129	129	129	129	129	129	129	127	131	129	127	129	127
D7S558	107	107	107	107	107	107	107	107	107	107	107	107	107	103	107	115
D7S657	246	246	260	246	246	246	246	246	246	246	246	246	246	260	250	258

Mutations: A=742C>T B=1089delA C=1314delT D=990T>A

Figure 1. Disease haplotype in Mexican-American CCM1 families. Numbers in the first column represent markers used within the CCM1 interval. Other columns indicate sizes in nucleotides of the disease-associated alleles in each of the families. *D7S2410* and *D7S689* mark proximal and distal boundaries of the CCM1 critical region, respectively. Physical location of the *KRIT1* gene is shown between *D7S1789* and *085-411*. The affected alleles within boxes indicate shared haplotypes within the interval. The region contained between markers *085-222* and *D7S689* comprises the shared disease haplotype in the Mexican-American families analyzed.

Both sporadic and familial forms of CCM have been identified. A familial form exhibits autosomal dominant inheritance with incomplete penetrance (OMIM 116860). However, a recent report suggests that up to 75% of the so-called sporadic cases are in reality members of an affected family, with asymptomatic vascular lesions in relatives masking the autosomal dominant segregation pattern (7). This suggests that the population incidence of familial CCM may be significantly underestimated.

Genetic linkage studies revealed a locus for CCM (*CCM1*) on chromosome 7q (8–10). Two additional loci, mapping to chromosomes 7p and 3q, respectively, have also been identified (11). Crossovers in multigenerational families define an ~4 cM critical region for *CCM1*, bordered by *D7S2410* and *D7S689* (12). The apparent high prevalence of the disease in families of Mexican-American heritage (1,13) and evidence for a common disease haplotype in affected kindreds from this population (12,14) suggest a possible founder effect.

Provided with a convincing genetic localization of the *CCM1* gene and with a strong genomic infrastructure for this region of chromosome 7, including a complete physical and rapidly evolving sequence map, we sought to identify the *CCM1* gene by a positional cloning strategy. Here we report the outcome of this effort, which has successfully identified *KRIT1* as the *CCM1* gene.

RESULTS

Refinement of the *CCM1* critical region

Two previously unpublished non-Hispanic families were ascertained. Family 2 is an American Caucasian family. While showing a clear autosomal dominant inheritance pattern for CCM, initial linkage analysis of this family was ambiguous. In part, this reflected the fact that a number of the family members were of uncertain affectation status (unpublished data). Family 55, a small Canadian Caucasian family, initially showed weak linkage to the *CCM1* locus (lod 0.74 at $\emptyset = 0$) and strongly negative linkage to both the *CCM2* and *CCM3* loci (11; unpublished data). Families 2 and 55 did not allow additional refinement of the established *CCM1* critical interval (12).

Our previous studies revealed a common disease haplotype among three apparently unrelated Mexican-American *CCM1* families (12). To utilize potential ancestral crossovers for further refinement of the critical interval, we collected 18 additional CCM families of Mexican-American descent. A series of new microsatellite-based genetic markers were developed from the available genomic sequence (see below) and used to analyze the available *CCM1* families. Haplotype construction for the Mexican-American families confirmed the shared disease haplotype (Fig. 1), consistent with the findings of others (15,16). Although most Mexican-American families shared alleles over a broad interval encompassing the entire critical region based on crossovers, two families (12 and 1200) appeared to share a smaller portion of the interval (Fig. 1), suggesting that the lineage of this genomic segment derives from ancestral crossovers. *D7S689*, a region of 450 kb, and the new marker *085-222* (Fig. 1) bordered the smallest common region. Thus, genetic mapping suggests two regions of interest: a larger ~2 Mb region based on recombinants in families and bordered by *D7S2410* and *D7S689* and an ~450 kb region contained within the telomeric end of the larger region bordered by *085-222* at the centromeric end (Fig. 1).

Physical mapping and sequencing of the *CCM1* critical region

We reported previously the construction of a complete yeast artificial chromosome (YAC)-based physical map of the

Downloaded from https://academic.oup.com/hmg/article/8/12/2325/661862 by guest on 16 August 2022

