2608–2613 Nucleic Acids Research, 1995, Vol. 23, No. 14

cDNA structure, alternative splicing and exon–intron organization of the predisposing tuberous sclerosis (*Tsc2*) gene of the Eker rat model

Toshiyuki Kobayashi, Masae Nishizawa, Youko Hirayama, Etsuko Kobayashi and Okio Hino*

Department of Experimental Pathology, Cancer Institute, 1-37-1 Kami-ikebukuro, Toshima-ku, Tokyo 170, Japan

Received May 3, 1995; Revised and Accepted June 12, 1995

ABSTRACT

The Eker rat hereditary renal carcinoma (RC) is an excellent example of a Mendelian dominant predisposition to a specific cancer in an experimental animal. We recently reported that a germline insertion in the rat homologue of the human tuberous sclerosis gene (*TSC2*) gives rise to the dominantly inherited cancer in the Eker rat model. We now describe the entire cDNA (5375 bp without exons 25 and 31) and genomic structure of the rat *Tsc2* gene. The deduced amino acid sequence (1743 amino acids) shows 92% identity to the human counterpart. Surprisingly, there are a great many (≥41) coding exons with small sized introns spanning only ~35 kb of genomic DNA. Two alternative splicing events [involving exons 25 (129 bp) and 31 (69 bp)] make for a complex diversity of the *Tsc2* product. The present determination of the *Tsc2* gene and establishment of strong conservation between the rat and man provide clues for assessing unknown gene functions apart from that already predicted from the GTPase activating proteins (GAP3) homologous domain and for future analysis of intragenic mutations in tumors using methods such as PCR–SSCP and for insights into diverse phenotypes between species.

INTRODUCTION

The hereditary renal carcinoma (RC) in the rat, originally reported by R. Eker in 1954, is an example of a Mendelian dominantly inherited predisposition to a specific cancer in an experimental animal (1). Recently, the predisposing gene of the Eker rat was mapped to the proximal part of rat chromosome 10 (2,3). We have established a new conserved linkage group on rat 10q and human 16p13.3 whereby the Eker mutation was found to be tightly linked to the tuberous sclerosis (*Tsc2*) gene (4) and finally identified a germline insertion in the *Tsc2* gene (5,6). At the histological level, RCs develop through multiple stages from early preneoplastic lesions (phenotypically altered tubules) to adenomas in virtually all heterozygotes by the age of 1 year (7). We previously reported that ionizing radiation induces additional tumors (large adenomas and carcinomas), with a linear dose–response relationship (7). Loss of heterozygosity (LOH) at chromosome 10, where the predisposing *Tsc2* gene is localized, was found in the RCs which developed from hybrid F1 rats carrying the Eker mutation (2,8), indicating that in heterozygotes at least two events (one inherited, one somatic) are necessary to produce large adenomas and carcinomas. We have recently detected LOH of the wild-type allele even in the earliest preneoplastic lesions, e.g., phenotypically altered renal tubules (9), supporting the hypothesis that a second somatic mutation (second hit) might be a rate-limiting step for renal carcinogenesis in the Eker rat model of dominantly inherited cancer, as well as a tumor suppressor nature for the *Tsc2* gene function. Investigation of extra-renal primary tumors occurring in Eker rats late in life additionally revealed pituitary adenomas, probable hemangiosarcomas of the spleen, and probable leiomyosarcomas of the uterus (10,11). The Eker rat thus bears a single gene mutation with dominant predisposition and develops tumors in four different organs, although the predilection for extra-renal tumors is not as complete as with RCs. The phenotype of tuberous sclerosis in humans differs from that in the Eker rat, except for the occurrence of RCs (in man renal angiomyolipomas are more common) (5). At present, we do not have a good explanation for the variation in phenotypic manifestations, but the Eker rat model should provide insights into species-specific differences in tumorigenesis and/or phenotype-specific mutations.

The function of the *TSC2* gene product (called 'tuberin' in the human case) is not yet understood, although it contains a short amino acid sequence homology to ras family GTPase activating proteins (GAP3) (12). To establish gene function(s), it is essential that we know the primary structure of the rat *Tsc2* gene. Therefore, in the present study, we determined the organization of the entire rat *Tsc2* cDNA and genomic DNA.

MATERIALS AND METHODS

RNA and DNA isolation

Tissue total RNAs were isolated by guanidium–isothiocyanate/acid phenol method from adult male rats (Long Evans, KIWA Breeding Laboratory, Japan) (5). Poly(A)$^+$ RNAs were isolated using oligo (dT) latex beads (Nippon Roche) (5). Genomic DNA

* To whom correspondence should be addressed

was isolated from an adult male rat liver by the SDS/proteinase K method followed by phenol extraction as described earlier (13).

