Oncogene (1998) 17, 2761–2770

http://www.stockton-press.co.uk/onc

Genomic architecture and transcriptional activation of the mouse and human tumor susceptibility gene *TSG101*: Common types of shorter transcripts are true alternative splice variants

Kay-Uwe Wagner[1], Patricia Dierisseau[1], Edmund B Rucker III[1], Gertraud W Robinson[1] and Lothar Hennighausen*[,1]

[1]*Laboratory of Genetics and Physiology, National Institute of Diabetes, Digestive, and Kidney Diseases, National Institutes of Health, Building 8, Room 101, Bethesda, Maryland 20892-0822, USA*

The functional inactivation of the tumor susceptibility gene *tsg101* in mouse NIH3T3 cells leads to cell transformation and the formation of metastatic tumors in nude mice. We cloned, mapped and sequenced the mouse *tsg101* gene and further identified a processed pseudogene that is 98% identical to the *tsg101* cDNA. Based on Northern blot analysis, *tsg101* is expressed ubiquitously in mouse tissues. A comparison of the coding region of the mouse *tsg101* gene with the human *TSG101* cDNA revealed that both the mouse and human gene encode ten additional highly conserved amino acids at the N-terminus. Based on the mouse *tsg101* genomic structure, we predicted four additional introns within the human *TSG101* gene. Their location was confirmed using PCR and sequencing analysis. The presence of these so far unidentified introns now explains published data on aberrantly spliced mRNA products that were frequently observed in primary breast tumors. We show that a majority of shorter *TSG101* transcripts are not the result of aberrant splicing events, but represent a fraction of true alternative splice variants. Finally, we examined *tsg101* expression patterns during different stages of mammary gland development and in different transgenic mouse models for breast tumorigenesis.

Keywords: TSG101 protein; mice; human; genomic DNA; pseudogene; mammary gland

Introduction

The tumor susceptibility gene 101 (*tsg101*) was discovered in mouse NIH3T3 fibroblasts using a random homozygous knockout approach (Li and Cohen, 1996). The functional inactivation of this gene leads to cell transformation *in vitro* and to metastasizing tumors *in vivo* when transformed mouse fibroblasts were transplanted into nude mice. Reversal of the neoplastic properties was obtained *in vitro* after deleting the gene that transactivated the antisense promoter using Cre-lox recombination. These results suggested that transformation and tumorigenesis were a direct consequence of TSG101 deficiency (Li and Cohen, 1996).

The *tsg101* cDNAs in mouse and human are 86% identical on the nucleotide level (Li *et al*., 1997). They both encode predicted proteins of about 43 kDa in size that show 94% similarity. On the subcellular level, the TSG101 protein is localized mainly in the cytoplasm (Zhong *et al*., 1997) but depending on the stage of the cell cycle it can be detected in the nucleus and it can co-localize with the mitotic spindle apparatus during cell division (Xie *et al*., 1998). Another indication that TSG101 may have different functions at specific stages of the cell cycle is the presence of highly conserved motifs, such as a coiled-coil domain that is proposed to interact with stathmin (Maucuer *et al*., 1995) and a proline-rich region known to exist in activation domains of transcription factors (Li *et al*., 1997). Furthermore, an N-terminal region is similar to the catalytic domain of ubiquitin-conjugating enzymes suggesting a potential role of TSG101 as a regulator in ubiquitin-mediated protein degradation (Koonin and Abagyan, 1997; Ponting *et al*., 1997). The mechanism underlying cell transformation after inactivation of both *tsg101* alleles is not known. Potentially as a regulator of protein degradation, TSG101 could influence the half-life of other important tumor suppressors and regulatory factors of the cell cycle. Since inactivation of TSG101 was observed to be associated with abnormal microtubule organizing centers and nuclear anomalies, it is suggested that the absence of the TSG101 protein or the lack of individual TSG101 domains may lead to chromosomal instability and, thereby to a progression of tumorigenesis (Xie *et al*., 1998).