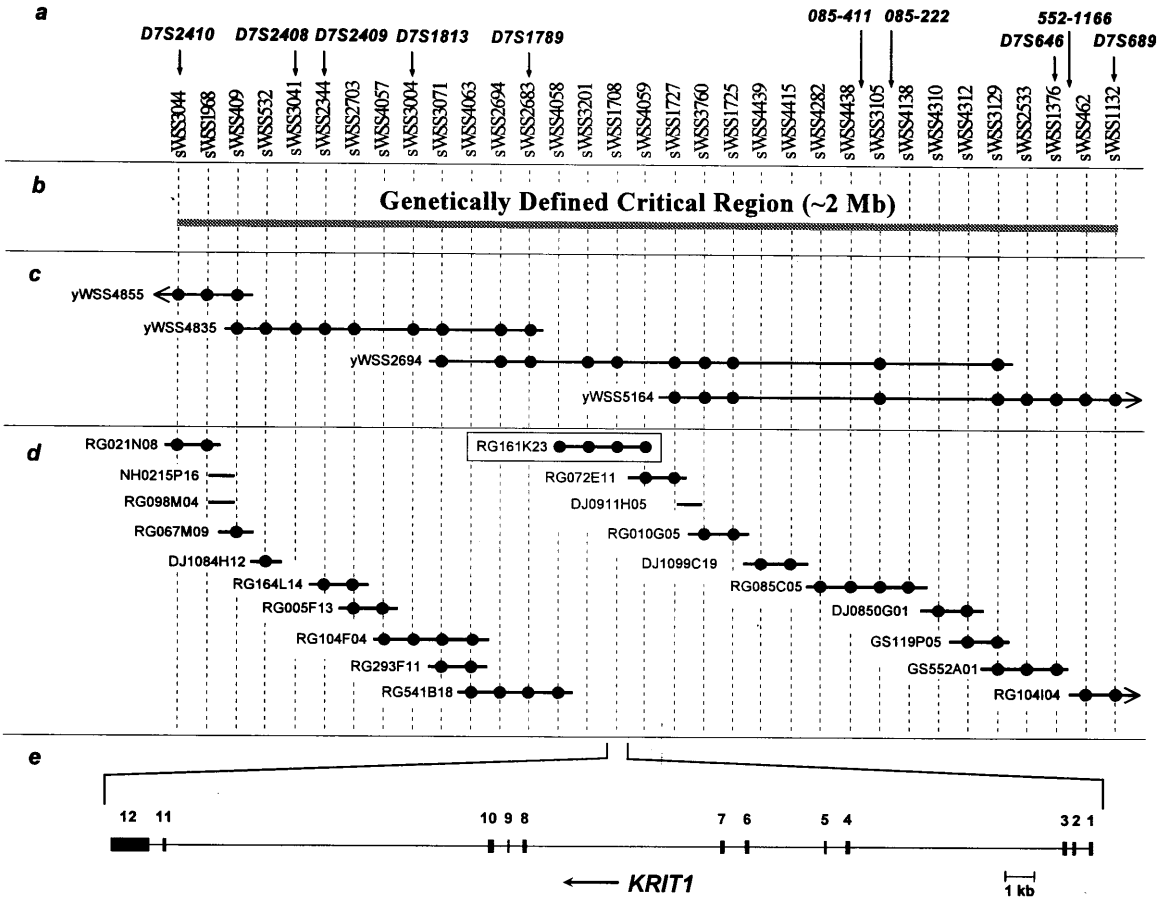

Figure 2. Physical map of the *CCM1* critical region. A clone-based physical map of the *CCM1* critical region, defined as the interval encompassing the genetic markers *D7S2410* and *D7S689* (**b**) (see text for additional details) is depicted, oriented with the centromeric end leftward and 7q telomeric end rightward. The relative positions of 33 mapped STSs (12,17; http://genome.nhgri.nih.gov/chr7) are indicated (**a**), with detailed information about the STSs available in GenBank. A minimal set of YACs spanning the interval is shown in (**c**), with the complete YAC contig map reported previously (12,17). The 20 BACs/PACs sequenced by the Washington University Genome Sequencing Center, which together provide complete contiguity across the interval, are depicted in (**d**). The presence of a closed circle indicates that the corresponding STS is contained within that clone (note that the YACs were not tested for the presence of some STSs, in which case no circle is present at the expected position in the YAC). The clone containing *KRIT1*, BAC RG161K23, is indicated with a box. The long range intron–exon structure of the 12 exon *KRIT1* gene is schematically illustrated in (**e**).

region of human chromosome 7 encompassing the *D7S2410–D7S689* interval (12). This map and its associated sequence-tagged sites (STSs) were used to develop a corresponding bacterial clone [bacterial artificial chromosome (BAC) and P1-derived artificial chromosome (PAC)] contig. The genetically defined *CCM1* critical region is ~2.0 Mb in size and represented by a minimal set of 20 overlapping BAC/PAC clones (Fig. 2). These were subjected to systematic sequencing at the Washington University Genome Sequencing Center.

Absence of mutations within the narrowest critical region

All known and computer predicted exons mapping within the narrowest (450 kb) region suggested by the apparent shared disease haplotype in Mexican-American families were analyzed. Although >120 potential exons were evaluated, no CCM-associated mutations were identified. We recognized that two families that might not share a common ancestral mutation primarily define the narrow region. Since the 7q telomeric boundary of the Mexican-American disease haplotype is the same as that identified by crossovers in Caucasian families, we extended our efforts in the centromeric direction.

KRIT1 mutations in Mexican-American families

One of the known genes within the larger candidate region is *KRIT1*, which encodes a novel Krev-1/rap1a binding protein (17). Alignment of the *KRIT1* cDNA sequence with the genomic sequence (BAC RG161K23) revealed that the ORF is distributed over 12 exons extending over 37.7 kb of genomic DNA (Fig. 2e). Probands from Mexican-American families 10 and 200 were found to harbor a single base transition (742C→T) in exon 6 of *KRIT1*, changing a Gln to a premature termination codon (Fig. 3a). Significantly, the proband from family 12 (the third Mexican-American family in the initial screening panel) does not harbor this mutation. This base substitution was not detected in 45 control individuals (90 chromosomes) of Mexican-American descent. Co-segregation of the mutation with disease status was assessed by conformation-sensitive gel electrophoresis (18). The mutation was present only in affected individuals or in unaffected family

Downloaded from https://academic.oup.com/hmg/article/8/12/2325/661862 by guest on 16 August 2022

members containing the disease haplotype (and therefore assumed to be asymptomatic mutation carriers). This mutation was identified in a majority (16 of 21) of Mexican-American families analyzed (families 11, 13, 14, 15, 16, 123, 152, 128, 300, 500, 600, 800, 900 and 1100).

Distinct mutations in *KRIT1* were identified in other Mexican-American families (Table 1). Family 400 carries a single base pair deletion (1089delA) in exon 8, resulting in a frameshift and premature termination. Family 12 carries a single base pair deletion (1314delT) in exon 10, leading to a frameshift and premature termination. Family 1200 carries a single base pair transversion (990T→A) in exon 8, converting a Tyr to a termination codon. Mutations in *KRIT1* were not identified in two families from this group: one Mexican-American and one collected entirely in Mexico. Although both families share some alleles for relevant markers within the *CCM1* critical region with the other Mexican-American families, neither family shares an extensive portion of the disease haplotype with the others (data not shown). Both families are too small to conclusively exhibit or exclude linkage to the *CCM1* locus.

Analysis of the genetic haplotype data for the Mexican-American families reveals a correlation between the *KRIT1* mutations and the common disease haplotype for the interval between markers *D7S1789* and *552-1166* (Fig. 1). All families with the 742C→T mutation share the common disease haplotype in this interval; for some families, the shared haplotype extends for additional markers mapping in both directions (Fig. 1). These data demonstrate that a founder mutation does exist in this population and that the majority of Mexican-American *CCM1* chromosomes are descended from a common ancestor.

Further analysis of haplotypes and *KRIT1* mutations reveals that family 400 appears to share the common disease haplotype between *D7S1789* and *552-1166*, a region that includes *KRIT1*, but harbors a unique *KRIT1* mutation (Fig. 1). This apparent sharing of the disease haplotype must reflect identity by state and not by descent. Families 12 and 1200 also harbor unique *KRIT1* mutations and share a decreasing portion of the common disease haplotype that does not include the region immediately surrounding *KRIT1*. This also likely reflects identity by state and not by descent. Importantly, inclusion of the latter two families in defining the most delimited common disease haplotype erroneously diverted the initial gene search towards the region just telomeric of *KRIT1*.

