

mount height and radius estimates.

20. The marine portions of the ETOPO5 data set derive from SYNAPS [R. J. Van Wyckhouse, *Tech. Rep. TR-233* (U.S. Naval Oceanographic Office, NOO-Washington, DC, 1973)] and contain numerous artifacts caused by the combination of poor data coverage, gridding of contours instead of depth soundings, and inappropriate gridding methodology [W. H. F. Smith, *J. Geophys. Res.* **98**, 9591 (1993)].

21. R. A. Duncan and D. A. Clague, in *The Ocean Basins and Margins*, A. E. M. Nairn, F. G. Stehli, S. Uyeda, Eds. (Plenum, New York, 1985), pp. 89–121; R. D. Jarrard and D. A. Clague, *Rev. Geophys.* **15**, 57 (1977); C. Y. Yan and L. W. Kroenke, *Proc. ODP Sci. Results* **130**, 697 (1993).

22. J. P. Morgan, W. J. Morgan, E. Price, *J. Geophys. Res.* **100**, 8045 (1995).

23. U. S. ten Brink, *Geology* **19**, 397 (1991).

24. P. Wessel and L. W. Kroenke, *Nature* **387**, 365 (1997).

25. R. D. Müller, W. R. Roest, J. Y. Royer, L. M. Gahagan, J. G. Sclater, *J. Geophys. Res.* **102**, 3211 (1997).

26. Robust regression of the envelope in Fig. 3 gives  $VGG(\Delta t) = 61.8 + 10.8\sqrt{\Delta t}$ . This is inverted to yield

the empirical relation

$$\text{pseudo age} = \text{seafloor age} - \left[ \frac{VGG(\Delta t) - 61.8}{10.8} \right]^2 \quad (2)$$

27. The flexural modeling also allowed numerical estimation of moat volumes.

28. R. Batiza, *Earth Planet. Sci. Lett.* **60**, 195 (1982).

29. W. H. F. Smith, thesis, Columbia University (1990).

30. The Ontong Java plateau was emplaced during two distinct episodes at ~121 Ma and ~89 Ma [D. Bercovic and J. Mahoney, *Science* **266**, 1367 (1994)]; the Manahiki plateau also formed at ~123 Ma, whereas the Hess rise (90 to 100 Ma) and the Mid-Pacific Mountains (75 to 130 Ma) have longer ranges or ages. The oldest plateau is Shatsky rise (138 to 145 Ma) [R. Larson and P. Olson, *Earth Planet. Sci. Lett.* **107**, 437 (1991)].

31. I thank W. Smith for providing the VGG grid. Supported by NSF grant EAR-9303402. School of Ocean and Earth Science and Technology, University of Hawaii, contribution no. 4517.

17 April 1997; accepted 13 June 1997

## Identification of the Tuberous Sclerosis Gene TSC1 on Chromosome 9q34

Marjon van Slegtenhorst, Ronald de Hoogt, Caroline Hermans, Mark Nellist, Bart Janssen, Senno Verhoef, Dick Lindhout, Ans van den Ouweland, Dicky Halley • Janet Young, Mariwyn Burley, Steve Jeremiah, Karen Woodward, Joseph Nahmias, Margaret Fox, Rosemary Ekong, John Osborne, Jonathan Wolfe, Sue Povey • Russell G. Snell, Jeremy P. Cheadle, Alistair C. Jones, Maria Tachataki, David Ravine, Julian R. Sampson • Mary Pat Reeve, Paul Richardson, Friederike Wilmer, Cheryl Munro, Trevor L. Hawkins • Tiina Sepp, Johari B. M. Ali, Susannah Ward, Andrew J. Green, John R. W. Yates • Jolanta Kwiatkowska, Elizabeth P. Henske, M. Priscilla Short, Jonathan H. Haines, Sergiusz Jozwiak, David J. Kwiatkowski\*

Tuberous sclerosis complex (TSC) is an autosomal dominant disorder characterized by the widespread development of distinctive tumors termed hamartomas. TSC-determining loci have been mapped to chromosomes 9q34 (*TSC1*) and 16p13 (*TSC2*). The *TSC1* gene was identified from a 900-kilobase region containing at least 30 genes. The 8.6-kilobase *TSC1* transcript is widely expressed and encodes a protein of 130 kilodaltons (hamartin) that has homology to a putative yeast protein of unknown function. Thirty-two distinct mutations were identified in *TSC1*, 30 of which were truncating, and a single mutation (2105delAAAG) was seen in six apparently unrelated patients. In one of these six, a somatic mutation in the wild-type allele was found in a TSC-associated renal carcinoma, which suggests that hamartin acts as a tumor suppressor.