Construction and screening of cDNA libraries

An oligo (dT)-primed, unidirectional adult rat kidney cDNA library was constructed using a λZAPII cDNA synthesis kit (Stratagene). For the random-primed cDNA library, first strand cDNA was synthesized from 2.5 μg of adult rat kidney poly(A)$^+$ RNA using random-hexamers. After second strand synthesis and adaptor ligation, the cDNAs were cloned into the *Eco*RI site of λZAPII (Stratagene). Both of these two libraries consisted of ~10^6 recombinant clones. Screening of cDNA libraries was carried out according to a standard procedure (14). A 3′ portion of the human *TSC2* cDNA (nucleotides 4651–5371) was used as a probe for initial screening (4,5).

RT–PCR

First strand cDNAs were synthesized from 4 μg of total RNAs using random-primers and SuperScript reverse transcriptase (Gibco-BRL). One μl aliquots of the resulting reaction mixtures were subjected to PCR in 25 μl reactions containing 67 mM Tris–HCl (pH 8.8), 16.6 mM $(NH_4)_2SO_4$, 1 mM $MgCl_2$, 200 μM of each of four dNTPs, 0.2 mg/ml gelatin, 0.45% Triton X-100, 25 pmol of each primer and 2 U *Tth* DNA polymerase (Boehringer Biotech). One cycle of 93°C for 5 min for initial denaturation, 35 cycles of 93°C for 1 min, 55°C for 1 min, 72°C for 1.5 min for amplification, and one cycle of 72°C for 5 min for final extension were performed. Primers used were RTSC27 (sense, 5′-CCATCACGTCATAGCCATGT-3′, nucleotides 2667–2686) and RTSC28 (anti-sense, 5′-TGGGCCATAGAGTTCTCATC-3′, nucleotides 3010–3029) for alternative splicing 1, RTSC18 (sense, 5′-GTCTAATGCCCTTATGGCTG-3′, nucleotides 3693–3712) and RTSC21 (anti-sense, 5′-AGGAGATGGCCCGCTCAAT-3′, nucleotides 4081–4099) for alternative splicing 2. Then, products were separated by 2% low melting point agarose gel electrophoresis and transferred onto nylon membranes. Southern hybridization using ^{32}P-labelled internal oligonucleotide probes was used to detect specific amplification products. Probes used were RTSC50 (sense, 5′-TCCAATGTCCTCCTGTCCTT-3′, nucleotides 2752–2771) and ALSP6 (anti-sense, 5′-CCACATGTTCAGATACACTGA-3′, nucleotides 2939–2959 of the alternatively spliced exon) for alternative splicing 1, RTSC24 (anti-sense, 5′-AATGCTGCCTCAAAGTCCTC-3′, nucleotides 3952–3971) and ALSP1 (sense, 5′-TCTTTCTCCTCCCTGTACCA-3′, nucleotides 3820–3839 of the alternatively spliced exon) for alternative splicing 2. Pre-hybridization, hybridization and washing were carried out as described earlier except for the hybridization and washing temperature (55°C) (5). DNA fragments from each band were subcloned into the *Sma*I site of pBluescript SK(+) for sequence analysis.

5′-RACE

First strand cDNA was synthesized from 4 μg of adult brain total RNA using 5 pmol of a sequence-specific primer, RTSC36 (anti-sense, 5′-TCCAAAGTGCCTCCACTGC-3′, nucleotides 229–247) in a 20 μl reaction. After dG tailing with terminal deoxinucleotidil transferase, one tenth of the product was subjected to PCR using an internal primer, RTSC37 (anti-sense, 5′-CACAGATCTGCCCTATCATC-3′, nucleotides 177–196), and $(dC)_{12}$ primer. Conditions of PCR was essentially the same as described for RT–PCR except for the annealing temperature (50°C) during the amplification cycles. The PCR product was subcloned into the *Sma*I site of pBluescript SK(+) for sequence analysis.

Screening of a rat genomic cosmid library

A rat genomic cosmid DNA library (derived from the Wistar strain) was obtained from Dr T. Oda (Hamamatsu University School of Medicine) (15). Colonies (~10^5) were plated and screened with the rat *Tsc2* cDNA probes. Cosmid DNA was isolated from the positive colonies and their inserts analyzed by restriction enzyme digestion.

Sequencing reaction and analysis

Nucleotide sequences of cDNA clones were determined using a T7 sequencing kit (Pharmacia) with [^{32}P]- or [^{35}S]dCTP (16). Nucleotide sequences of the exon–intron boundaries were determined using cosmid DNAs (100 ng/each reaction) as templates and a SequiTherm sequencing kit (Epicentre Technologies) according to the supplier's instructions. To determine sequences of the exon–intron boundaries, various sequence specific primers were designed from the cDNA sequence. After obtaining initial sequence information, next primers were designed to cover entire coding exons and exon–intron boundaries. Direct sequencing of PCR products were performed using a SequiTherm sequencing kit. Sequence analysis was carried out using the GCG package (17).

