Human Molecular Genetics, 1996, Vol. 5, No. 1 **131–137**

# Comparative analysis and genomic structure of the tuberous sclerosis 2 (TSC2) gene in human and pufferfish

**Magitha M. Maheshwar, Richard Sandford[1], Mark Nellist[+], Jeremy P. Cheadle, Barbara Sgotto[1], Mark Vaudin[2] and Julian R. Sampson***

Institute of Medical Genetics, University of Wales College of Medicine, Cardiff, CF4 4XN, UK, [1]Department of Medicine, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, UK and [2]Department of Genetics and Genome Sequencing Center, Washington University School of Medicine, St Louis, MO 63108, USA

Received September 9, 1995; Revised and Accepted October 27, 1995

**Germ-line mutations of the TSC2 tumour suppressor gene have been identified in humans with tuberous sclerosis and in the Eker rat. Tuberin, the human TSC2 gene product, has a small region of homology with rap1GAP and stimulates rap1 GTPase activity *in vitro*, suggesting that one of its cellular roles is to function as a GTPase activating protein (GAP). We have undertaken a comparative analysis of the TSC2 gene in human and the pufferfish, *Fugu rubripes*. In addition to the GAP domain, three other regions of the proteins are highly conserved (peptide sequence similarity >80%). These regions are likely to represent further functional domains. To facilitate analysis of mutations within these domains we have determined the genomic structure of the human TSC2 gene. It comprises 41 exons, including exon 31 which was absent from the originally described spliceoform of the human TSC2 transcript and was identified following exon prediction from Fugu genomic sequence. These findings support the proposal of the Fugu genome as a tool for human gene analysis.**

## INTRODUCTION

In humans, constitutional mutations affecting the TSC2 gene on chromosome 16 are found in individuals with the autosomal dominant trait tuberous sclerosis, MIM 191100 (1). Linkage studies in families with tuberous sclerosis (TSC) show that approximately 50% segregate a mutant TSC2 gene, while an unidentified gene which maps to chromosome 9 (designated TSC1) is implicated in the remainder (2–5). With the exception of a recently described contiguous gene syndrome involving deletion of both the TSC2 gene and the adjacent PKD1 gene on chromosome 16 (6), phenotypes associated with TSC1 and TSC2 mutations appear indistinguishable. Both are characterized by development of unusual hamartomatous growths affecting many organs, especially the brain, skin, kidneys and heart (7). Loss of heterozygosity (LOH) across the TSC1 and TSC2 chromosomal regions in TSC associated hamartomas (8–10) suggests that cell proliferation and differentiation become perturbed when both TSC1 or TSC2 alleles are mutated and that the genes act as tumour suppressor genes serving related functions. The wide range of tissues in which TSC associated hamartomas develop implies a fundamental role for both TSC genes in regulating cell proliferation and differentiation. It is likely that study of the TSC2 gene and its protein product tuberin (and when identified, the TSC1 gene and its product) will provide new insights into these processes.

The rat homologue of the TSC2 gene has recently been characterized (11) and a constitutional insertion mutation has been identified in the Eker strain (12,13), which exhibits a predisposition to renal cell carcinoma, splenic haemangioma and uterine leiomyoma or leiomyosarcoma (14). Analogous pathology is only occasionally seen in humans with TSC, suggesting that further studies of TSC2 might also provide insights into species-specific differences in tumorigenesis and/or phenotype-specific mutations.

The human TSC2 gene was identified by positional cloning (1). It maps to a gene-rich region of 16p13.3, approximately 2.25 Mb from the telomere and immediately adjacent to the PKD1 gene (15). The 5.5 kb transcript spans an estimated 43 kb of genomic sequence and encodes a novel, predicted ~198 kDa protein (tuberin). Fifty-eight amino acids near the carboxy terminus are homologous to the catalytic domain of the GTPase activating protein GAP3 (rap1GAP) (1). GAPs are regulators of the GTP binding and hydrolysing activity of the Ras superfamily of proteins that help to regulate cell growth, proliferation and differentiation. Further homologies between tuberin and other proteins have not been identified and alternative approaches will be required to elucidate other functional domains. Strategies

*To whom correspondence should be addressed

[+]Present address: Department of Clinical Genetics, Erasmus University, 3015 GE Rotterdam, The Netherlands

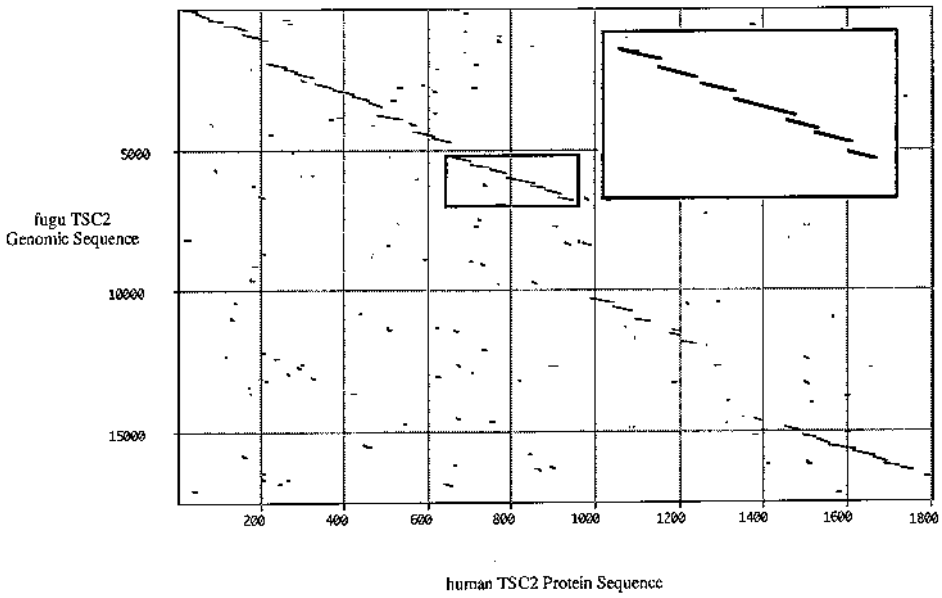


**Figure 1.** Protein–DNA matrix plot of the human TSC2 predicted protein against the Fugu TSC2 genomic sequence following three frame translation. Analysis was performed using the MacVectorTM analysis package. A window size of 30, minimum percentage score of 20, and hash value of 2 were used with a pam250 matrix. The homology between the two sequences was sufficient to allow identification of all exons. The enlarged insert panel demonstrates the program's ability to identify individual exons as regions of homology.

which might prove useful in this regard include characterization of mutations at the TSC2 locus in humans and other species and identification of regions of the gene which exhibit a high level of conservation in distantly related species.