TSC is a systemic disorder in which hamartomas occur in multiple organ systems, particularly the brain, skin, heart, lungs, and kidneys (1; 2). In addition to its distinct clinical presentation, two features serve to distinguish TSC from other familial tumor syndromes. First, the tumors that occur in TSC are very rare in the general population, such that several TSC lesions are, by them-

selves, diagnostic of TSC. Second, TSC hamartomas rarely progress to malignancy. Only renal cell carcinoma occurs at increased frequency in TSC (~2.5%) and with earlier age of onset; it appears to arise in TSC renal hamartomas, termed angiomyolipomas (3). Nonetheless, TSC can be a devastating condition, as the cortical tubers (brain hamartomas) frequently cause epilep-

sy, mental retardation, autism, or attention deficit-hyperactive disorder, or a combination of these conditions (1, 4).

TSC affects about 1 in 6000 individuals, and ~65% of cases are sporadic (5). Linkage of TSC to chromosome 9q34 was first reported in 1987, and this locus was denoted *TSC1* (6). Later studies provided strong evidence for locus heterogeneity (7) and led to the identification of chromosome 16p13 as the site of a second TSC locus (denoted *TSC2*) (8). The *TSC2* gene was identified by positional cloning, and the encoded protein, denoted tuberlin, contains a domain near the COOH-terminus with homology to a guanosine triphosphatase (GTPase) activating protein (GAP) for rap1, a Ras-related GTPase (9).

The focal nature of TSC-associated hamartomas has suggested that *TSC1* and *TSC2* may function as tumor suppressor genes. The occurrence of inactivating germline mutations of *TSC2* in patients with tuberous sclerosis (9–11) and of loss of heterozygosity (LOH) at the *TSC2* locus in about 50% of TSC-associated hamartomas (12–14) supports a tumor suppressor function for *TSC2*. In contrast, LOH at the *TSC1* locus has been detected in <10% of TSC-associated hamartomas (13, 14), suggesting the possibility of an alternative pathogenic mechanism for lesion development in patients with *TSC1* disease.

As part of a comprehensive strategy to identify *TSC1*, we identified 11 microsatellite markers from the 1.4-Mb *TSC1* region and developed an overlapping contig (with only a single gap of 20 kb) of cosmid, P1

### The *TSC1* Consortium:

M. van Slegtenhorst, R. de Hoogt, C. Hermans, M. Nellist, B. Janssen, S. Verhoef, D. Lindhout, A. van den Ouweland, D. Halley, Department of Clinical Genetics, Erasmus University and University Hospital, Rotterdam, Netherlands.

J. Young, M. Burley, S. Jeremiah, K. Woodward, J. Nahmias, M. Fox, R. Ekong, J. Wolfe, S. Povey, MRC Human Biochemical Genetics Unit and Galton Laboratory, University College of London, London NW1 2HE, UK.

J. Osborne, University of Bath, Bath BA2 7AY, UK.

R. G. Snell, J. P. Cheadle, A. C. Jones, M. Tachataki, D. Ravine, J. R. Sampson, Institute of Medical Genetics, University of Wales College of Medicine, Cardiff CF4 4XN, Wales, UK.

M. P. Reeve, P. Richardson, F. Wilmer, C. Munro, T. L. Hawkins, Whitehead Institute, MIT Center for Genome Research, Cambridge, MA 02139, USA.

T. Sepp, J. B. M. Ali, S. Ward, A. J. Green, J. R. W. Yates, Departments of Pathology and Medical Genetics, University of Cambridge, Addenbrooke's NHS Trust, Cambridge CB2 2QQ, UK.

M. P. Short, Department of Child Neurology, University of Chicago School of Medicine, Chicago, IL 60637, USA.

J. H. Haines, Molecular Neurogenetics Unit, Massachusetts General Hospital, 149 13th Street, Boston, MA 02129, USA.

S. Jozwiak, Division of Child Neurology, Children's Health Center, 04-736 Warsaw, Poland.

J. Kwiatkowska, E. P. Henske, D. J. Kwiatkowski, Division of Experimental Medicine and Medical Oncology, Brigham and Women's Hospital, Boston, MA 02115, USA.

\*To whom correspondence should be addressed. E-mail: kwiatkowski@calvin.bwh.harvard.edu

artificial chromosome (PAC), and bacterial artificial chromosome (BAC) clones (15). Figure 1 shows the *TSC1* region (16, 17), including limiting centromeric and telomeric markers, as derived from analyses of affected individuals (solid arrows) from families with individual lod scores of  $>2$  (18). These limits are also consistent with the information available from LOH studies (13). Two additional recombination events were identified in unaffected individuals (open arrows), also from families with lod scores of  $>2$  (19). In each of these families, two individuals from different generations carried the same recombinant chromosome, and all four had no evidence of TSC. Because the penetrance of TSC is nearly 100% (2), we concentrated our search within the 900-kb region between markers D9S2127 and DBH.