KRIT1 mutations in non-Hispanic Caucasian families

There is no evidence for a common disease haplotype in non-Hispanic Caucasian families (12; unpublished data). Not surprisingly, distinct *KRIT1* mutations were found in each of the four Caucasian *CCM1* families (Table 1). Family 20 carries the same mutation (1089delA in exon 8; Fig. 3b) as that present in family 400 of the Mexican-American kindreds, although the families are unrelated by disease haplotype and ancestry. The mutation consists of the loss of a single A in a run of six A residues, which may explain its independent origin. Families 1 and 2 harbor different splice site mutations. Family 1 exhibits a single base pair transition at the exon–intron border in intron 4 (IVS4+1G→A) that disrupts the invariant splice donor site from gt to at (Fig. 3c). Family 2 exhibits a single base pair tran-

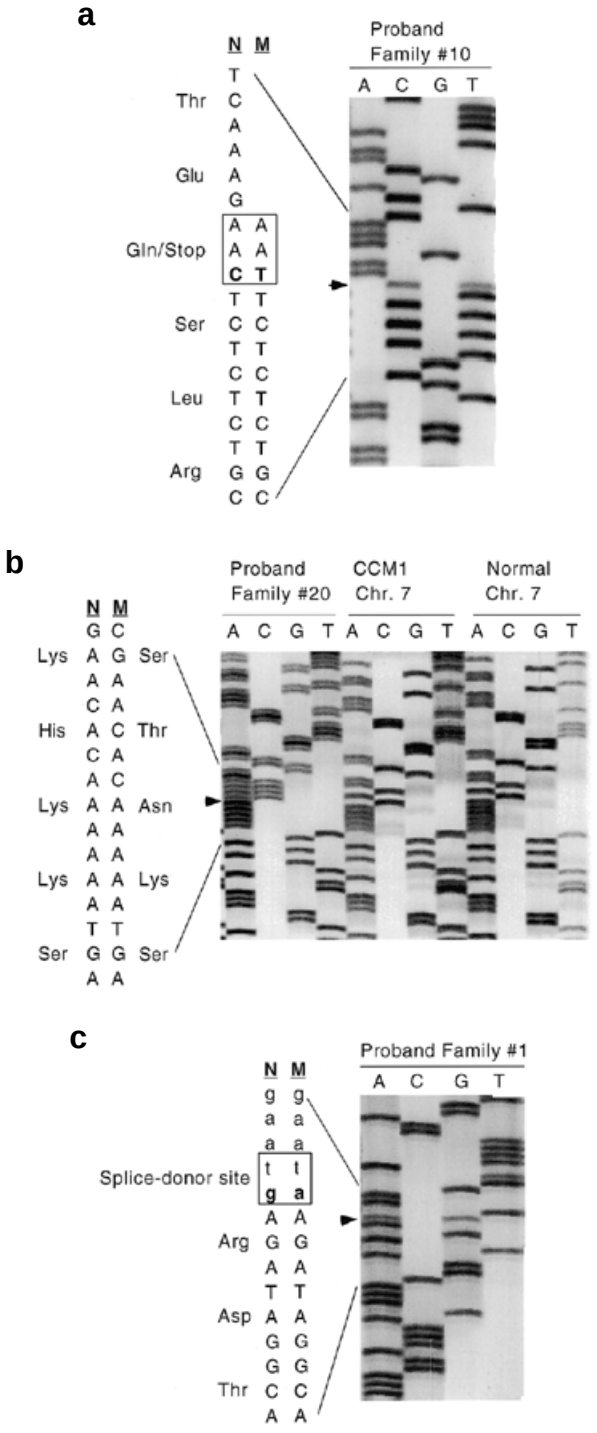

Figure 3. Three *KRIT1* mutations in CCM1 families. (**a**) Genomic sequence from an affected member (proband) of family 10 showing the common Mexican-American mutation. This is a C→T transition (arrowhead) in exon 6 changing a Gln to a premature termination codon and was seen in 16 of the 21 Mexican-American families analyzed. (**b**) Genomic sequence from an affected member (proband) of family 20 showing a single base deletion (1089delA) in exon 8 leading to a frameshift. The sequences of the individual affected and unaffected chromosomes 7 isolated from somatic cell hybrids (see Materials and Methods) from the same individual show the single base deletion (arrowhead) only present in the affected chromosome and absent in the normal. (**c**) Genomic sequence from an affected member (proband) of family 1 showing a single base G→A transition (arrowhead) that alters the invariant splice donor site in intron 4. N, normal; M, mutant.

Downloaded from https://academic.oup.com/hmg/article/8/12/2325/661862 by guest on 16 August 2022

Table 1. *KRIT1* mutations in CCM1 families

Family	Mutation	Predicted effect
Mexican-American		
10, 11, 13, 14, 15, 16, 123, 128, 152, 200, 300, 500, 600, 800, 900 and 1100	742C→T	Premature termination (Q248X)
12	1314delT	Frameshift in codon 438
400	1089delA	Frameshift in codon 363
1200	990T→A	Premature termination (Y330X)
Caucasian		
1	IVS4+1G→A	Invariant splice donor site altered
2	IVS8-2A→G	Invariant splice acceptor site altered
20	1089delA	Frameshift in codon 363
55	281C→G	Premature termination (S94X)

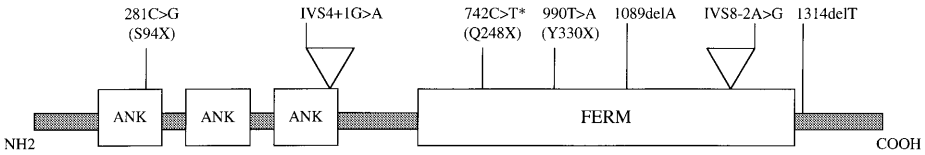

Figure 4. Putative structural motifs in KRIT1 as predicted by the comparative sequence analysis program SMART (Simple Modular Architecture Research Tool v.3.0). The ankyrin repeats are distributed from amino acid residue 80 to 177 and the FERM domain from residue 209 to 433. The locations of the mutations with respect to the structural domains of KRIT1 are represented (742C→T represents the common Mexican-American mutation).

sition at the intron–exon border in intron 8 (IVS8-2A→G) that disrupts the invariant splice acceptor site from ag to gg. A single base pair transversion (281C→G in exon 3) in family 55 changes a Ser to a termination codon.