The haploid genome of the Japanese pufferfish, *Fugu rubripes*, comprises ~400 Mb (16), compared with ~3000 Mb in man, but is estimated to contain a similar number of genes. The compact nature of the Fugu genome reflects the small relative size of intergenic and intronic sequences and a paucity of repetitive DNA. These characteristics have led to the proposal of *Fugu rubripes* as an efficient model for vertebrate genome analysis (16). We have characterized the Fugu homologue of the TSC2 gene and have undertaken a comparative analysis of TSC2 in human and Fugu. Four regions of coding sequence, including the rap1GAP homologous region, showed a particularly high level of conservation. These may represent critical functional or structural domains toward which further studies, such as mutation analysis might be usefully directed. To facilitate the identification and analysis of potentially revealing mutations, such as missense mutations, we have determined the intron–exon structure of the human TSC2 gene.

## RESULTS

### Isolation and sequencing of the Fugu TSC2 gene

Genomic clones containing the Fugu homologue of the TSC2 gene were identified as previously described (Sandford *et al.* manuscript submitted). Briefly, human TSC2 cDNA clones 4B2 and 4.9 (1) were used to screen a Fugu genomic library constructed in lambda 2001 (16). Three overlapping lambda clones were identified and fragments of these used to isolate a single cosmid, 295C6, from a gridded library containing 40 000 clones (G. Elgar, unpublished). This clone was also found to contain part of the PKD1 gene homologue and was sequenced in its entirety (Sandford *et al.* manuscript in preparation). Exon 1 was found to be interrupted by the vector cloning site. An overlapping cosmid, 52G18, was isolated and the sequence of exon 1 completed using specific walking primers.

### Predicted intron–exon structure of Fugu TSC2

The intron–exon structure of the Fugu homologue of TSC2 was predicted using the genomic sequence from cosmid 295C6. The BLAST algorithm (17) and direct comparison to the human predicted protein sequence after three frame translations with MacVectorTM were used to search for Fugu sequences with homology to the human TSC2 cDNA. Searches using the known consensus sequences for Fugu splice sites (18) were then made to predict positions of Fugu intron–exon boundaries, which were also analysed using the Netgene (19) and Grail (20) programs. Homology between Fugu genomic sequence and human tuberin sequence predicted 41 Fugu exons comprising a transcript of similar size to the previously published human TSC2 transcript (Fig. 1). GC content and Grail (20) analysis also predicted a further Fugu exon of 90bp (exon 31) for which no homologous sequence was present in the published human transcript (Fig. 2). Alternative splicing of exon 31 in Fugu was confirmed by RT–PCR using pooled RNAs.

### Intron–exon structure of the human TSC2 gene

Sequence data from genomic sub-clones of the cosmids LADS4 and ZDS5 which together span the human TSC2 locus (see Fig. 3 and Methods) were compared with the human TSC2 cDNA. The genomic sequences at points of divergence were compared with the 3′ and 5′ splice site consensus sequences defined by Shapiro and Senapthy (21), identifying 59 intron–exon boundaries. A further 19 boundaries were identified by direct sequencing of PCR products amplified from the genomic TSC2 locus.

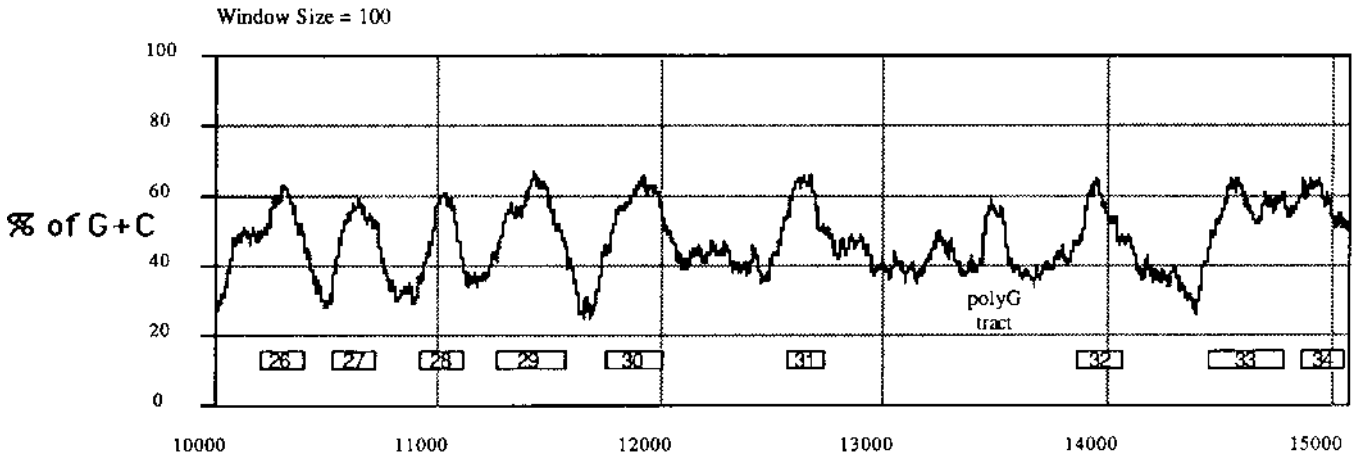


**Figure 2.** Percentage GC content of the Fugu TSC2 genomic region containing the newly identified exon 31. Using the MacVectorTM program and a window size of 100 discrete GC peaks of greater than 60% can be seen to correspond to exons (open boxes) predicted by homology searching. The demonstration of a possible exon 31 in the Fugu gene led to the discovery of this exon in the human gene. Both have been confirmed as being alternatively spliced by either RT–PCR or cDNA sequencing. The same exon has been shown to exist in the rat gene.