In a search for further positional information, we looked for large deletions and rearrangements by means of pulsed-field gel electrophoresis (Fig. 1) (9) and through

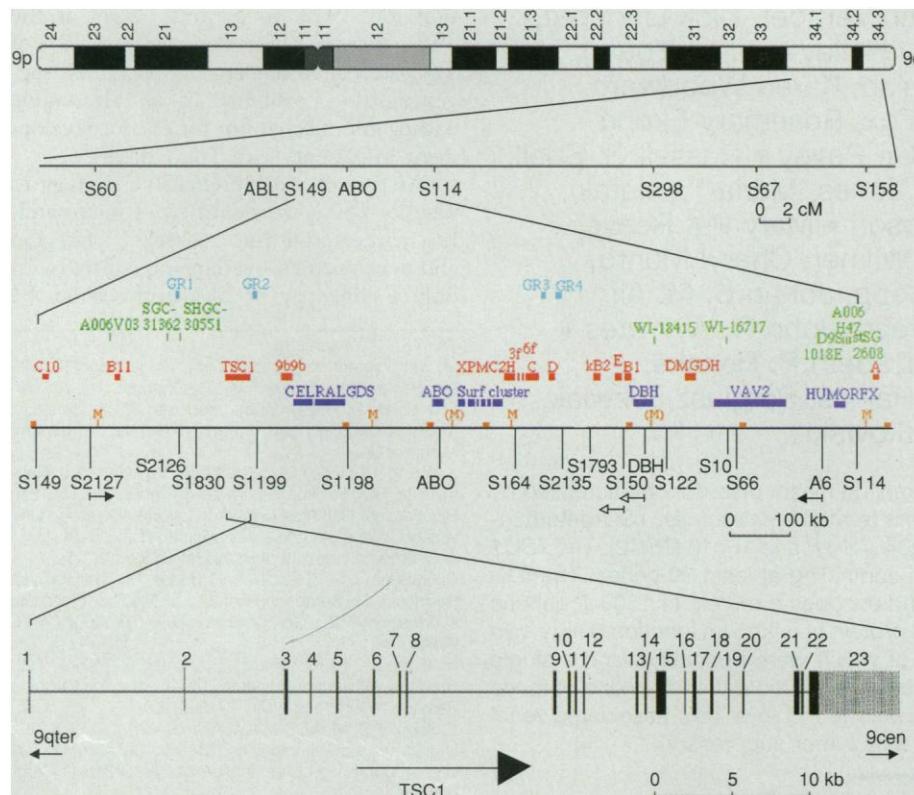
analysis of patient-derived hybrid cell lines retaining a single chromosome 9 bearing a *TSC1* mutation (20). No abnormalities were detected, and we therefore began a systematic gene-by-gene analysis.

Several techniques were used to identify genes in the *TSC1* region, which proved to be relatively gene-rich. Using a combination of exon trapping (21), cDNA selection, expressed sequence tag (EST) mapping, and whole-cosmid hybridization (22), we identified 142 exons and 13 genes between D9S1199 and D9S114. In all, 30 genes were identified or mapped to the 900-kb critical region.

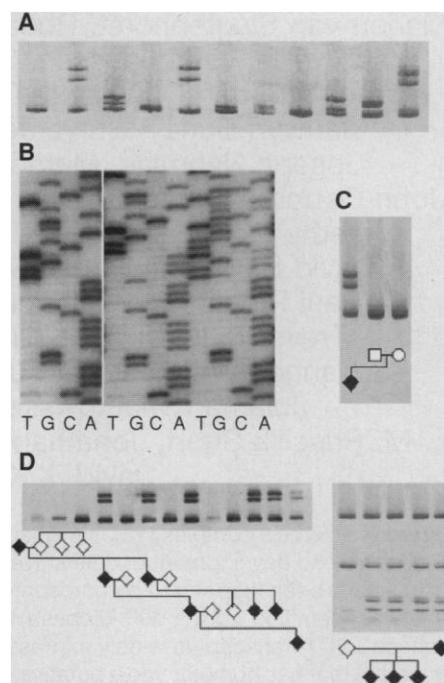
In parallel, we began sequencing the entire contig (23). We used the polymerase chain reaction (PCR) to amplify putative (24) and confirmed exons found in 208 kb of sequence on a screening set of 60 DNA samples from 20 unrelated familial TSC cases with linkage to 9q34, and 40 sporadic TSC cases (18). Amplification products were analyzed for heteroduplex formation

using weakly denaturing polyacrylamide gels (25). The 62nd exon screened demonstrated mobility shifts in 10 of the 60 patient samples (Fig. 2A).