Genomic PCR analysis

Cosmid DNAs (2 ng/each reaction) and genomic DNA (100 ng/each reaction) were used for intron amplification PCR with various combinations of primers on the exons. PCR was performed as described in RT–PCR analysis except for a longer extension time (2 min) during amplification cycles. Products were separated in 1% agarose gels to determine their lengths. Sizes of each intron were estimated by subtraction of the length of the exon portion from the size of the genomic PCR product (5).

RESULTS

Cloning and primary structure of the rat Tsc2 cDNA

Initially, we screened the oligo (dT)-primed cDNA library from an adult rat kidney using a fragment covering a 3′ portion of the human *TSC2* cDNA as a probe (4,5). Four positive clones were isolated from ~3×10^5 recombinant clones. The sequence of the longest clone, 7a1 was established to extend from the nucleotide corresponding to nucleotide 2428 of the human *TSC2* cDNA to the 3′ poly(A) tail (Fig. 1A). However, it lacks the 126 bp corresponding to that from nucleotides 2856–2981 of the human *TSC2* cDNA (see below). To isolate cDNAs covering the 5′ part of the *Tsc2* transcript, ~2×10^5 recombinant clones from a random-primed adult rat kidney cDNA library were screened using the 0.7 kb *Eco*RI fragment from the 5′ part of 7a1 as a probe. One of three positive clones, 2a2, had a 3.3 kb fragment covering most of the 5′ part of the *Tsc2* transcript, but still lacked a putative translational initiation codon (Fig. 1). Therefore, we cloned the remaining 5′ terminal part of the rat *Tsc2* cDNA by 5′-RACE. Several 5′-RACE products contained an identical sequence homologous to the 5′-terminal part of the human *TSC2* cDNA and has a putative translation initiation codon at the identical position to the human sequence (12). The entire length of the rat *Tsc2* cDNA comprising these clones (7a1, 2a2 and 5′-RACE clones) is 5375 bp (Fig. 1).

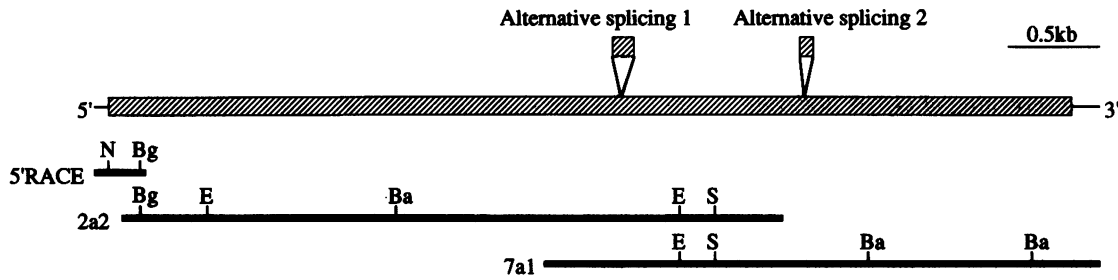


Figure 1. Structure of the rat *Tsc2* cDNA. Restriction maps of the two cDNA clones and one 5′ RACE clone are shown below the schematic representation of the rat *Tsc2* cDNA structure. The hatched box indicates the coding region deduced from the nucleotide sequence. Two insertions of the alternatively spliced exons found by RT–RCR are also shown above. Restriction enzymes are *Nco*I (N), *Bgl*II (Bg), *Eco*RI (E), *Bam*HI (Ba) and *Sma*I (S). The entire nucleotide sequence of the rat *Tsc2* cDNA clone and two alternative splicing exons was deposited in EMBL/GenBank/DDBJ database (accession no. D50413).

The identity of its sequence to the human *TSC2* cDNA is ~86% other than the missing 126 bp described above. The 5′- and 3′-non-coding regions are 35 and 108 bp, respectively. A poly(A) addition signal (5′-AATAAA-3′) is present 24 bp upstream from the poly(A) tail. A polypeptide consisting of 1743 amino acid (aa) residues is predicted from this cDNA sequence, with ~92% aa sequence identity to the human *TSC2* product other than the 42 aa residues encoded by the missing 126 bp. The GAP3 homologous domain (39 aa residues) is conserved between man and the rat with one conservative amino acid substitution (valine to isoleucine in the rat) (data not shown) (12). Although the presence of a leucine-zipper like domain in the amino terminal part of the human *TSC2* product was suggested, one of the leucine repeats is substituted by proline (aa 95) in the rat *Tsc2* aa sequence (data not shown). A direct repeat of nine amino acid residues is completely identical to that of the human *TSC2* product (data not shown) (12). A stretch of 41 aa residues mostly divergent from the human sequence was found between amino acids 1098 and 1138 (exon 29; see below) with a total of 23 residues showing replacement. In both human and rat sequences, this region is rich in proline, glycine and alanine residues (data not shown).