In total, seven distinct *KRIT1* mutations were found among the *CCM1* kindreds examined (Table 1). Mutations fall into one of three classes: frameshifts, premature chain terminations and splice site mutations (Fig. 4). The combined data suggest that these are all loss-of-function alleles. Thus far, no missense mutations have been identified.

Computational analysis of the KRIT1 protein

A BLASTP search of *KRIT1* against the non-redundant collection of GenBank coding sequences reveals many significant matches to proteins containing either the ankyrin domain (19) or the so-called FERM domain (20,21). However, the two best matches (apart from the match of *KRIT1* with itself) are to two uncharacterized genes. Residues 35–312 of KRIT1 align with the protein encoded by the predicted nematode gene CAB03514 (Z81143) at an *E* value of ~10^{-20} and residues 252–472 align with a conceptual translation from a nematode expressed sequence tag (EST) CAB02995 (Z81069) at an *E* value of ~10^{-13}.

To better understand the domain organization of KRIT1, a search of the SMART database of protein domains was conducted. This revealed that KRIT1 contains at least three ankyrin domains, beginning at residue 80 and ending at residue 177, and one FERM (or B41) domain, spanning residues 209–433 (Fig. 4). Furthermore, this domain composition appears to be unique to KRIT1, with no other proteins in the SMART database containing both ankyrin and FERM domains. We note that Serebriiskii *et al.* (17) originally reported that KRIT1 has four ankyrin domains, but only three are detected using the strict SMART criteria.

DISCUSSION

The positional cloning-based identification of *KRIT1* as the *CCM1* gene was facilitated by the availability of the complete genomic sequence of the region (22). The sequence provided new microsatellite markers for refined haplotype construction, detailed information about known genes and clues about computer predicted expressed sequences. Since few known genes mapped to the interval, we relied on gene prediction programs to identify putative exons for mutation screening. Of note, although the *CCM1* locus ultimately proved to be a previously known gene, all 12 *KRIT1* exons were identified by at least one of the set of gene prediction programs utilized (23).

The strategy to concentrate our efforts on families of Mexican-American descent for the purpose of mapping ancestral crossovers proved to be sound, although the assumption that all such families were related to a common ancestor proved to be incorrect. The inappropriate inclusion of putative ancestral recombination events suggested by two families diverted the initial search ~200 kb telomeric of *KRIT1*. These families falsely appeared to share a reduced portion of the common disease haplotype with the majority of the other Mexican-American *CCM1* families. Some of the microsatellite markers used for determining disease haplotypes were of low heterozygosity, with alleles for some of the markers being relatively common in the Mexican-American population. Linkage disequilibrium is an alternative approach for fine mapping of disease genes when a founder effect is suspected, with this approach less influenced by the haplotypes of one or two specific families. Such an approach was used by other investi-

Downloaded from https://academic.oup.com/hmg/article/8/12/2325/661862 by guest on 16 August 2022

gators searching for the *CCM1* gene, with their data also predicting that it resided a few cM telomeric of *KRIT1* (15). Nonetheless, the great majority of *CCM1* alleles in Mexican-American families consist of the identical 742C→T (Q248X) *KRIT1* mutation. This observation has potential clinical significance for presymptomatic diagnosis of CCM in the Mexican-American population.

A central question of great interest is the function of the *KRIT1*-encoded protein. Clues about KRIT1 function come partly from its initial isolation as a binding partner of Krev-1/rap1A in the yeast interaction trap assay (17). Krev-1/rap1a was originally isolated due to its ability to partially revert the transformed phenotype of Kirsten sarcoma virus-transformed cells and later shown to be a member of the Ras family of GTPases with tumor suppressing activity for the *ras* oncogenes (24–26). The protein residues in both Krev-1/rap1a and Ras that are responsible for interaction with effectors are identical, indicating that these two proteins mediate opposing effects through common downstream partners like Raf-1, B-Raf, and RalGDS (26–31). The homology between the Krev-1/rap1a protein and Ras, its ability to interact with most cellular effectors of Ras and its ability to revert the oncogenic effect of Ras *in vivo* suggests that Krev-1/rap1a behaves as a tumor suppressor in the cell, possibly by sequestering Ras–GAP and acting as a competitive inhibitor of Ras–GAP interaction. Whether the *KRIT1* mutations activate or inactivate Krev-1/rap1a signaling remains to be determined. Krev-1/rap1a homologs in *Drosophila* (Ras3) and yeast (RSR/BUD1) also regulate GTPase signal transduction cascades controlling cellular differentiation, morphology, polarity and cytoskeletal structure (32,33). Some of these same regulatory circuits for these fundamental characteristics of the cell may be conserved in mammalian species.

Other potential clues about KRIT1 function can be gained by analyzing the structural motifs of the protein. The ~33 residue ankyrin repeat has been described in >400 proteins with diverse functions and, in a number of cases, is known to mediate protein–protein interactions (34,35). The ~225 residue FERM domain is present in a superfamily of proteins, including the band 4.1 protein originally isolated from human erythrocyte ghosts and the ezrin/radixin/moesin (ERM) protein family that regulates the interaction of plasma membranes with the actin cytoskeleton. However, it should be noted that most FERM-containing proteins possess an N-terminal FERM domain, whereas the FERM domain of KRIT1 resides towards the C-terminus (21). Furthermore, the combination of a FERM domain with ankyrin repeats is apparently unique.

The expression profiles of *KRIT1* and Krev-1/rap1a only hint at a role in vascular development. Krev-1/rap1a is expressed in a number of tissues that includes the small blood vessels of many organs (36). Previously described northern analysis of *KRIT1* showed weak expression in heart, muscle and brain, but not in a number of other tissues (17). Our analysis shows varying levels of expression in heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas (data not shown), suggesting that *KRIT1* expression is not restricted to vascular or cerebrovascular tissue. The *KRIT1* expression pattern during development has yet to be reported. As CCMs may be present at birth, they may develop *in utero*, decades before any clinical presentation. A better understanding of *KRIT1* expression in a developmental and anatomical context is required to establish why the CCM phenotype is restricted to the vasculature of the brain.