Sequence analysis revealed seven small frameshifting deletions (three identical), one nonsense mutation, one missense change, and one polymorphism that did not change the encoded amino acid (Fig. 2B). Eight of the nine mutations were from the 20 familial cases tested, and only one mutation was seen among the 40 sporadic cases (Fig. 2C). Analysis of samples from other family members confirmed that each of the familial mutations segregated with TSC and that a frameshift mutation had occurred de novo in the sporadic case (Fig. 2D). The recurrent mutation, 2105delAAAAG, was identified in two apparently unrelated familial cases and a sporadic case. Haplotype analysis of the families, using markers flanking the mutation (D9S2126, D9S1830, and D9S1199, Fig. 1), confirmed that the three mutations were of independent origin.



**Fig. 1.** The *TSC1* region on chromosome 9. The ideogram (top) represents a normal G-banded metaphase chromosome 9, with the *TSC1* region located at 9q34. The male genetic map (next line) shows selected anchor polymorphic loci mapped to 9q34. The detailed physical map of the candidate region (next level) shows the positions of polymorphic markers and key recombination events in affected members (filled arrows) and unaffected members (open arrows) of families showing linkage of TSC to 9q34; the approximate positions of Mlu I (M) sites (with sites that partially cut in genomic DNA shown in parentheses) and of probes used to screen the region for rearrangements in patients with TSC by means of pulsed-field gel electrophoresis (orange boxes); genes previously mapped to the *TSC1* candidate region (blue boxes); novel cDNAs isolated from the region (red boxes); ESTs mapped to the region (green); and additional putative genes predicted by GRAIL analysis of genomic sequence (light blue boxes). There was a single 20-kb gap in the contig near D9S1793. The map of the *TSC1* gene (bottom) shows the 23 exons, of which exons 3 to 23 are coding.



**Fig. 2.** Identification of mutations in *TSC1* exon 15. (A) Heteroduplex analysis. Control sample (left lane) is followed by 10 samples with a shift. (B) Sequence analysis demonstrating 2105delAAAAG mutation. The sequence reactions were done in antisense orientation, so that reading from the top down (b2083 to 2124 of the normal sequence is shown), the allele sequenced on the left has the deletion, the middle sequence is a normal allele, and the sequence on the right is the heteroduplex product with both alleles. (C) In a sporadic case, the heteroduplex mobility shift is not present in either parent. (D) Segregation of heteroduplex mobility shifts in a large family with TSC (left) and digestion of amplification products with Mwo I in another family (right) demonstrates segregation of the 2105delAAAAG mutation with the disease.

The exon with mutations was part of a transcriptional unit identified by earlier gene discovery efforts (26). The full sequence of the *TSC1* gene was determined by comparison of genomic sequence and cDNA clone sequence, including clones obtained by 5' rapid amplification of cDNA ends (RACE). The *TSC1* gene consists of 23 exons, of which the last 21 contain coding sequence and the second is alternatively spliced (Fig. 1, bottom). The open reading frame (ORF) of the longest transcript begins at nucleotide 162, and the likely initiator ATG codon occurs at nucleotide 222. The first stop codon is at nucleotide 3738, leaving a 4.5-kb 3' untranslated region. Northern (RNA) blot analysis with a coding region probe (nucleotides 1100 to 2100) revealed a major 8.6-kb transcript that was widely expressed and was particularly abundant in skeletal muscle (Fig. 3).

The predicted *TSC1* protein, which we call hamartin, consists of 1164 amino acids with a calculated mass of 130 kD (Fig. 4). The protein is generally hydrophilic and has a single potential transmembrane domain at amino acids 127 to 144 (27) as well as a probable 266-amino acid coiled-coil region beginning at position 730 (28). Database searches identified a possible homolog of *TSC1* in the yeast *Schizosaccharomyces pombe* (GenBank accession number Q09778), a hypothetical 103-kD protein, but there were no strong matches with vertebrate proteins (29).

Because the initial screen identified a high frequency of mutations in exon 15, we studied this exon in a large sample of patients. Mutations in exon 15 [559 base pairs (bp), 16% coding region] were identified in 8 of 55 (15%) familial DNA samples with linkage to the *TSC1* region, and in 15 of 607 (2.5%) DNA samples from sporadic patients or families uninformative for linkage (Table 1). A

screen for mutations in all coding exons in 20 familial cases and 152 sporadic patients yielded eight mutations in each group (40% and 5%, respectively). In total, 19 mutations were found in coding exons other than exon 15. No mutations have been detected thus far in exons 3 to 6, 8, 11 to 14, 16, or 21 to 23. Of the 32 distinct mutations seen in 42 different patients or families, five were recurrent. Thirty were predicted to be truncating, one was a possible missense mutation, and one was a

splice site mutation. Analysis of a renal cell carcinoma from a *TSC* patient with germline mutation 2105delAAAG revealed a somatic mutation, 1957delG, in the wild-type *TSC1* allele (30). A giant cell astrocytoma from another patient with germline mutation 1942delGGinsTTGA had retained the mutant allele but lost the wild-type allele.