The 40 exons defined by these approaches accounted for all 5754 nucleotides of the published cDNA sequence. An additional, alternatively spliced exon (exon 31) which was not represented in the original TSC2 transcript was also identified. This exon was predicted by analysis of Fugu genomic sequence (see above). RT–PCR amplification of human lymphoblastoid RNA revealed no evidence for the exon, but 4A2, one of eight human fetal brain cDNA clones spanning this part of the transcript, contained an additional 69 bp of coding sequence homologous to the alternatively spliced exon in Fugu. Genomic PCR products spanning the novel exon were amplified from cosmid ZDS5 and the exon and flanking intronic regions were confirmed by direct sequencing. Thus a total of 41 human TSC2 exons were identified (Table 1). A combination of PCR, restriction digestion and Southern hybridization were used to estimate the sizes of 28 introns for which sequence data were incomplete, and to confirm the positions of exons on the genomic map (Fig. 3). Partial genomic sequence across the 41 exons and flanking intronic sequence has been deposited with GenBank (accession numbers L48517–L48546).

### Interspecies comparison

The known TSC2 gene spans ~43 kb of genomic DNA in human and only 17.3 kb in Fugu. However, in both cases transcript size (~5.5 kb) including putative translation initiation codons and polyadenylation signals, and intron–exon organization are very similar. The difference in genomic coverage reflects smaller intron size in the Fugu (62–1928 bp in Fugu, and 79 to >4 kb in humans). The only structural divergence noted between the Fugu and human TSC2 genes was the presence of an additional Fugu intron. Optimal alignment indicated that this intron of 163 bp was inserted in phase 0 at amino acid number 1435 of the human transcript (WSASGE--DSRGQP), splitting exon 33, the largest of the human exons, into two Fugu exons denoted 33a and 33b. Otherwise both position and phase of predicted Fugu introns were identical to their directly determined human counterparts (Table 1).

After optimal alignment using the CLUSTAL program (22), the TSC2 transcripts in Fugu and humans showed 65% sequence homology at nucleotide level. GC content was 60% in man and 54% in Fugu. Optimal alignment at the protein level revealed 60% identity of amino acid residues, with 79% similarity if conservative changes were included. Comparison of the predicted Fugu and human peptide sequences is shown in Figure 4. The PLOTSIMILARITY program from the GCG package was used to assess variation in similarity across both predicted proteins as a 75 amino acid residue window was moved along the aligned sequences. Four regions of high conservation were identified. These include the GAP-related domain (human residues 1593–1631) and the sequences flanking this, two small regions between residues 750 and 1100, and the N-terminal portion of the molecule.

### Secondary structure and protein similarity predictions

Analysis of the human TSC2 predicted protein sequence has previously identified hydrophobic regions which could represent membrane spanning domains (1). Hydropathy of the Fugu and human predicted proteins was examined using the Kyte and Doolittle method (23) and both N-termini were found to be hydrophobic. Alignments of the two sequences were analysed for transmembrane domains using TMAP (24) and PredictProtein (25–27), but none were reliably predicted.

Apart from the small region of tuberin showing sequence homology to rap1GAP, no similarities to known proteins have been found (1). Extensive searches for structural and functional protein domains were made with the aligned Fugu and human sequences (see Materials and Methods). Evolutionary conservation can increase the power of these analyses, but no similarity to other protein motifs or domains was found.

## DISCUSSION

The comparative analysis we have undertaken in humans and Fugu has highlighted four regions of the TSC2 gene product which have remained highly conserved (>80% at the amino acid