A working model would be that cerebral cavernous malformations are benign vascular tumors that develop due to an alteration in an important growth control pathway(s) involving Krev-1/rap1a. In support of this, the Krev-1/rap1a pathway is also altered in tuberous sclerosis type 2 (TSC2), an autosomal dominant condition characterized by benign neurocutaneous tumors. Significantly, TSC2, like CCM, is often associated with seizures. The TSC2 gene product, tuberin, functions as a tumor suppressor protein by acting as a GTPase-activating protein for Krev-1/rap1a (37). Somatic inactivation or loss of heterozygosity of the wild-type tuberin allele in TSC2-associated tumors leads to constitutive activation of Krev-1/rap1a and to unregulated growth (38,39). The focal nature of CCMs suggests that these vascular tumors may develop due to loss of the wild-type *KRIT1* allele in the developing cerebral vasculature. A two-hit model, with a central tumor suppressor function of KRIT1, would be similar to the mechanism of tuberin inactivation in TSC2-associated tumors. Alternatively, the etiology of vascular lesion development may relate to a mechanical or environmental trigger in the cerebral vasculature, coupled with the underlying genetic defect in *KRIT1*. Whatever the mechanism, the data reported here point to a key role for the KRIT1 and the Krev-1/rap1a pathway in the complex process of angiogenesis and, potentially, in other forms of cerebrovascular disease.

MATERIALS AND METHODS

Families and DNA samples

We ascertained and collected DNA samples from 25 multigeneration families with CCM under appropriate IRB-approved informed consent. Twenty-one of these are of Mexican-American descent and four are Caucasian. Clinical details, pedigrees and genetic linkage data for families 1, 20, 10, 200 and 300 have been reported previously (12). Two previously unpublished Caucasian families (2 and 55) are included in these data, as described above. Eighteen new unpublished Mexican-American families were collected ranging in size from seven members (four affected) to a single affected individual. One of these families was collected entirely in Mexico. Diagnosis of most patients with CCM was based on detailed neurological evaluation that included clinical features, familial occurrence, MRI examination and, when possible, post-surgery histopathology.

DNA was purified from whole blood using established methods (Puregene; Gentra Systems, Research Triangle Park, NC). Each sample was assigned a consecutive sample number, entered into a secure database and identified only by that number for all subsequent analyses. Individual affected and unaffected chromosomes 7 were isolated from patient monocytes from a number of affected individuals and maintained as somatic cell hybrids in a rodent cell line (RJK88). DNA from these cell lines was used for exon amplification and mutation analysis when available.

Downloaded from https://academic.oup.com/hmg/article/8/12/2325/661862 by guest on 16 August 2022

Table 2. Intron–exon organization of the *KRIT1* coding region[a]

Exon/intron	Exon length (bp)	Starting position in cDNA	Acceptor splice site	Donor splice site
1	108	1		CAAATCGGgtaagagt
2	116	109	ctttctagGTAGATAA	GAAGACAAgtactgttt
3	144	225	atgattagGGAACGAC	GCATGCTGgtaaatgg
4	157	369	ttatacagGTATGGAA	CGGATAGAgtaagtta
5	108	526	atttctagCATATAAC	CAAACCAgtaagaat
6	157	634	tcttgtagTATGAAAA	AACCTCAgtaagaaa
7	152	791	tttttcagGCCTTCAA	AAAAACAGgtttgctt
8	167	943	cctttaagATTGAAGA	TTCCTAAAgtaagtat
9	88	1110	gcttacagTGAAGAAA	AATACAAGgtaagctg
10	207	1198	gtttttagAATCTCAG	AAACTAAGgtagattt
11	117	1405	tgttgcagGCTTTACT	CAAAACAGgtaagtat
12	66	1522	ctttgtagGCTGGTCT	

[a]The intronic and exonic sequence are in lower case and upper case, respectively.

Table 3. Primers for PCR amplification and sequencing of *KRIT1* exons

Exon	Forward primer (sense)[a]	Reverse primer (antisense)[a]	Product size
1	TCACTTTATTCCAGCTTTATTCC	AAAACGTCTTTTAAATCAGAGC	260
2	TGACAAAGCTCTTAATGGGT	GACTACAATGCATACAAATTGC	261
3	AAACAATTTTTACAGTCCTGTTG	AGAACAGTCTTGAAAAGAAGGA	280
4	CATTTCAGATGATCTTTTTAGG	TGTCATTACTTGTTATTCACTGCT	286
5	ATTGGATGACATTTTCCCTT	AGCCATCTAATCGTCTTTCC	281
6	AGCACATGAAGTTGAAGGAA	CCCAAAAAGGAATAATGAGG	292
7	GAAGTGCAGACAGTTTAATACAAA	CTCAACAGATTTTGTGCATTT	293
8	GCTTTTTCTTTTCCCATATT	TAGCACAAGACCATGCATAA	283
9	CGTTACTGAAAGCCATTTGT	CAGGACTATAAATTTAATCTACCTCTG	281
10	CAATGGTACATTTTCCTTTCA	AGGTTGGTACTGTTGTTTTAACT	329
11	CTGAACTATTATATTTAGAGCAGACA	CACAATAGTTTATGAAGTCCAAA	261
12	CCCAATGTCATGAATTTCC	GCTCGGCCAAAAGTAATA	233

[a]All primers read 5′→3′. Only forward primers were used for direct sequencing of the exons.

Genetic analyses

New families were genotyped as described (40). New microsatellite markers (*085-411*, *085-222*, *552-1166* and others not shown) were generated from genomic sequence generated in the candidate interval. The primer sequences for the newer markers are: *085-411*, forward, cactttaaaagatcacagaatctc, reverse, aggcacagctgagttatgat; *085-222*, forward, catctactagaatgcccagc, reverse, caaatggttccttcctttct; *552-1166*, forward, gtctgagcagttaaaatggc, reverse, gttgagtgtccacctgagat. Two-point lod scores were calculated for the largest Caucasian families using LINKAGE v.5.03b (41). Haplotypes were generated both manually and using the Cyrillic software package (Cherwell Scientific, Oxford, UK).

Construction of a bacterial clone-based physical map

Human DNA-containing BAC (42) and PAC clones (43) were isolated, analyzed and assembled into contigs as described previously (44–46). Specifically, this involved hybridization (47) and PCR screening (48) of available human genomic libraries, PCR analysis of isolated clones to establish their STS content and restriction enzyme digest-based fingerprint analysis (49). For contig expansion, sequence was obtained from the insert ends of clones residing at contig ends and used to isolate additional BACs/PACs. Clones are named with a prefix reflecting their library of origin: RG, Research Genetics human BAC library (Huntsville, AL; see http://www.resgen.com); GS, Genome Systems human BAC library (St Louis, MO; see http://www.genomesystems.com); DJ, Roswell Park Cancer Institute PAC library (see http://bacpac.med.buffalo.edu).