Our results support the hypothesis that *TSC1* functions as a tumor suppressor gene. First, the majority of mutations are likely to

**Fig. 4.** Predicted amino acid sequence of the *TSC1* protein, hamartin. A potential transmembrane domain (amino acids 127 to 144) and a coiled-coil domain (amino acids 730 to 965) are underlined. The *TSC1* genomic sequence and the cDNA sequence have been deposited in GenBank (accession numbers AC002096 and AF013168, respectively).

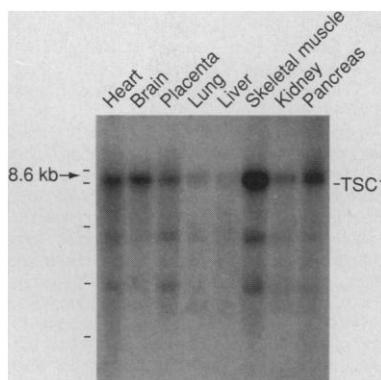
```

MAQQANVGEL LAMLDSPMLG VRDDVTVAFK ENLNSDRGPM LVNTLVDDYYL ETSSQPALHI 60
LTTLQEPHDK HLLDRINEYV GKAATRLSIL SLLGHVIRLQ PSWKHKLQA PLLPSLLKCL 120
KMDTDVYVLT TGVVLLITML EMLPQSGKQH LLDFDFIFGR LSSWCLKPKG HVAEVLVHL 180
HASVYALFHR LYGMVPCNFV SFLRSHYSMK ENLETFFEVV KPMMEHVRIH PELVTGSKDH 240
ELDPRWRKRL ETHDVIIECA KISLDPTFAS YEDGYVSQSH ISARFPHRSA DVTTSPYADT 300
QNSYGCATST PYSTRMLMLL NMPGQLPQTL SSPSTRLITE PPQATLWSPS MVCGMTTPPT 360
SPGNVPPDLS HPYSKVFQTT AGGKGTPLGT PATSPPPAPL CHSDDYVHIS LPQATVTPPR 420
KEERMSDARP CLHRQHHLN DRGSEEPGGS KGSVTLSDLE GFLGLDASEE DSIEKDKREA 480
AISRELSAIT TAAEAPVVRP GGFDSPPFYR SLPGSQKTH SAASSSQGAS VNPELHSSL 540
DKLGPDTPKQ APTPIDLPCG SADESPAGDR ECQTSLETSI FPTSPPCKIPP PTRVGFSGGQ 600
PPPYDHLFEV ALPKTAHFHFV IRKTEELLK AGNTEEDGV PSTSPMEVLD RLIQQGADAH 660
SKELNKLPLP SKSVDTHTFG GSPSPDEIRT LRDQLLLHN QLLYERFKRQ QHALRNRRL 720
RVVKAALAE EHNAAMKDL KLOEKDLOM KVSLOKROAR YNQLQORDT MYLKLHSGIR 780
OLOHDBREFF NOSQELQTKL EDCRNMIAEL RIELKANNK YCHTELLSQ YSOKLSNSES 840
VOCOMRFLNR QLLVLGEVNE LYLEOLONKH SDTTRKVEYM KAAYRKELEK NRSHVLOQTO 900
RLDTSOKRL ELESHLAKKD HLLLEOKKYL EDYKLOARGO LOAABSRYEA OKRITOVFEL 960
EILDLYGRLE KQGLLKLLEE EKAEAAEAAE ERLDCCNDGC SDSMVGHNEE ASGHNGETKT 1020
PRPSSARGSS GSRGGGGSS SSELSTPEK PPHQAGPFS SRWETTMGEA SASIPTTVGS 1080
LPSSKSLGM KARELFRNKS ESQCDEDMT SLSSELKTE LGKDLGVEAK IFLNLDGPHF 1140
SPPFPDSVGQ LHIMDYNETH HEHS 1164
    
```

**Table 1.** All mutations found in *TSC1*. Both heteroduplex and single-strand conformation polymorphism (33) gels were used to search for mutations after the initial screening. F, familial; S, sporadic.