**Table 1.** Intron–exon structure of TSC2 in human and fugu

| Human | | | | Fugu | | | Intron Phase | Human 3' Splice junction | Human 5' Splice junction |
|---|---|---|---|---|---|---|---|---|---|
| Exon | Nucleotide Nos. | Size | Amino acid Nos. | Exon | Size | Amino acid Nos. | | | |
| 1 | 1-156 | >156[#] | 1-46 | 1 | 135 | 1-45 | 0 | - | CTGAGA/gtgagtgagctacc |
| 2 | 157-243 | 87 | 47-75 | 2 | 87 | 46-74 | 0 | tcatctctctccag/GAACTG | GAAGAG/gtaggtttatccag |
| 3 | 244-354 | 111 | 76-112 | 3 | 111 | 75-111 | 0 | tctgctggtgacag/CACGCA | GGGCAG/gtaaggccagggcg |
| 4 | 355-499 | 145 | 113-160 | 4 | 148 | 112-160 | I | gcttttgtctttag/GGCGAG | AGCTGG/gtgggtgccacctt |
| 5 | 500-617 | 118 | 161-200 | 5 | 118 | 161-200 | II | gacttctccccccag/CTGACT | GGTTCA/gtaagaaaagaatt |
| 6 | 618-666 | 49 | 201-216 | 6 | 49 | 201-216 | 0 | ggtgctccctgcag/GATGAT | ATAGAG/gtcagtgcctcccc |
| 7 | 667-792 | 126 | 217-258 | 7 | 126 | 217-258 | 0 | tccctctccaccag/GTCTCC | TGGAAG/gtgggtttctgaa |
| 8 | 793-866 | 74 | 259-283 | 8 | 74 | 259-283 | II | attgtgtctcgcag/CTGATG | GGACAG/gtgagtgtggtggg |
| 9 | 867-993 | 127 | 284-325 | 9 | 127 | 284-325 | 0 | gggaacacttttag/AGCCTA | TACCAG/gtaaggcggtttct |
| 10 | 994-1137 | 144 | 326-373 | 10 | 144 | 326-373 | 0 | gctcgtgttcccag/GCCATG | CTCCAG/gtgggtgggggca |
| 11 | 1138-1275 | 138 | 374-419 | 11 | 138 | 374-419 | 0 | gcctggtactgcag/ACCTTG | AGGCCT/gtgagacccctcc |
| 12 | 1276-1379 | 104 | 420-454 | 12 | 104 | 420-454 | II | tcttctttgacag/GAGTCC | CTTCAG/gtaggggtcctctg |
| 13 | 1380-1461 | 82 | 455-481 | 13 | 82 | 455-481 | 0 | ctgccgtcccgcag/GAGCGA | TATGAG/gtgcgtgtcaggcg |
| 14 | 1462-1617 | 156 | 482-533 | 14 | 156 | 482-533 | 0 | ttgtgcctgtgcag/GAGGAG | GAGAAG/gtgagagccgttgt |
| 15 | 1618-1734 | 117 | 534-572 | 15 | 120 | 534-573 | 0 | cttctctttgaaag/GTGATG | CTTCAG/gtgggtgttctgca |
| 16 | 1735-1857 | 123 | 573-613 | 16 | 123 | 574-614 | 0 | tcctctctctgcag/ACCAAG | CTGCAG/gtatggtggctggg |
| 17 | 1858-1964 | 107 | 614-649 | 17 | 107 | 615-650 | II | cctcttcctgacag/GCCTTT | CTACAT/gtacgcggacctgc |
| 18 | 1965-2115 | 151 | 650-699 | 18 | 163 | 651-704 | 0 | tcttgcttctgcag/GGAGCC | AAGCAG/gtgagtgggccgg |
| 19 | 2116-2238 | 123 | 700-740 | 19 | 123 | 705-745 | 0 | tcctctcctcgcag/GAGTCT | TCCATG/gtaccatgccggcc |
| 20 | 2239-2373 | 135 | 741-785 | 20 | 135 | 746-790 | 0 | gatttggtcatcag/CTTTCA | AAACAG/gtaggagaggtcag |
| 21 | 2374-2563 | 190 | 786-848 | 21 | 190 | 791-853 | I | cctgtctcctgcag/CGCGAG | TGTCCA/gtgagtccccgccc |
| 22 | 2564-2657 | 94 | 849-880 | 22 | 94 | 854-885 | II | ctccattaccgcag/CTCTGG | CTCCAA/gtgagtggtcgccc |
| 23 | 2658-2760 | 103 | 881-914 | 23 | 103 | 886-919 | 0 | ccttctcatctcag/GTTTAA | ACTAAG/gtgggctcagggcg |
| 24 | 2761-2855 | 95 | 915-946 | 24 | 95 | 920-951 | II | gctcactctgccag/GGCCTG | CAAGAG/gtacggcatgcggg |
| 25 | 2856-2984 | 129 | 947-989 | 25 | 141 | 952-998 | II | acttacccagatag/TCTGAG | CCGCAG/gtagcggactgtcg |
| 26 | 2985-3149 | 165 | 990-1044 | 26 | 162 | 999-1052 | II | gcctctccctccag/CAGGAT | GAAGAG/gtccaggcggcact |
| 27 | 3150-3302 | 153 | 1045-1095 | 27 | 159 | 1053-1105 | II | tccctgtcttctag/GTCTCC | GTCGAG/gtgactgcaccttc |
| 28 | 3303-3415 | 113 | 1096-1133 | 28 | 113 | 1106-1142 | I | gattctcttctcag/CTCCAG | TGTCGG/gtgagccttggccc |
| 29 | 3416-3628 | 213 | 1134-1203 | 29 | 243 | 1143-1223 | I | cccgtcttcttcag/GGGGCC | CCACAG/gtactgggctaggg |
| 30 | 3629-3832 | 204 | 1204-1271 | 30 | 228 | 1224-1299 | I | gccgcctccgtcag/GGAACA | ACACAG/gtgagtggcatggc |
| 31 | 3833-3901 | 69 | 1272-1294 | 31 | 90 | 1300-1329 | I | gccttcccttgcag TGGCCT | GGGCAG gtatcgcctctcag |
| 32 | 3902-4023 | 122 | 1295-1335 | 32 | 158 | 1330-1382 | 0 | ctgtgtcctcccag/ACTCCG | AGCAGG/gtgagtgtggctca |
| 33 | 4024-4511 | 488 | 1336-1498 | 33a<br>33b | 249<br>167 | 1383-1465<br>1466-1521 | 0[*]<br>II | atggtccttttctag/TCGTCC | CCCCAG/gtgggcctcttgct |
| 34 | 4512-4587 | 76 | 1499-1523 | 34 | 79 | 1522-1547 | 0 | ccctccctgtgcag/TTTCGT | AATGAG/gtaggcgtggctcc |
| 35 | 4588-4680 | 93 | 1524-1554 | 35 | 87 | 1548-1576 | 0 | tctgtgtgccacag/TCACAG | GGCCAG/gtgaggctgcaaaa |
| 36 | 4681-4867 | 187 | 1555-1616 | 36 | 187 | 1577-1638 | I | gtcagtcccgacag/AGCAAC | TGCAAG/gtacggcctggcgc |
| 37 | 4868-5007 | 140 | 1617-1663 | 37 | 140 | 1639-1685 | 0 | ctgctcaccctcag/CCGTCT | ATCAAG/gtgagtgagggccg |
| 38 | 5008-5086 | 79 | 1664-1689 | 38 | 79 | 1686-1711 | I | tcttgtccgggcag/GGCCAG | GGAAAG/gtagggccgggtgg |
| 39 | 5087-5178 | 92 | 1690-1720 | 39 | 92 | 1712-1742 | 0 | ccaagtctccccag/ACATGG | GCAAAT/gtgagtgggggtgg |
| 40 | 5179-5277 | 99 | 1721-1753 | 40 | 99 | 1743-1775 | 0 | tgcctacgtcccag/ATGGCC | CAGCGG/gtagggaatatggg |
| 41 | 5278-5439 | 162 | 1754-1807 | 41 | 165 | 1776-1830 | - | gcctctgccttcag/ATCTGC | - |

[*]Human exon 1 contains at least 19 nucleotides upstream of the translation initiation codon.