Identification of candidate genes in the *CCM1* critical region

Several strategies were used to identify genes within the *CCM1* critical region. The genomic sequence of the critical region (Fig. 2a) was analyzed for the presence of known genes or matching ESTs. This was supplemented by the use of the gene prediction programs GRAIL II (50), GENESCAN (51) and MZEF (52). Data were assimilated and organized using a computational approach for analyzing and managing genomic sequence (23).

CCM1 mutation panel

In total, 25 different CCM families were used for mutation analysis, 21 of which were of Mexican-American descent and four non-Hispanic Caucasian. An initial smaller screening panel included the probands of the four non-Hispanic *CCM1*

Downloaded from https://academic.oup.com/hmg/article/8/12/2325/661862 by guest on 16 August 2022

families, expected to harbor distinct mutations, and those from only three of the 21 Mexican-American families, assumed to harbor the same mutation. The largest of both groups of families have been shown to be linked to the *CCM1* locus (8,10,12); others were smaller but consistent with linkage, while a few of the Mexican-American kindreds were too small for genetic linkage analysis. Each individual exon (known or computer predicted) was amplified by PCR and subjected to direct DNA sequencing.

PCR amplification of exons

KRIT1 exon-containing fragments were PCR amplified from genomic DNA (Table 2). Primers were designed from the genomic sequence flanking the exons using the program PRIMER v.3.0 (http://www-genome.wi.mit.edu) ensuring that the amplified project contained putative splice junctions (Table 3). All the fragments were amplified using the following amplification cycle: initial denaturation at 95°C for 5 min followed by 35 cycles of 94°C for 30 s, 56°C for 60 s and 72°C for 60 s.

Mutation detection and analysis

PCR-amplified products were separated in low melting point agarose gels and specific fragments were excised, crushed, freeze–thawed and purified by spinning and retrieving the supernatant. Direct DNA sequencing of the amplified genomic DNA fragment was carried out with a ThermoSequenase Cycle Sequencing kit (Amersham Life Science, Cleveland, OH) utilizing ^{33}P terminator labeling. Sequencing reaction products were resolved on denaturing polyacrylamide gels and loaded so as to have all the individual ddATP, ddCTP, ddGTP or ddTTP lanes together in sets. Segregation of the mutation with affected phenotype was carried out by either PCR amplification and sequencing of the appropriate exon for each family or by conformation-sensitive gel electrophoresis of the amplified fragments in high resolution 15% polyacylamide gels containing ethylene glycol (10%) and formamide (15%) (18).

Computer modeling of *KRIT1* structure

The KRIT1 protein sequence (accession no. NP_004903; 529 residues) was retrieved from www.ncbi.nlm.nih.gov/Entrez/ and used for all subsequent analyses. Sequence similarity searching was carried out using the BLAST family of programs (53). Domain analysis was performed using SMART (the Simple Modular Architecture Research Tool) available at http://coot.embl-heidelberg.de/SMART (54,55).

GenBank accession numbers

KRIT1 mRNA, U90268; BAC, RG161K23 and AC000120.

ACKNOWLEDGEMENTS

We are grateful to the patients and their families for their participation in this study and the Washington University Genome Sequencing Center for their generation and open dissemination of human genomic sequence data. We also thank Marco Marra, John McPherson, Brad Barbazuk and Tammy Kucaba for collaborative assistance in bacterial clone-based physical mapping and Blaine Hart (University of New Mexico) for neuroimaging and diagnostic confirmation. This work was supported by American Heart Association Scientist Development Grant 9630236N to E.W.J. and Grant-in-Aid 95006970 to D.A.M. D.A.M. is an Established Investigator of the American Heart Association.