Exon	Number of patients screened*		Mutations	Patients			
	F	S					
7	20	152	865delTT	1S			
9	39	230	966delA	1S			
			970T → G, L250X	1F			
			993G → T, E258X	1F			
			1112T → G, Y297X	1F			
10	20	152	1207delCT	1S			
			1746C → T, R509X	1F, 2S			
15	55	607	1750delCA	1S			
			1801delAG	1F			
			1892del23	1S			
			1929delAG	2S			
			1942delGGinsTTGA	1S			
			1981A → G, K585R	1F			
			2009delT	1S			
			2041delTT	1F			
			2060delA	1S			
			2105delAAAG	4F, 2S			
			2122delAC	2S			
			2126delAG	1F			
			2176delTG	1F			
			17	45	296	2295C → T, R692X	1F
						2324duplGT TACTC	1F
						2332delAT	1S
						2395insA	1S
18	45	296	2448C → T, Q743X	1S			
			2519del23bp	1S			
			2540delC	1F			
			2577C → T, R786X	1F, 1S			
			2583G → T, E788X	1F			
			2691delAC	1S			
			2724-1 G → T	1F			
19	39	230	2730insA	1S			
20	39	230					

\*Families are defined as those with linkage to the *TSC1* region and negative linkage to the *TSC2* region. Sporadics include both sporadic cases and cases from families without linkage information. Exon structure and primer information are provided at <http://expmed.bwh.harvard.edu/projects/tsc/>.



**Fig. 3.** Northern blot analysis of *TSC1* expression. Each lane contained 2 μg of polyadenylated RNA from adult human organs, and the probe consisted of base pairs 1100 to 2100 of the *TSC1* gene. Minor hybridization signals of size 4 and 2.5 kb are also seen.

inactivate protein function. Second, in two TSC-associated tumors we have shown that loss of the wild-type *TSC1* allele occurred through LOH or intragenic somatic mutation. The paucity of LOH for the *TSC1* region found in patient lesions (13, 14) may reflect the same mutational spectrum seen in the germline of TSC patients with a high frequency of small mutations causing inactivation of the second allele. It is also possible that there is a greater frequency of *TSC2*- versus *TSC1*-associated disease among the sporadic cases providing the lesions analyzed. This is suggested by the low frequency of mutations we have detected in *TSC1* in sporadic cases. However, in families suitable for linkage analysis, about half show linkage to *TSC1* and half to *TSC2* (16, 31).

The mutations observed in *TSC1* consist of small deletions, small insertions, and point mutations. No genomic deletions or rearrangements in *TSC1* were detected by Southern (DNA) blot analysis of 250 TSC patients. This restricted mutational spectrum may reflect an intrinsic tendency for this type of mutation in this region of the genome. Alternatively, it may reflect selection against more disruptive mutations such as large deletions, which would involve neighboring genes.

The mechanism by which loss of hamartin expression produces TSC lesions is unknown. It is likely that hamartin and tuberin participate in the same pathway of cellular growth control, because the clinical features of *TSC1* and *TSC2* disease are so similar (31). Tuberin has modest GAP activity for both rap1 and rab5, members of the Ras superfamily of small GTPases. The physiological function of the rap1 GTPase is not understood, whereas rab5 is thought to be involved in early endosomal transport. Tuberin-deficient rat embryo fibroblasts display increased endocytosis, which suggests that the rab5 interaction of tuberin has physiological relevance (32). It is unclear how a deficiency of GAP activity for rap1 or rab5, if that is the critical function of tuberin, leads to hamartoma development. The sequence homology of hamartin to a putative *S. pombe* protein suggests that it may participate in an evolutionarily conserved pathway of eukaryotic cell growth regulation. The identification of *TSC1* will enable analysis of the functions of both hamartin and tuberin, and may permit further insight into the molecular pathogenesis of TSC.