[†] In Fugu, a phase 0 intron, absent in the human gene, separates exons 33a and 33b. Human nucleotide numbers 1–3832 (amnio acids 1–1271) correspond to the previously published transcript (1). Thereafter they differ due to the present of the additional exon (exon 31).

level) over some 400 million years of evolution. These include the previously recognized GAP homologous region (1) and an area of ~200 amino acid residues which surrounds this. It is likely that this conserved area as a whole contributes to any functional GAP domain. Preliminary *in vitro* biochemical evidence for (rap1) GAP activity has been demonstrated for tuberin (28), but additional cellular roles are expected. In this regard the three remaining regions that are highly conserved between human and Fugu deserve further investigation.

Prediction of intron–exon structure from genomic sequence at the Fugu TSC2 locus helped us to identify a novel, alternatively spliced exon in humans which was absent from the previously reported TSC2 transcript (1). The additional exon was identified by the presence of a GC rich genomic region 3′ to exon 30 in the Fugu. By characterizing eight human fetal brain cDNA clones we confirmed that the additional exon (exon 31) was also present in some human transcripts. The exon corresponded to the alternatively spliced exon 31 of the recently characterized rat TSC2 gene (11). Sequencing of the exon 31 containing splice form showed that 67 of 69 nucleotides were identical to the homologous rat exon. It has already been established that exon 25 shows tissue specific alternative splicing in humans and mouse (29) and that it is alternatively spliced in the rat (11). Using RT–PCR we showed that exons 25 and 31 are also alternatively spliced in Fugu. Conservation of these splice forms is suggestive of underlying functional correlates which remain to be established.

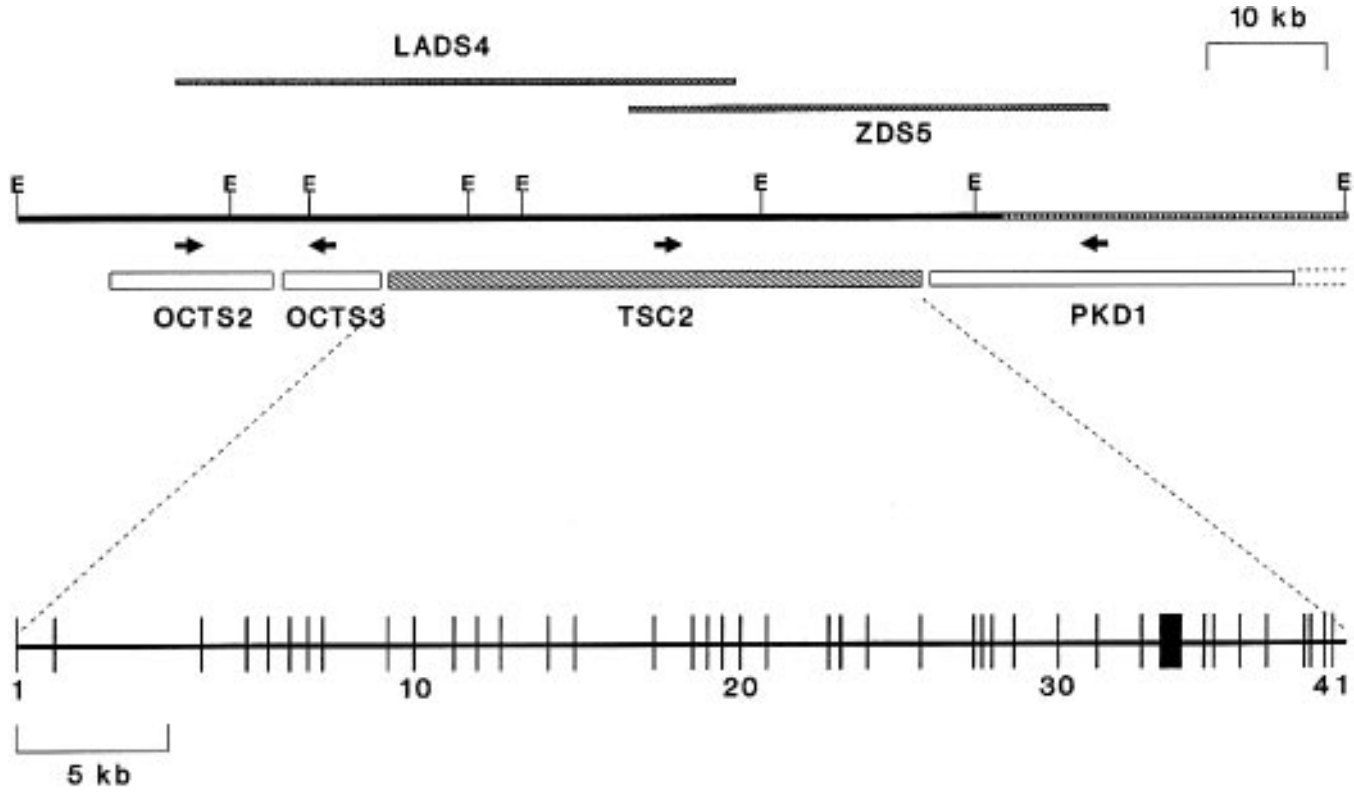


**Figure 3.** The human TSC2 locus. Top, The positions of cosmid clones LADS4 and ZDS5 are shown above a genomic map indicating restriction sites for*Eco*RI (E). Part of the genomic region spanned by ZDS5 is duplicated more proximally on the short arm of chromosome 16 (vertical hatches). The genomic distribution and direction of transcription (arrows) of the TSC2 gene and the adjacent genes, OCTS2, OCTS3 and PKD1 are shown below the restriction map. Bottom, genomic distribution of 41 coding exons (vertical bars) of the human TSC2 gene.