REFERENCES

1. Rigamonti, D., Hadley, M.N., Drayer, B.P., Johnson, P.C., Hoenig-Rigamonti, K., Knight, J.T. and Spetzler, R.F. (1988) Cerebral cavernous malformations. Incidence and familial occurrence. *N. Engl. J. Med.*, **319**, 343–347.
2. Gil-Nagel, A., Wilcox, K.J., Stewart, J.M., Anderson, V.E., Leppik, I.E. and Rich, S.S. (1995) Familial cerebral cavernous angioma: clinical analysis of a family and phenotypic classification. *Epilepsy Res.*, **21**, 27–36.
3. Tomlinson, F.H., Houser, O.W., Scheithauer, B.W., Sundt, T.M.Jr, Okazaki, H. and Parisi, J.E. (1994) Angiographically occult vascular malformations: a correlative study of features on magnetic resonance imaging and histological examination. *Neurosurgery*, **34**, 792–799 (Discussion 799–800).
4. Zabramski, J.M., Wascher, T.M., Spetzler, R.F., Johnson, B., Golfinos, J., Drayer, B.P., Brown, B., Rigamonti, D. and Brown, G. (1994) The natural history of familial cavernous malformations: results of an ongoing study. *J. Neurosurg.*, **80**, 422–432.
5. Kondziolka, D., Lunsford, L.D. and Kestle, J.R. (1995) The natural history of cerebral cavernous malformations. *J. Neurosurg.*, **83**, 820–824.
6. Maraire, J.N. and Awad, I.A. (1995) Intracranial cavernous malformations: lesion behavior and management strategies. *Neurosurgery*, **37**, 591–605.
7. Labauge, P., Laberge, S., Brunereau, L., Levy, C., Tournier-Lasserve, E. and Societe Francaise de Neurochirurgie (1998) Hereditary cerebral cavernous angiomas: clinical and genetic features in 57 French families. *Lancet*, **352**, 1892–1897.
8. Dubovsky, J., Zabramski, J.M., Kurth, J., Spetzler, R.F., Rich, S.S., Orr, H.T. and Weber, J.L. (1995) A gene responsible for cavernous malformations of the brain maps to chromosome 7q. *Hum. Mol. Genet.*, **4**, 453–458.
9. Gunel, M., Awad, I.A., Anson, J. and Lifton, R.P. (1995) Mapping a gene causing cerebral cavernous malformation to 7q11.2–q21. *Proc. Natl Acad. Sci. USA*, **92**, 6620–6624.
10. Marchuk, D.A., Gallione, C.J., Morrison, L.A., Clericuzio, C.L., Hart, B.L., Kosofsky, B.E., Louis, D.N., Gusella, J.F., Davis, L.E. and Prenger, V.L. (1995) A locus for cerebral cavernous malformations maps to chromosome 7q in two families. *Genomics*, **28**, 311–314.
11. Craig, H.D., Gunel, M., Cepeda, O., Johnson, E.W., Ptacek, L., Steinberg, G.K., Ogilvy, C.S., Berg, M.J., Crawford, S.C., Scott, R.M., Steichen-Gersdorf, E., Sabroe, R., Kennedy, C.T.C., Mettler, G., Beis, M.J., Fryer, A., Awad, I.A. and Lifton, R.P. (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.
12. Johnson, E.W., Iyer, L.M., Rich, S.S., Orr, H.T., Gil-Nagel, A., Kurth, J.H., Zabramski, J.M., Marchuk, D.A., Weissenbach, J., Clericuzio, C.L., Davis, L.E., Hart, B.L., Gusella, J.F., Kosofsky, B.E., Louis, D.N., Morrison, L.A., Green, E.D. and Weber, J.L. (1995) 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.*, **5**, 368–380.
13. Kattapong, V.J., Hart, B.L. and Davis, L.E. (1995) Familial cerebral cavernous angiomas: clinical and radiologic studies. *Neurology*, **45**, 492–497.
14. Gunel, M., Awad, I.A., Finberg, K., Steinberg, G.K., Craig, H.D., Cepeda, O., Nelson-Williams, C. and Lifton, R.P. (1996) Genetic heterogeneity of inherited cerebral cavernous malformation. *Neurosurgery*, **38**, 1265–1271.
15. Gunel, M., Awad, I.A., Finberg, K., Anson, J.A., Steinberg, G.K., Batjer, H.H., Kopitnik, T.A., Morrison, L., Giannotta, S.L., Nelson-Williams, C. and Lifton, R.P. (1996) A founder mutation as a cause of cerebral cavernous malformation in Hispanic Americans. *N. Engl. J. Med.*, **334**, 946–951.
16. Polymeropoulos, M.H., Hurko, O., Hsu, F., Rubenstein, J., Basnet, S., Lane, K., Dietz, H., Spetzler, R.F. and Rigamonti, D. (1997) Linkage of the locus for cerebral cavernous hemangiomas to human chromosome 7q in four families of Mexican-American descent. *Neurology*, **48**, 752–757.
17. Serebriiskii, I., Estojak, J., Sonoda, G., Testa, J.R. and Golemis, E.A. (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.

Downloaded from https://academic.oup.com/hmg/article/8/12/2325/661862 by guest on 16 August 2022