## REFERENCES AND NOTES

1. M. Gomez, *Tuberous Sclerosis* (Raven, New York, 1988); M. R. Gomez, *Ann. N.Y. Acad. Sci.* **615**, 1 (1991).
2. D. J. Kwiatkowski and M. P. Short, *Arch. Dermatol.* **130**, 348 (1994); E. Roach et al., *J. Child Neurol.* **7**, 221 (1992); D. W. Webb and J. P. Osborne, *J. Med. Genet.* **28**, 417 (1991).
3. J. Bjornsson, M. P. Short, D. J. Kwiatkowski, E. P. Henske, *Am. J. Pathol.* **149**, 1201 (1996); J. A. Cook, K. Oliver, R. F. Mueller, J. Sampson, *J. Med. Genet.* **33**, 480 (1996).
4. A. Hunt and C. Shepherd, *J. Autism Dev. Disord.* **23**, 323 (1993); S. L. Smalley, P. E. Tanguay, M. Smith, G. Gutierrez, *ibid.* **22**, 339 (1992).
5. J. P. Osborne, A. Fryer, D. Webb, *Ann. N.Y. Acad. Sci.* **615**, 125 (1991); J. R. Sampson, S. J. Scahill, J. B. Stephenson, L. Mann, J. M. Connor, *J. Med. Genet.* **26**, 28 (1989).
6. A. E. Fryer et al., *Lancet* **i**, 659 (1987).
7. L. A. J. Janssen et al., *Genomics* **8**, 237 (1990); H. Northrup et al., *Lancet* **ii**, 804 (1987); J. R. Sampson et al., *J. Med. Genet.* **26**, 511 (1989); R. Kandt et al., *Exp. Neurol.* **104**, 223 (1989).
8. R. Kandt et al., *Nature Genet.* **2**, 37 (1992).
9. European Tuberous Sclerosis Consortium, *Cell* **75**, 1305 (1993).
10. P. T. Brook-Carter et al., *Nature Genet.* **3**, 328 (1994).
11. P. J. Wilson et al., *Hum. Mol. Genet.* **5**, 249 (1996); A. Kumar et al., *ibid.* **4**, 1471 (1995); A. Kumar et al., *ibid.*, p. 2295; S. Verhoef et al., *Hum. Mutat.*, in press; R. Vrtel et al., *J. Med. Genet.* **33**, 47 (1996); Q. Wang et al., *Hum. Mutat.*, in press; A. Kumar et al., *ibid.* **9**, 64 (1997).
12. A. J. Green, M. Smith, J. R. Yates, *Nature Genet.* **6**, 193 (1994).
13. T. Sepp, J. R. W. Yates, A. J. Green, *J. Med. Genet.* **33**, 962 (1996); E. P. Henske et al., *Am. J. Hum. Genet.* **59**, 400 (1996).
14. C. Carbonara et al., *Genes Chromosomes Cancer* **15**, 18 (1996).
15. N. Hornigold et al., *Genomics* **41**, 385 (1997).
16. D. J. Kwiatkowski et al., *Cytogenet. Cell Genet.* **64**, 93 (1993); S. Povey et al., *Ann. Hum. Genet.* **58**, 177 (1994).
17. K. S. Au, J. Murrell, A. Buckler, S. H. Blanton, H. Northrup, *J. Med. Genet.* **33**, 559 (1996).
18. The diagnosis of tuberous sclerosis was made according to standard diagnostic criteria (2). Blood samples (obtained after informed consent) were used for DNA preparation, either directly or after creation of immortalized Epstein-Barr virus-transformed lymphoblastoid cell lines. Linkage to the *TSC1* region was inferred if a family demonstrated obligate recombination with markers within 2 centimorgans (cM) of *TSC2* and had positive lod scores (logarithm of the odds ratio for linkage) in analyses with 9q34 markers (Fig. 1). Families providing critical recombinant events were analyzed with multiple markers from 9q34, and haplotype analysis was performed manually to identify the site of recombination.
19. J. L. Haines et al., *Am. J. Hum. Genet.* **49**, 764 (1991); M. Nellist et al., *J. Med. Genet.* **30**, 224 (1993).
20. C. Jones, F. T. Kao, R. T. Taylor, *Cytogenet. Cell Genet.* **28**, 181 (1980).
21. Cosmid, BAC, or PAC DNA was digested with Pst I or Bam HI, and a library of subclones was prepared in pSP3L3 [A. J. Buckler et al., *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4005 (1991); E. P. Henske et al., *Ann. Hum. Genet.* **59**, 25 (1995)]. Exons identified in this manner were used to identify cDNA clones (i) by screening cDNA libraries by conventional methods, (ii) by screening GenBank and dbEST databases, and (iii) in reverse transcription PCR experiments. IMAGE clones were obtained from Research Genetics or the UK Human Genome Mapping Project (HGMP) Resource Centre.
22. Complementary DNA selection was performed using the end-ligation coincident sequence cloning method [A. J. Brookes et al., *Hum. Mol. Genet.* **3**, 2011 (1994)]. A normalized infant brain cDNA library [M. B. Soares et al., *Proc. Natl. Acad. Sci. U.S.A.* **91**, 9228 (1994)] was screened using whole cosmids as probes. A human fetal brain cDNA library (Clontech) was screened by standard phage plating and filter lift methods.