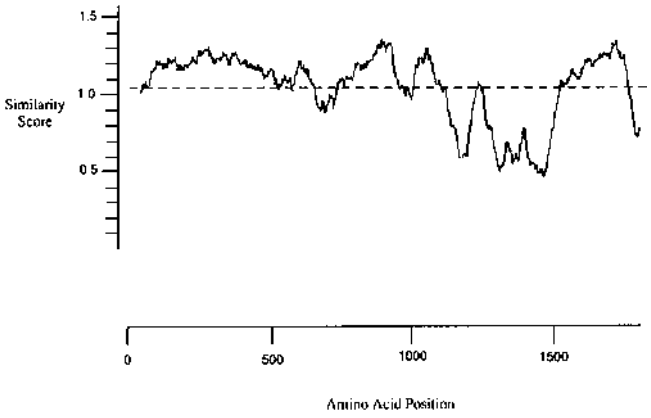


**Figure 4.** Similarity plot produced from the alignment of the TSC2 predicted protein sequences. Similarity was determined over a window of 75 residues using the GCG package program, PLOTSIMILARITY. 100% similarity scores 1.5. The scale along the bottom of the plot represents the consensus sequence after alignment and is therefore larger than the individual sequences. The dotted line shows the average similarity.

The predicted genomic structure of the Fugu TSC2 gene was extremely similar to that determined for the human TSC2 gene. The position and phase of each of 41 introns interrupting the human TSC2 transcript were identical in Fugu, the only divergence in organization being the occurrence of an additional intron (phase 0) of 163 bp splitting the coding sequence of human exon 33 into Fugu exons 33a and 33b. Given this level of conservation it is unsurprising that intron position and phase are completely conserved in the recently characterized rat TSC2 gene, which has no intron splitting the homologue of human exon 33 (11). For a very few genes, such as the Huntington's disease gene (30) and the G6PD gene (31) genomic structure have been determined in both Fugu and humans. In these cases complete conservation of intron–exon boundaries has been found throughout protein coding regions. The overall size of the human genome (and other mammalian genomes) is ~7.5-fold larger than that of Fugu, reflecting the larger size of both intergenic and intronic sequences in mammals. However, relative difference in size varies between genes. The TSC2 gene was 2.5-fold larger in humans than Fugu, compared with 7.5-fold for the HD gene (30) and 3.75-fold in the case of the G6PD gene (31). In contrast, the level of nucleotide sequence conservation within the coding regions of these genes is very similar, at 65% for TSC2, 69% for the HD gene and 71% for the G6PD gene.

It is possible that additional, alternatively spliced protein encoding exons of the TSC2 gene remain to be found. However, we do not predict the existence of such exons from our analysis of genomic sequence in Fugu. It is more likely that one or more leader exons (non-protein encoding) exist 5′ to the ATG codon identified as the probable translation initiation codon in man (1), mouse (Olsson *et al*. manuscript submitted), rat (11) and Fugu. A sequence corresponding to the splice site acceptor consensus lies

~30 bp upstream from this codon at the human TSC2 locus. As yet, the 5′UTR of the TSC2 gene has not been fully characterized in any organism, but delineation of this region, and the regulatory elements contained within it, is under investigation.

The data we present on genomic structure and sequence will assist with the identification of constitutional and somatic mutations at the human TSC2 locus. Some 5% of germ-line TSC2 mutations appear to be deletions detectable by pulse field gel electrophoresis or conventional Southern analysis (1,9). More detailed studies applying single strand conformation polymorphism (SSCP) and heteroduplex analysis may be useful both diagnostically and in the identification and characterization of functional domains of tuberin and vital residues within these. Unfortunately, the complex genomic structure of the human TSC2 gene with many small exons will make any exon screening approach to mutation identification labour intensive. As a wide spectrum of inactivating mutations might be expected, alternative approaches such as the protein truncation test (PTT) (32) may be more appropriate in a diagnostic setting. However, PTT does not detect missense mutations, which are potentially the most revealing in terms of function, and SSCP and related techniques will remain useful in the research-oriented laboratory.

## MATERIALS AND METHODS

### Sequencing of Fugu cosmid 295C6

The sequence of the Fugu TSC2 gene was obtained from the complete sequence of cosmid 295C6 and individual walking sequences from the adjacent cosmid 52G18. Sonicated cosmid 295C6 DNA was shotgun cloned into *Sma*I cut M13 vector. Sequence was obtained using both dye primer and dye terminator chemistries and automated detection with ABI 373 or 377 sequencers. Walking primers were used to confirm sequence on both strands and to obtain the complete sequence of exon 1 from cosmid 52G18.

### Human TSC2 cDNA clones

Four cDNA clones spanning the human TSC2 transcript have been described previously, 2A-6, 4.9, 4B2 and 1A1 (1). Additional cDNAs were identified by screening a human fetal brain cDNA library (Stratagene) with the established TSC2 cDNAs. Previously undescribed cDNAs are 4.1, 4A2, 4B5, 1.1, 1/1, 1/3, 3.3, 4B4, 2A1, 2/1, 2B1 and 1B1. These were mapped relative to cosmids LADS4 and ZDS5 and to each other. Eight of the clones spanned the region of the gene predicted to contain an additional exon (exon 31) in Fugu. These were investigated for evidence of the additional exon by PCR using the primers 3699F (5′-TGCCCCTGCAGGAGCTGTCTA-3′) and MP53 (5′-GCGCCTGTCCATGCCTAGC-3′). cDNA 4A2 was found to contain an additional 69 bp. The PCR product was sequenced directly.