Alternative splicing of the rat *Tsc2* gene

As noted above, we obtained cDNA clones which lack the 126 bp corresponding to the sequence from nucleotides 2856–2981 of human *TSC2* cDNA. To examine the possibility that this 126 bp is subjected to an alternative splicing event, we performed RT–PCR analysis using a primer set encompassing this region. As shown in Figure 2A ('alternative splicing 1'), two bands were obtained from both brain and kidney RNAs although ratios of upper and lower bands differed. Subclones from the upper band had a 126 bp insertion after nucleotide 2837 (Fig. 2B). As expected, this 126 bp sequence is 94% identical to that from nucleotides 2856–2981 of the human *TSC2* cDNA (12). Among 42 aa residues generated by insertion of this 126 bp, only two amino acid residues differ from the human sequence (Fig. 2B). From these observations, we predicted that this 126 bp might be an alternative splicing exon. However, analysis of the genomic sequence revealed that, together with this 126 bp, the following 5′-CAG-3′ is also derived from an alternatively spliced exon (thus, exon 25 is 129 bp) and that there is a differential usage of the 3′ exon splicing acceptor sites (Figs 2E and 3B). To coincide with this, sequences of subclones from the lower band were divided into two groups, one identical to the cDNA clone and the other lacking three bases (5′-CAG-3′; nucleotides 2967–2969). Without these three bases, a serine residue is absent from the amino acid sequence (Fig. 2C). Although we did not observe these acceptor site variants in subclones from the upper band, direct sequence analysis of the upper band indicated that there are also two acceptor site variants (Fig. 2D). Again, a serine residue is absent from the amino acid sequence without the three bases. Thus, four types of nucleotide sequence are generated by these alternative splicing and differential acceptor site usage events (Fig. 2E).

During our previous study for identification of the germ line *Tsc2* mutation in the Eker rat, we identified another alternatively spliced exon by RT–PCR (5). In Figure 2A (' alternative splicing 2'), a result of RT–PCR analysis to distinguish the two splicing variants is shown. Subclones from the lower band had a sequence identical to the cloned cDNA, whereas those from the upper band contained an additional 69 bp insertion after nucleotide 3814 (Fig. 2B). This 69 bp exon has a novel sequence as no corresponding one has been described in the human *TSC2* gene (12). Furthermore, searching of the EMBL/GenBank data base showed no homologous sequence in known DNA and RNA sequences. The insertion site (after nucleotide 3814) of this 69 bp in the *Tsc2* cDNA is an exon junction site as we described earlier (5). By genomic sequence analysis, we detected splicing donor (5′-GT-3′) and acceptor sites (5′-AG-3′) adjacent to this 69 bp exon (see below).

Exon–intron structure of the rat Tsc2 gene

To perform further analysis of the *Tsc2* mutations in Eker rat tumors ('second hit') at the sequence level, we determined exon–intron boundaries of the rat *Tsc2* gene. We isolated two overlapping genomic cosmid clones (cosTsc2-1 and cos 2d1) covering the entire coding region of the *Tsc2* gene (Fig. 3A). Using these cosmid DNAs as templates, exon–intron boundaries were determined by direct sequence analysis with various sequence-specific primers designed from the *Tsc2* cDNA sequence. There are 41 exons, including two alternatively spliced exons, in the rat *Tsc2* gene (Fig. 3). All introns start with the splicing donor site consensus, 5′-GT-3′, and end with the acceptor site consensus, 5′-AG-3′. The GAP3 homology region is divided into two exons, 36 and 37, at the codon for the second alanine residue. The most diverged 41 aa stretch described above is located within one exon (exon 29). The direct repeat of nine amino acid residues is also encoded by one exon (exon 3). We estimated the size of each intron by PCR-based assay or sequence analysis (Fig. 3). Calculating from all these data, the size of the rat *Tsc2* gene is