23. Cosmid (15) DNA was sheared and subcloned into M13mp18. Single clear plaques were picked using an automated picking device (PBA Technologies, Cambridge, UK) and expanded with JM101, and phage supernatant was collected. M13 DNA isolation was performed with the Sequatron robotic system [T. L. Hawkins et al., *Science* **276**, 1887 (1997)] following the solid-phase reversible immobilization protocol [T. L. Hawkins, T. O'Connor, A. Roy, C. Santillan, *Nucleic Acids Res.* **22**, 4543 (1994)]. Dye primer DNA sequencing used energy transfer primers and thermostable polymerase chain reaction (Amersham), and electrophoresis was performed on Applied Biosystems 377 DNA sequencers. Gel films were extracted, signal-processed, and bases called with the program Trout (available from genome.wi.mit.edu/distribution/software/trout) and were assembled with Alewife, a sequence assembly package. Typically, 1200 reads from a single cosmid assembled into one to three contigs, which were then finished by directed primer walking and directed selection of reverse reads from existing M13 templates to span sequence gaps. All sequence data and protocols were available during the sequence process from our Web site, <http://www-seq.wi.mit.edu>.
24. Genomic sequence was analyzed with the program GRAL2 to identify possible exons and gene models [Y. Xu, R. Mural, M. Shah, E. Uberbacher, *Gene Eng. Principles Methods* **16**, 241 (1994)]. Putative transcriptional units were also identified by BLAST searches of public databases and comparison with our own collection of cDNA clones and exon trapping products.
25. Oligonucleotide primers were designed to be external to exons by 40 to 60 bp where possible. DNA products with mobility shifts on heteroduplex analysis [F. J. Couch et al., *Nature Genet.* **13**, 123 (1996); A. Ganguly, M. J. Rock, D. J. Prockop, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10325 (1993)] were subjected to sequence analysis of both strands.
26. A transcriptional unit denoted B2 was identified from a brain cDNA library by hybridization with a cosmid near D9S1830 (22). A 4.5-kb cDNA clone was sequenced and contained no ORF. Database searches showed that this clone was the 3' portion of a 6.8-kb cDNA clone (K1AA0243) [T. Nagase et al., *DNA Res.* **3**, 321 (1996)], which contained a 2.0-kb ORF. After the discovery of mutations in exon 15 of B2, 5' RACE was performed with the Marathon cDNA kit (Clontech), using oligonucleotides derived from *TSC1* cDNA clones or inferred by analysis of genomic sequence information (24). RACE and other cDNA clones were sequenced fully on both strands by means of Taq cycle sequencing methodology.
27. R. F. Smith, B. A. Wiese, M. K. Wojzynski, D. B. Davison, K. C. Worley, *Genome Res.* **6**, 454 (1996).
28. A. Lupas, M. Van Dyke, J. Stock, *Science* **252**, 1162 (1991).
29. A BLAST search of the NCBI combined protein database using the *TSC1* amino acid sequence gave the best match at a probability of  $2.2 \times 10^{-18}$  with the Q09778 sequence, with several regions of homology throughout the sequence but no areas of particularly strong match. The next lowest probability was  $5.2 \times 10^{-8}$ .
30. Allele-specific amplification was performed on the renal cell carcinoma DNA using a primer specific for the normal allele, not bearing the 2105delAAAG mutation. Sequence analysis of this product indicated that it contained the 1957delG mutation.
31. S. Povey et al., *Ann. Hum. Genet.* **58**, 107 (1994); B. Janssen et al., *Hum. Genet.* **94**, 437 (1994).
32. R. Wienecke, A. Konig, J. E. DeClue, *J. Biol. Chem.* **270**, 16409 (1995); G. H. Xiao, F. Shoarinejad, F. Jin, E. A. Golemis, R. S. Yeung, *ibid.* **272**, 6097 (1997); R. Wienecke et al., *Oncogene* **13**, 913 (1996).
33. M. Orita, Y. Suzuki, T. Sekiya, K. Hayashi, *Genomics* **5**, 874 (1989).
34. We thank the hundreds of TSC patients and their families for blood samples and financial support (particularly V. Whittemore and W. Watts), and numerous clinicians (particularly A. Oranje, H. Stroink, M. Wildervanck de Blecock, H. Kros, B. Prahli-Andersen, and H. L. J. Tanghe) for assistance; H. Galjaard and P. Harper for support; R. Bakker, S. Ramakhan, D. Humphrey, Q. Wang, M. Maheshwar, A. L. W. Hesseling-Janssen, N. Thomas, and R. Slomski for technical support; and the UK HGMP Resource Centre and the European Collection of Animal Cell Cultures. Supported by the NIH, National Tuberous Sclerosis Association, Tuberous Sclerosis Association of Great Britain, Dutch Organization for Scientific Research (NWO), Dutch Tuberous Sclerosis Association (STSN), Dutch Kidney Foundation, Westminster Medical School Research Trust (Royds Fund), Bath Unit for Research into Paediatrics, and New Zealand Health Resources Council.

11 June 1997; accepted 9 July 1997