18. Korkko, J., Annunen, S., Pihlajamaa, T., Prockop, D.J. and Ala-Kokko, L. (1998) Conformation sensitive gel electrophoresis for simple and accurate detection of mutations: comparison with denaturing gradient gel electrophoresis and nucleotide sequencing. *Proc. Natl Acad. Sci. USA*, **95**, 1681–1685.
19. Sedgwick, S.G. and Smerdon, S.J. (1999) The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem. Sci.*, **24**, 311–316.
20. Tsukita, S. and Yonemura, S. (1997) ERM proteins: head-to-tail regulation of actin–plasma membrane interaction. *Trends Biochem. Sci.*, **22**, 53–58.
21. Chishti, A.H., Kim, A.C., Marfatia, S.M., Lutchman, M., Hanspal, M., Jindal, H., Liu, S.C., Low, P.S., Rouleau, G.A., Mohandas, N., Chasis, J.A., Conboy, J.G., Gascard, P., Takakuwa, Y., Huang, S.C., Benz, E.J.Jr, Bretscher, A., Fehon, R.G., Gusella, J.F., Ramesh, V., Solomon, F., Marchesi, V.T., Tsukita, S., Hoover, K.B. *et al.* (1998) The FERM domain: a unique module involved in the linkage of cytoplasmic proteins to the membrane [letter]. *Trends Biochem. Sci.*, **23**, 281–282.
22. Bouffard, G.G., Idol, J.R., Braden, V.V., Iyer, L.M., Cunningham, A.F., Weintraub, L.A., Touchman, J.W., Mohr-Tidwell, R.M., Peluso, D.C., Fulton, R.S., Ueltzen, M.S., Weissenbach, J., Magness, C.L. and Green, E.D. (1997) A physical map of human chromosome 7: an integrated YAC contig map with average STS spacing of 79 kb. *Genome Res.*, **7**, 673–692.
23. Kuehl, P.M., Weisemann, J.M., Touchman, J.W., Green, E.D. and Boguski, M.S. (1999) An effective approach for analyzing 'prefinished' genomic sequence data. *Genome Res.*, **9**, 189–194.
24. Kitayama, H., Sugimoto, Y., Matsuzaki, T., Ikawa, Y. and Noda, M. (1989) A ras-related gene with transformation suppressor activity. *Cell*, **56**, 77–84.
25. Frech, M., John, J., Pizon, V., Chardin, P., Tavitian, A., Clark, R., McCormick, F. and Wittinghofer, A. (1990) Inhibition of GTPase activating protein stimulation of Ras-p21 GTPase by the Krev-1 gene product. *Science*, **249**, 169–171.
26. Zhang, K., Noda, M., Vass, W.C., Papageorge, A.G. and Lowy, D.R. (1990) Identification of small clusters of divergent amino acids that mediate the opposing effects of ras and Krev-1. *Science*, **249**, 162–165.
27. Marshall, M.S., Davis, L.J., Keys, R.D., Mosser, S.D., Hill, W.S., Scolnick, E.M. and Gibbs, J.B. (1991) Identification of amino acid residues required for Ras p21 target activation. *Mol. Cell. Biol.*, **11**, 3997–4004.
28. Maruta, H., Holden, J., Sizeland, A. and D'Abaco, G. (1991) The residues of Ras and Rap proteins that determine their GAP specificities. *J. Biol. Chem.*, **266**, 11661–11668.
29. Herrmann, C., Horn, G., Spaargaren, M. and Wittinghofer, A. (1996) Differential interaction of the ras family GTP-binding proteins H-Ras, Rap1A and R-Ras with the putative effector molecules Raf kinase and Ral-guanine nucleotide exchange factor. *J. Biol. Chem.*, **271**, 6794–6800.
30. Urano, T., Emkey, R. and Feig, L.A. (1996) Ral-GTPases mediate a distinct downstream signaling pathway from Ras that facilitates cellular transformation. *EMBO J.*, **15**, 810–816.
31. Hu, C.D., Kariya, K., Kotani, G., Shirouzu, M., Yokoyama, S. and Kataoka, T. (1997) Coassociation of Rap1A and Ha-Ras with Raf-1 N-terminal region interferes with ras-dependent activation of Raf-1. *J. Biol. Chem.*, **272**, 11702–11705.
32. Pizon, V., Chardin, P., Lerosey, I., Olofsson, B. and Tavitian, A. (1988) Human cDNAs rap1 and rap2 homologous to the *Drosophila* gene Dras3 encode proteins closely related to ras in the 'effector' region. *Oncogene*, **3**, 201–204.
33. Ruggieri, R., Bender, A., Matsui, Y., Powers, S., Takai, Y., Pringle, J.R. and Matsumoto, K. (1992) RSR1, a ras-like gene homologous to Krev-1 (smg21A/rap1A): role in the development of cell polarity and interactions with the Ras pathway in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **12**, 758–766.
34. Sedgwick, S.G. and Smerdon, S.J. (1999) The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem. Sci.*, **24**, 311–316.
35. Bork, P. (1993) Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? *Proteins*, **17**, 363–374.
36. Wienecke, R., Maize, J.C.Jr, Reed, J.A., de Gunzburg, J., Yeung, R.S. and DeClue, J.E. (1997) Expression of the TSC2 product tuberin and its target Rap1 in normal human tissues. *Am. J. Pathol.*, **150**, 43–50.
37. Wienecke, R., Konig, A. and DeClue, J.E. (1995) Identification of tuberin, the tuberous sclerosis-2 product. Tuberin possesses specific Rap1GAP activity. *J. Biol. Chem.*, **270**, 16409–16414.
38. Maheshwar, M.M., Cheadle, J.P., Jones, A.C., Myring, J., Fryer, A.E., Harris, P.C. and Sampson, J.R. (1997) The GAP-related domain of tuberin, the product of the TSC2 gene, is a target for missense mutations in tuberous sclerosis. *Hum. Mol. Genet.*, **6**, 1991–1996.
39. Soucek, T., Pusch, O., Wienecke, R., DeClue, J.E. and Hengstschlager, M. (1997) Role of the tuberous sclerosis gene-2 product in cell cycle control. Loss of the tuberous sclerosis gene-2 induces quiescent cells to enter S phase. *J. Biol. Chem.*, **272**, 29301–29308.
40. Weber, J.L., Wang, Z., Hansen, K., Stephenson, M., Kappel, C., Salzman, S., Wilkie, P.J., Keats, B., Dracopoli, N.C., Brandriff, B.F. *et al*. (1993) Evidence for human meiotic recombination interference obtained through construction of a short tandem repeat-polymorphism linkage map of chromosome 19. *Am. J. Hum. Genet.*, **53**, 1079–1095.
41. Lathrop, G.M. and Lalouel, J.M. (1984) Easy calculations of lod scores and genetic risks on small computers. *Am. J. Hum. Genet.*, **36**, 460–465.
42. Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y. and Simon, M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl Acad. Sci. USA*, **89**, 8794–8797.
43. Ioannou, P.A., Amemiya, C.T., Garnes, J., Kroisel, P.M., Shizuya, H., Chen, C., Batzer, M.A. and de Jong, P.J. (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet.*, **6**, 84–89.
44. Everett, L.A., Glaser, B., Beck, J.C., Idol, J.R., Buchs, A., Heyman, M., Adawi, F., Hazani, E., Nassir, E., Baxevanis, A.D., Sheffield, V.C. and Green, E.D. (1997) Pendred syndrome is caused by mutations in a putative sulphate transporter gene (PDS). *Nature Genet.*, **17**, 411–422.
45. Ellsworth, R.E., Ionasescu, V., Searby, C., Sheffield, V.C., Braden, V.V., Kucaba, T.A., McPherson, J.D., Marra, M.A. and Green, E.D. (1999) The CMT2D locus: refined genetic position and construction of a bacterial clone-based physical map. *Genome Res.*, **9**, 568–574.
46. Zenklusen, J.C., Weintraub, L.A. and Green, E.D. (1999) Construction of a high-resolution physical map of the ~1-Mb region of human chromosome 7q31.1–q31.2 harboring a putative tumor suppressor gene. *Neoplasia*, **1**, 16–22.
47. Vollrath, D. (1999) DNA markers for physical mapping. In Birren, B., Green, E.D., Hieter, P., Klapholz, S., Myers, R.M., Riethman, H. and Roskams, J. (eds), *Genome Analysis: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Vol. 4, pp. 187–215.
48. Green, E.D. and Olson, M.V. (1990) Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. *Proc. Natl Acad. Sci. USA*, **87**, 1213–1217.
49. Marra, M.A., Kucaba, T.A., Dietrich, N.L., Green, E.D., Brownstein, B., Wilson, R.K., McDonald, K.M., Hillier, L.W., McPherson, J.D. and Waterston, R.H. (1997) High throughput fingerprint analysis of large-insert clones. *Genome Res.*, **7**, 1072–1084.
50. Xu, Y., Mural, R., Shah, M. and Uberbacher, E. (1994) Recognizing exons in genomic sequence using GRAIL II. *Genet. Eng.*, **16**, 241–253.
51. Burge, C. and Karlin, S. (1997) Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.*, **268**, 78–94.
52. Zhang, M.Q. (1997) Identification of protein coding regions in the human genome by quadratic discriminant analysis. *Proc. Natl Acad. Sci. USA*, **94**, 565–568 [Erratum. *Proc. Natl Acad. Sci. USA*, **94**, 5495].
53. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
54. Schultz, J., Milpetz, F., Bork, P. and Ponting, C.P. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl Acad. Sci. USA*, **95**, 5857–5864.
55. Ponting, C.P., Schultz, J., Milpetz, F. and Bork, P. (1999) SMART: identification and annotation of domains from signalling and extracellular protein sequences. *Nucleic Acids Res.*, **27**, 229–232.

Downloaded from https://academic.oup.com/hmg/article/8/12/2325/661862 by guest on 16 August 2022