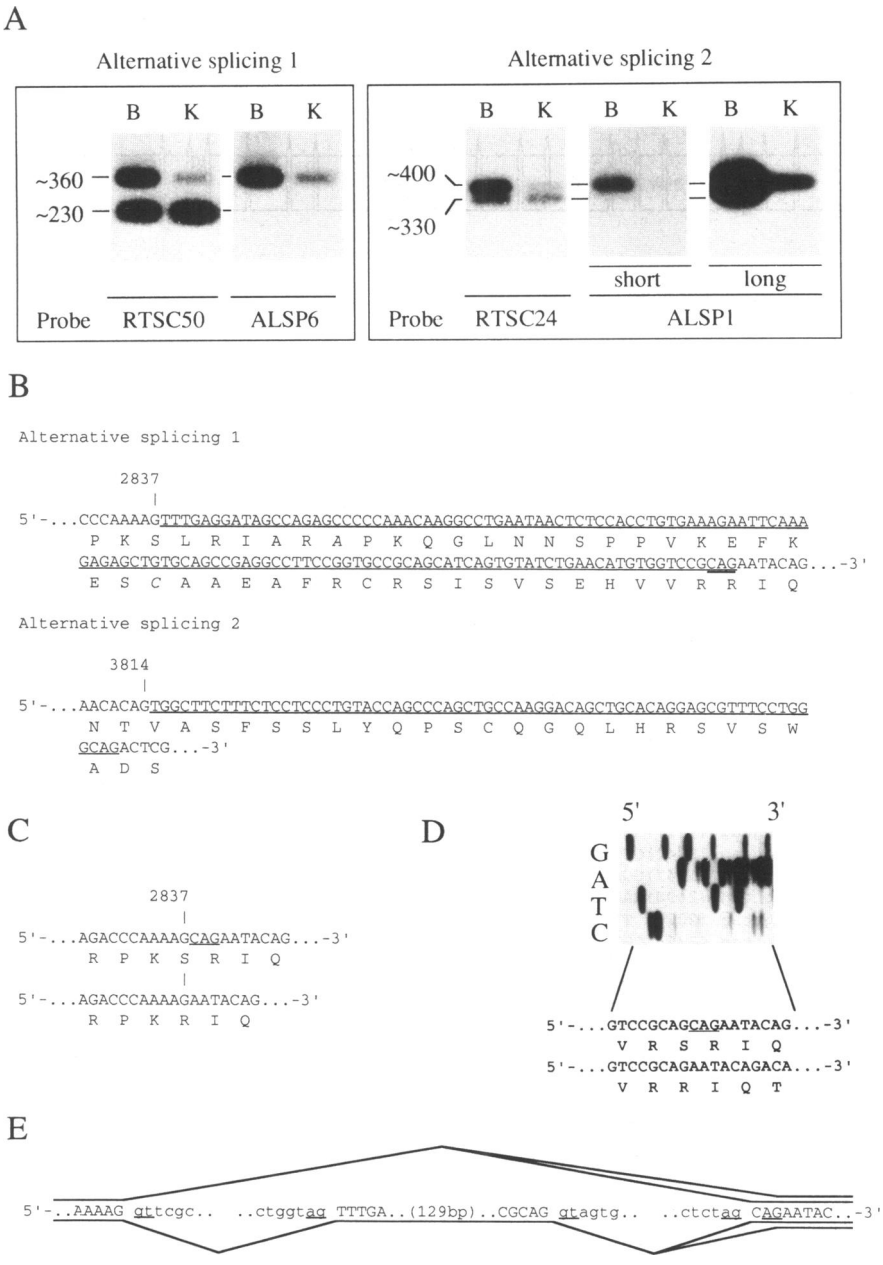


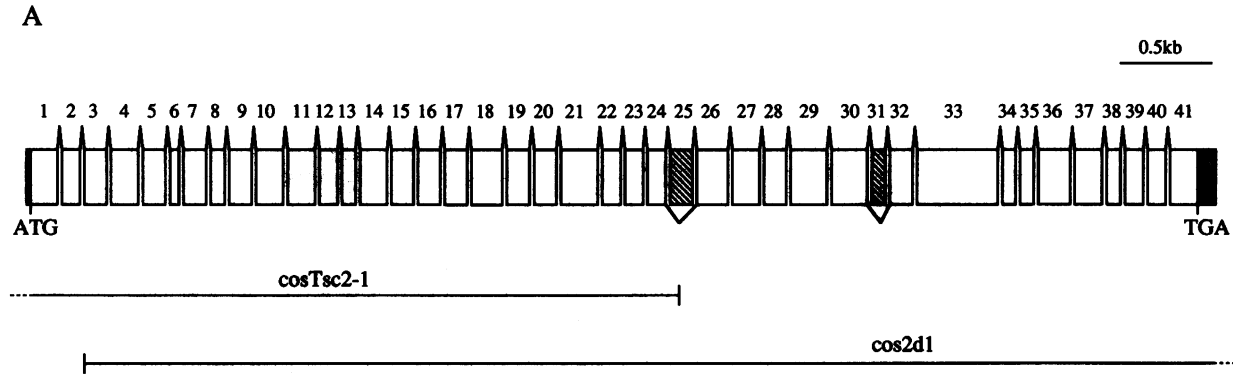
Figure 2. Alternative splicing of rat *Tsc2*. (**A**) RT–PCR using two primer sets, RTSC27/RTSC 28 (alternative splicing 1) and RTSC18/RTSC21 (alternative splicing 2). Lanes: B, brain and K, kidney. Probes used for Southern blot are indicated below each panel. ALSP6 and ALSP1 are specific to the alternatively spliced exon. RTSC50 and RTSC24 are the retained sequences in all variant spliced exons. Long and Short indicate the long exposure and short exposure of autoradiogram, respectively. Sizes (bp) of each band are shown on the left side. (**B**) The nucleotide and deduced amino acid sequences of two alternatively spliced sequences. The nucleotide numbers of insertion sites are shown above each sequence. Underlining denotes the alternatively spliced sequences. Bold underlining CAG is derived from exon 25, but not exon 26. For alternative splicing 1, amino acid (aa) residues different from human counterparts are indicated by italics. (**C**) The nucleotide and deduced amino acid sequences of two acceptor site variants from the lower band in RT–PCR for alternative splicing 1. The nucleotide number of the guanine residue before the exon junction site is shown above the sequence. The underlining denotes the residues subject to alternative acceptor site usage. (**D**) A result of direct sequence analysis of the upper band from RT–PCR for alternative splicing 2. Nucleotide and deduced amino acid sequences are shown below. The underlining denotes the nucleotide subject to alternative acceptor site usage. (**E**) Exon–intron organization and splicing events for alternative splicing 1. Exon and intron sequences are shown by upper and lower case letters, respectively. The splicing donor and acceptor sites are marked by underlining.