### Genomic clones at the human TSC2 locus

*Pst*I and *Taq*I restriction fragments from cosmids LADS4 and ZDS5 (1), were shotgun cloned into pBluescript (Stratagene). Recombinants were picked on to duplicate grids and transferred to Hybond membrane. Exon containing clones were identified by colony hybridization with TSC2 cDNA fragments according to the method of Grunstein and Hogness (33) and the positive genomic clones were assigned to their relative positions in the cDNA.

### Splice junction analysis of human TSC2

Initially, intron–exon boundaries were determined by single stranded sequencing of the *Pst*I and *Taq*I shotgun clones using the M13 reverse primer (30 boundaries). Further sequencing of these clones using additional intragenic primers identified an additional 29 boundaries (primer sequences available upon request). The boundary sequences were completed by direct sequencing of PCR products generated from genomic clones at the TSC2 locus. PCR amplification was carried out in a 50 µl total volume containing 0.2 mM of each dNTP, 25 pmol of each primer and 1 unit of Amersham *Taq* polymerase in standard PCR buffer (10 mM Tris–HCl, 50 mM KCl, 2 mM $MgCl_2$). PCR conditions used for each primer set were similar (95°C 2 min followed by 30 cycles of 95°C 1 min, 55°C–60°C 1 min, 72°C 2 min). Direct DNA sequencing was carried out using the PCR product sequencing kit (Amersham), using exonuclease I and shrimp alkaline phosphatase for pre-enzymic treatment of the PCR products. The sequences at all splice sites were analysed for agreement with the 3′ and 5′ splicing consensus sequences using the method described by Shapiro and Senapathy (21).

### Restriction map construction

An approximate genomic map was constructed by hybridization of TSC2 cDNA fragments to restriction digests of cosmid and human genomic DNA. These results were confirmed by hybridization of exon-containing shotgun clones to the same digests. Finer mapping was achieved by restriction digest analysis of the shotgun clones and PCR products derived from the TSC2 genomic locus. In many instances restriction mapping data were confirmed by sequence analysis.

### RT–PCR

Total RNA from Fugu pooled tissues and from human lymphoblastoid lines for use in RT–PCR experiments was isolated using guanidinium thiocyanate–phenol–chloroform extraction (34). First-strand cDNA was obtained from 1 µg of RNA using the Promega Reverse Transcription System according to the supplied protocols. One-twentieth of the reaction volume was used in subsequent PCR reactions with Fugu or human specific primers.

### Primer synthesis

Primers were synthesized by either Severn Biotech Ltd, Kidderminster or Oswel DNA Service, Edinburgh.

### Database searches and computer analysis

Sequence analysis and exon prediction was carried out with MacVectorTM, GRAILII (20), and Netgene (19). Homology searches to identify sequence, motif and structural similarity were performed with Blastn, Blastp, PredictProtein (25–27), TMAP (24), MOTIF (Prosite directory, ref. 35), and BLOCKS by direct access to the servers via the worldwide web. Sequence alignments were performed using the GCG package.

## ACKNOWLEDGEMENTS

The authors would like to thank their colleagues in the European Chromosome 16 Tuberous Sclerosis Consortium and also Sydney Brenner, Sam Aparicio, Anne-Marie Frischauf and Peter Harper. This work was supported by the Tuberous Sclerosis Association (GB), the Medical Research Council, the London Law Trust and the Dutch Kidney Foundation.

## ABBREVIATIONS

TSC: tuberous sclerosis; TSC2: tuberous sclerosis gene 2; PKD1 polycystic kidney disease gene 1; GAP: GTPase activating protein; bp: base pair; kb: kilobase pair; Mb megabase pair; kDa: kilodalton; RT–PCR: reverse transcription–polymerase chain reaction.

## REFERENCES

1. The European Chromosome 16 Tuberous Sclerosis Consortium (1993) Identification and characterisation of the tuberous sclerosis gene on chromosome 16. *Cell*, **75,** 1305–1315.
2. Sampson, J.R., Janssen, L.A.J., Sandkuijl, L.A. and the Tuberous Sclerosis Collaborative Group (1992) Linkage investigation of three putative tuberous sclerosis determining loci on chromosomes 9q, 11q and 12q. *J. Med. Genet.*, **29,** 861–866.
3. Kwiatkowski, D.J., Armour, J., Bale, A.E., Fountain, J.W., Goudie, D., Haines, J.L., Knowles, M.A., Pilz, A., Slaugenhaupt, S. and Povey, S. (1993) Report on the second international workshop on human chromosome 9. *Cytogenet. Cell Genet.*, **64,** 94–106.
4. Povey, S., Burley, M.W., Attwood, J., Benham, F., Hunt, D., Jeremiah, S.J., Franklin, D., Gillet, G., Malas, S., Robson, E.B., Tippett, P., Edwards, J.H., Kwiatkowski, D.J., Super, M., Mueller, R., Fryer A., Clarke, A., Webb D. and Osborne J. (1994) Two loci for tuberous sclerosis: one on 9q34 and one on 16p13. *Ann. Hum. Genet.*, **58,** 107–127.
5. Janssen, B., Sampson, J., van der Est, M., Deelen, W., Verhoef, S., Daniels, I., Hesseling, A., Brook-Carter, P., Nellist, M., Lindhout, D., Sandkuijl, L. and Halley, D. (1994) Refined localisation of TSC1 by combined analysis of 9q34 and 16p13 data in 14 tuberous sclerosis families. *Hum. Genet.*, **94,** 437–440.
6. Brook-Carter, P.T., Peral, B., Ward, C.J., Thompson, P., Hughes, J., Maheshwar, M., Nellist, M., Gamble, V., Harris, P.C. and Sampson, J.R. (1994) Deletion of the TSC2 and PKD1 genes associated with severe infantile polycystic kidney disease—a contiguous gene syndrome. *Nature Genet.*, **8,** 328–332.
7. Gomez, M.R. (1988) *Tuberous Sclerosis*, 2nd edition. Raven Press, New York.
8. Green, A.J., Smith, M. and Yates, J.R.W. (1993) Loss of heterozygosity on chromosome 16p13.3 in hamartomas from tuberous sclerosis patients. *Nature Genet.*, **6,** 193–196.
9. Green, A.J., Johnson, P.H. and Yates, J.R.W. (1994) The tuberous sclerosis gene on chromosome 9q34 acts as a growth suppressor. *Hum. Mol. Genet.*, **3,** 1833–1834.
10. Carbonara, C., Longa, L., Grosso, E., Borrone, C., Garre, M.G., Brisigotti, M. and Migone, M. (1994) 9q34 loss of heterozygosity in a tuberous sclerosis astrocytoma suggests a growth suppressor-like activity also for the TSC1 gene. *Hum. Mol. Genet.*, **3,** 1829–1832.
11. Kobayashi, T., Nishizawa, M., Hirayama, Y., Kobayashi, E. and Hino, O. (1995) cDNA structure, alternative splicing and exon–intron organisation of the predisposing tuberous sclerosis (Tsc2) gene of the Eker rat model. *Nucleic Acids Res.*, **23,** 2608–2613.
12. Kobayashi, T., Hirayama, Y., Kobayashi, E., Kubo, Y. and Hino, O. (1995) A germline insertion in the tuberous sclerosis (TSC2) gene gives rise to the Eker rat model of dominantly inherited cancer. *Nature Genet.*, **9,** 70–74.
13. Yeung, R.S., Xiao, G-H., Jin, W-C., Testa, J.R. and Knudson, A.G. (1994) Predisposition to renal cell carcinoma in the Eker rat is determined by germ-line mutation of the tuberous sclerosis 2 (TSC2) gene. *Proc. Natl Acad. Sci. USA*, **91,** 11413–11416.
14. Everitt, J.I., Goldsworthy, T.L., Wolf, D.C. and Walker, C.L. (1992) Hereditary renal cell carcinoma in the Eker rat: a rodent familial cancer syndrome. *J. Urol.*, **148,** 1932–1936.
15. The European Polycystic Kidney Disease Consortium (1994) The polycystic kidney disease 1 gene encodes a 14kb transcript and lies within a duplicated region on chromosome 16. *Cell*, **77,** 881–894.
16. Brenner, S., Elgar, G., Sandford, R., Macrae, A., Venkatesh, B. and Aparicio, S. (1993) Characterisation of the pufferfish (Fugu) genome as a compact model vertebrate genome. *Nature*, **366,** 265–268.
17. Altschul, S.F., Gish, W., Miller, M., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215,** 403–410.
18. Elgar, G., Sandford, R., Aparicio, S., Venkatesh, B., Macrae, A. and Brenner, S. (1995) Comparative genomic analysis using the Japanese Puffer fish (*Fugu rubripes*). *Trends Genet.*, in press.
19. Brunak, S., Engelbrecht, J. and Knudsen, S. (1991) Prediction of human mRNA donor and acceptor sites from the DNA sequence. *J. Mol. Biol.*, **220,** 49–65.
20. Xu, Y., Mural, R. J., Shah, M. B. and Uberbacher, E. C. (1994) Recognising exons in genomic sequence using GRAIL II. In *Genetic Engineering: Principles and Methods*, Vol 15. Jane Setlow (Ed.). Plenum Press.
21. Shapiro, M.B. and Senapathy, P. (1988) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.*, **15,** 7155–7174.
22. Higgins, D.G. (1994) CLUSTAL V: multiple alignment of DNA and protein sequences. *Methods Mol. Biol.*, **25,** 307–318.
23. Kyte, J. and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, **157,** 105–132.
24. Persson, B. and Argos, P. (1994) Prediction of transmembrane segments in proteins utilising multiple sequence alignments. *J. Mol. Biol.*, **237,** 182–192.
25. Rost, B. and Sander, C. (1993) Prediction of protein structure at better than 70% accuracy. *J. Mol. Biol.*, **232,** 584–599.
26. Rost, B. and Sander, C. (1994) Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins*, **19,** 55–72.
27. Rost, B., Casadio, R., Fariselli, P. and Sander, C. (1995) Prediction of helical transmembrane segments at 75% accuracy. *Protein Sci.*, **4,** 521–533.
28. 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.
29. Xu, L., Sterner, C., Maheshwar, M., Wilson, P.J., Nellist, M., Short, P., Haines, J., Sampson, J.R. and Ramesh, V. (1995) Alternative splicing of the tuberous sclerosis 2 (TSC2) gene in human and mouse tissues. *Genomics*, **27,** 475–480.
30. Baxendale, S., Abdulla, S., Elgar, G., Buck. D., Berks, M., Micklem, G., Durbin, R., Bates, G., Brenner, S., Beck, S. and Lehrach H. (1995) Comparative sequence analysis of the human and pufferfish Huntington's disease genes. *Nature Genet.*, **10,** 67–76.
31. Mason., P.J., Stevens, D.J., Luzzatto, L., Brenner, S. and Aparicio, S. (1995) Genomic structure and sequence of the Fugu rubripes glucose-6-phosphate dehydrogenase gene (G6PD). *Genomics*, **26,** 587–591.
32. Roest, P.A.M., Roberts, R.G., Sugino, S., van Ommen, G-J. and den Dunnen, J.T. (1993) Protein truncation test (PTT) for rapid detection of translation-terminating mutations. *Hum. Mol. Genet.*, **2,** 1719–1721.
33. Grunstein, M. and Hogness, D.S. (1975) Colony hybridisation: a method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl Acad. Sci. USA*, **72,** 8961.
34. Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Anal. Biochem.*, **162,** 156–159.
35. Bairoch, A. (1992) Prosite: a dictionary of sites and patterns in proteins. *Nucleic Acids Res.*, **20,** 2013–2018.