estimated to be ~35 kb without the promoter region. This value is similar to that (30–40 kb) estimated from results of genomic Southern blot analysis with *Tsc2* cDNA probes (T. K. *et al.*, unpublished result).

DISCUSSION

In the present study, we analyzed the structure of cDNAs including several alternative splicing variants and the exon–intron organization

A



B

5'- -3'

	(1) ATG..	Exon 1:	>138bp)..	TGAGA(138)	gtgagtgagc(1.8kbp)...
...tttcctcctttctag	(139)	GAACT..	Exon 2: 87bp)..	AAGAG(225)	gtgggtttgc(0.8kbp)...
...ttttcgttgaggcag	(226)	CATGC..	Exon 3: 111bp)..	GACAG(336)	gtaaggagct(0.7kbp)...
...tgtctttgtctttag	(337)	GGTGA..	Exon 4: 145bp)..	ACTGG(481)	gtaggtgtca(0.7kbp)...
...gcttttctcctacag	(482)	CAGAG..	Exon 5: 118bp)..	GTTCA(599)	gtaagtaaga(1.1kbp)...
...ttcttcttcctctag	(600)	CATGA..	Exon 6: 49bp)..	TCGAG(648)	gttagtaact(0.4kbp)...
...ccacccctcctgcag	(649)	GTGTC..	Exon 7: 126bp)..	GGAAG(774)	gtaggtgtcg(1.1kbp)...
...tgttttgtttttcag	(775)	CTGAT..	Exon 8: 74bp)..	AACAG(848)	gtgagtgtgt(1.1kbp)...
...ggggaacattttcag	(849)	ATCCT..	Exon 9: 127bp)..	ATGAG(975)	gtaaggagat(1.0kbp)...
...gtggccctgtttcag	(976)	GCCAT..	Exon 10: 144bp)..	TCCAG(1119)	gtaagtcaga(0.8kbp)...
...ctcttggtaccgtag	(1120)	AACCT..	Exon 11: 138bp)..	GGCCT(1257)	gtaaggagcc(0.4kbp)...
...gttttcctttgacag	(1258)	GAATC..	Exon 12: 104bp)..	TTCAG(1361)	gtagcggagc(0.3kbp)...
...ctgcttcttccccag	(1362)	GAACG..	Exon 13: 82bp)..	ACGAG(1443)	gtgtgcacgg(1.2kbp)...
...ttgtgattgatgcag	(1444)	GAGGA..	Exon 14: 156bp)..	AAAAG(1599)	gtgagagcca(0.5kbp)...
...ctggcctcttggaag	(1600)	GTGAT..	Exon 15: 117bp)..	TTCAG(1716)	gtgggtgctc(3.0kbp)...
...actctttctctgcag	(1717)	ACCAA..	Exon 16: 123bp)..	TGCAG(1839)	gtgagacacc(0.7kbp)...
...tttgctcccc ggcag	(1840)	GCCTT..	Exon 17: 107bp)..	TGCGC(1946)	gtatgtgtgg(0.3kbp)...
...catccgcctccacag	(1947)	GGAAC..	Exon 18: 151bp)..	AGCAG(2097)	gtgaggtggg(1.1kbp)...
...aaatccttcttctag	(2098)	GAGAC..	Exon 19: 123bp)..	CCATG(2220)	gtaacgcccc(0.8kbp)...
...tgtcttatatgttag	(2221)	CTTTC..	Exon 20: 135bp)..	GACAG(2355)	gtgagagagg(1.1kbp)...
...tcttgcctcccttag	(2356)	CGTGA..	Exon 21: 190bp)..	GTCTA(2545)	gtgagtatgt(0.7kbp)...
...gtccttgtgtcacag	(2546)	CCTTG..	Exon 22: 94bp)..	TCCAA(2639)	gtaagtctgt(203 bp)...
...ctcttcttccactag	(2640)	GTTCA..	Exon 23: 103bp)..	CTAAG(2742)	gttcgtccag(147 bp)...
...ctgctcctactccag	(2743)	GGTCT..	Exon 24: 95bp)..	AAAAG(2837)	gttcgcggtc(1.0kbp)...
...ttcctcgcctggtag	(2838)	TTTGA..	Exon 25: 129bp)..	CGCAG(2966)	gtagtggggc(1.3kbp)...
...ggtttctccctctag	(2967)	CAGAA..	Exon 26: 165bp)..	AAGAG(3131)	gtatgaggac(0.4kbp)...
...tttttactcttttag	(3132)	GTCCC..	Exon 27: 153bp)..	TCAAG(3284)	gtgaccgtcc(85 bp)...
...cacttctgtccttag	(3285)	CTCTG..	Exon 28: 113bp)..	GTCAG(3397)	gtaagactct(1.1kbp)...
...ctcccttgtttttag	(3398)	GGGGC..	Exon 29: 213bp)..	CACAG(3610)	gtactgactg(0.8kbp)...
...acccccta ccccccag	(3611)	GAAAC..	Exon 30: 204bp)..	CACAG(3814)	gtgagtgcag(0.8kbp)...
...gtggttaccttgcag	(3815)	TGGCT..	Exon 31: 69bp)..	GGCAG(3883)	gtattggcgc(1.2kbp)...
...tctgtgtcctctcag	(3884)	ACTCG..	Exon 32: 131bp)..	GCAGG(4014)	gtgagtgtca(0.3kbp)...
...aaattgtctttgcag	(4015)	TCTTC..	Exon 33: 491bp)..	CCTAG(4505)	gtgggcacaa(0.3kbp)...
...atgctttccctgcag	(4506)	CTTTG..	Exon 34: 76bp)..	ACGAG(4581)	gtaggccagc(194 bp)...
...tctgtgtgtctacag	(4582)	TCCTT..	Exon 35: 87bp)..	GCCAG(4668)	gtaaggccct(0.8kbp)...
...tggatttttccacag	(4669)	AGCAG..	Exon 36: 187bp)..	GCAAG(4855)	gtgtgtactg(152 bp)...
...tcacccaccctacag	(4856)	CTGTT..	Exon 37: 140bp)..	TTAAG(4995)	gtgagtgaga(0.7kbp)...
...ctctcgtctgggcag	(4996)	GGCCA..	Exon 38: 79bp)..	GAAAG(5074)	gtgaggccca(65 bp)...
...agtgtgtctccacag	(5075)	ACATG..	Exon 39: 92bp)..	CAAAT(5166)	gtgagtagtg(82 bp)...
...ctctgtcctctccag	(5167)	ATGGC..	Exon 40: 99bp)..	AGCGG(5265)	gtaggaagga(90 bp)...
...tttctctctctctag	(5266)	ATCCG..	Exon 41: >165bp)..	TGA (5430)	

Figure 3. The exon–intron organization of rat *Tsc2*. (**A**) Schematic representation of the *Tsc2* exons. The boxes with numbers denote the coding exons. Two striped boxes are alternatively spliced exons. Positions of the translational initiation codon (ATG) and stop codon (TAG) are marked below the first and last exon, respectively. The 5′ and 3′ non-coding regions found in the cDNA sequence are shown by filled areas. Lines under the scheme denote the areas covered by two overlapping cosmid clones. (**B**) Nucleotide sequences of the exon–intron boundaries of the *Tsc2*. Each line represents are exon (upper case latters) with the surrounding intron sequences (lower case letters). On both sides of the parentheses for the number and size of each exon, the first and last five nucleotides are shown. For exon 1 and exon 41, only the size of the coding regions is shown. The nucleotide numbers of the first and last residue of each exon, denoting the first 'A' of the initiation codon as 'nt 1', are shown in parentheses before and after them, respectively. The size of each intron is shown in parentheses on the right side.

of the rat *Tsc2* gene. The rat *Tsc2* product shows a 92% amino acid identity to the human counterpart. The GAP3 homologous domain is well conserved, suggesting that it might be functionally important. In the Eker rat, a germ line insertional mutation (intron 30) of the *Tsc2* gene causes structural alteration of the *Tsc2* transcripts (5). Such mutant transcripts encode truncated products without one third of the carboxy terminal part of the normal *Tsc2* product (5). The GAP3 homologous domain is missing from the

coding sequence in the mutant *Tsc2* transcripts. We previously reported allelic loss at the *Tsc2* gene locus in RCs and even in the earliest preneoplastic lesions, e.g. phenotypically altered tubules developing in hybrid rats carrying the Eker mutation (8,9). One plausible mechanism of tumor development in the Eker rat is loss of the GAP3 homologous domain. However, other unknown functions should also be considered. The observed strong conservation between rat and man should provide clues to establishing unknown gene functions other than that already predicted from the existence of the GAP3 homologous domain. Currently we are trying to identify a novel function of the *Tsc2* gene. The most diverged exon 29 might code a 'hinge' or 'spacer' region of the *Tsc2* protein. A germ line insertion occurred at the intron 30 in the Eker rat.

The rat *Tsc2* gene consists of 41 exons with relatively small sized introns and the length of rat *Tsc2* gene covering the entire coding exons is ~35 kb, although we can not exclude the possibility of a few alternative spliced exons or non-coding exons in the 5′ upstream region. It was also suggested that the human *TSC2* gene might be <50 kb (12). Based on the present data, we can design exon amplification primers for PCR–SSCP analysis. We are now investigating the second hit (intragenic mutations including point mutations) in LOH-negative tumors.

We have clearly demonstrated that there are two alternatively spliced exons (exons 25 and 31) in the rat *Tsc2* gene. Although we observed a discrete band at ~5.5 kb a Northern blot analysis using the *Tsc2* cDNA probe, it may be a mixture of several splicing variants (5). The sequence of exon 25 is conserved in man, although it is not known whether this homologous exon is also subjected to alternative splicing. A mouse *Tsc2* cDNA clone without the sequence homologous to exon 25 was reported, indicating that a similar alternative splicing event might occur in mice (18). In addition, there is a differential usage of splicing acceptor sites during the splicing reaction between exons 24 and 26, or exons 25 and 26. The meaning and the importance of this differential acceptor usage remains for future study. Another alternatively spliced exon (exon 31) is a novel exon as a homologous sequence has not been reported in the human *TSC2* gene. It is noteworthy that there might be at least four different types of *Tsc2* transcripts in the rat, produced by differential usage combinations of exons 25 and 31. The function of the *Tsc2* products may diverge between splicing variants. As noted from RT–PCR analysis, both tissue- and cell-type specific expression of these splicing variants may occur. We are now analyzing such specificity using antibodies against peptides encoded by each alternatively spliced exon.

Human tuberous sclerosis is an autosomal dominant genetic disease characterized by phakomatosis with manifestations that include mental retardation, seizures and angiofibroma. Thus, the same gene shows diverse phenotypes between species. Although human tuberous sclerosis is known to be associated with development of RCs, nothing is known to date about the molecular mechanism. The Eker rat therefore continues to be a valuable experimental model for elucidating *TCS2* gene role in renal carcinogenesis, as well as studying species-specific phenotypes.

ACKNOWLEDGEMENTS

We gratefully acknowledge Drs Alfred G. Knudson, Haruo Sugano and Tomoyuki Kitagawa for their encouragement throughout this work. We also acknowledge Dr T. Oda for providing a cosmid DNA library and Drs H. Tsuchiya and Y. Kubo for their help. This work was supported in part by Grants-in-Aid for Cancer Research from the Ministry of Education, Science and Culture of Japan and The Council for Tobacco Research of the USA.

REFERENCES

1 Eker, R. and Mossige, J. (1961) *Nature* (London) **189**, 858–859.
2 Yeung, R.S., Buetow, K.H., Testa, J.R. and Knudson, A.G. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 8038–8042.
3 Hino, O., Mitani, H., Nishizawa, M., Katsuyama, H., Kobayashi, E. and Hirayama, Y. (1993) *Jpn. J. Cancer Res.* **84**, 1106–1109.
4 Hino, O., Kobayashi, T., Tsuchiya, H., Kikuchi, Y., Kobayashi, E., Mitani, H. and Hirayama, Y. (1994) *Biochem. Biophys. Res. Commun.* **203**, 1302–1308.
5 Kobayashi, T., Hirayama, Y., Kobayashi, E., Kubo, Y. and Hino, O. (1995) *Nature Genet.* **9**, 70–74.
6 Yeung, R.S., Xiao, G-H., Jin, F., Lee, W-C., Testa, J.R. and Knudson, A.G. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 11413–11416.
7 Hino, O., Klein-Szanto, A.J.P., Freed, J.J., Testa, J.R., Brown, D.Q., Vilensky, M., Yeung, R.S., Tartof, K.D. and Knudson, A.G. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 327–331.
8 Kubo, Y., Mitani, H. and Hino, O. (1994) *Cancer Res.* **54**, 2633–2635.
9 Kubo, Y.,Klimek, F., Kikuchi, Y., Bannasch, P. and Hino, O. (1995) *Cancer Res.* **55**, 989–990.
10 Everitt, J. I., Goldsworthy, T.L., Wolf, D.S. and Walker, C.L. (1992) *J. Urol.* **148**, 1932–1936.
11 Hino, O., Mitani, H., Katsuyama, H. and Kubo, Y. (1994) *Cancer Lett.* **83**, 117–121.
12 The European Chromosome 16 Tuberous Sclerosis Consortium. (1993) *Cell* **75**, 1305–1315.
13 Hino, O., Kitagawa, T. and Sugano, H. (1985) *Int. J. Cancer* **35**, 5–10.
14 Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Plainview, NY.
15 Oda, T., Funai, T. and Ichiyama, A. (1990) *J. Biol. Chem.* **265**, 7513–7519.
16 Sanger, F., Nicklen, S. and Coulson, A.A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
17 Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nucleic Acid Res.* **12**, 387–395.
18 Olsson, P.G., Southerland, H.F., Scneider, M.C., Reeders, S.T. and Frischauf, A.M. (1994) *Ann. Hum. Genet.* **58**, 230–231.