The human *TSG101* gene has been mapped to a specific chromosomal region on chromosome 11, p15.1-p15.2 (Li *et al*., 1997), that is associated with a loss of heterozygosity in different types of tumors such as breast cancer (Ali *et al*., 1987) and Wilms' tumor (Reeve *et al*., 1989). However, rearrangements and somatic mutations within the *TSG101* gene are rare events in human breast cancers (Steiner *et al*., 1997; Gayther *et al*., 1997; Lee and Feinberg, 1997; Wang *et al*., 1998). Instead of gene alterations, aberrant *TSG101* splice forms have been correlated to cell transformation (Gayther *et al*., 1997; Sun *et al*., 1997; Lee and Feinberg, 1997). In order to identify the structural basis of the aberrant transcripts, the intron-exon architecture of the *TSG101* gene needs to be determined. This information and the identification of regulatory elements within the promoter will provide new insights into *TSG101* transcriptional regulation, alternative and aberrant mRNA splicing events, resulting protein variants, and gene function in normal and transformed tissue.

*Correspondence: L Hennighausen
Received 19 August 1998; revised 15 October 1998; accepted 26 October 1998

Results

Genomic organization of the mouse tsg101 *gene*

The mouse *tsg101* gene was cloned as multiple overlapping restriction fragments from two individual BAC genomic clones. Briefly, a cDNA that was generated by RT–PCR from mammary tissue of a lactating dam was used to probe a 129/SvJ genomic BAC library (see Material and methods). Six positive BAC clones were analysed and five contained a novel processed *tsg101* pseudogene but not the actual *tsg101* gene. In order to distinguish the true gene from the pseudogene, we designed a PCR assay that amplified intron sequences (see Materials and methods). The sixth BAC clone contained the genuine *tsg101* gene. Since this BAC clone lacked the 5′ part of the *tsg101* gene a second overlapping BAC clone was isolated. Overlapping subclones were sequenced and the size of the mouse *tsg101* gene was determined to be 33.6 kb (GenBank AF060868). Furthermore, we identified additional introns, which had not been predicted from mapping studies (Li *et al*., 1997). In addition, we determined differences in the coding region of the mouse *tsg101* gene to the previously published mouse cDNA sequence (Li and Cohen, 1996). These differences lead to an expansion of the *tsg101* reading frame that encodes ten additional amino acids at the N-terminus, which are highly similar to the predicted human TSG101 protein.

The murine *tsg101* gene contains ten exons (Figure 1). Sequences of intron-exon junctions are summarized in Table 1. Evidently, the larger number of exons and introns of the murine *tsg101* gene compared to the human homologue is the result of four additional introns within a region considered to be the first exon of the human gene (Li *et al*., 1997). The first exon includes the 5′ untranslated sequence and the translation initiation codon. With 329 bp, exon 10 is the largest exon, but it contains almost 75% of its nucleotides (240 bp) as 3′ untranslated sequence. A comparison of the coding region to the previously published mouse cDNA sequence (Li and Cohen, 1996) revealed several differences. An insertion of two G residues between position 37 and 38 of the published mouse cDNA shifts the first ATG codon in position 33 in frame with the *tsg101* coding region. This results in the expansion of the *tsg101* open reading frame of exactly 30 nucleotides encoding ten additional amino acids at the N-terminus (Met-Ala-Val-Ser-Glu-Ser-Gln-Leu-Lys-Lys). The hypothesis that the ATG at position 33 is the preferred translational start site is supported by the fact that this ATG codon is flanked by a C-rich sequence motif in position −6 to −1 and a highly conserved G in positon +4 (Kozak, 1997). A mismatch at position 81 (T to C) is similar to the human sequence and did not result in any amino acid substitution. Some additional changes of single nucleotides were observed in the 3′ untranslated region immediately following the translational stop codon.

A prominent part of the TSG101 protein, an α-helix domain that forms a coiled-coil structure, was reported to be identical to CC2 and has been proposed to interact with stathmin (Maucuer *et al*.,

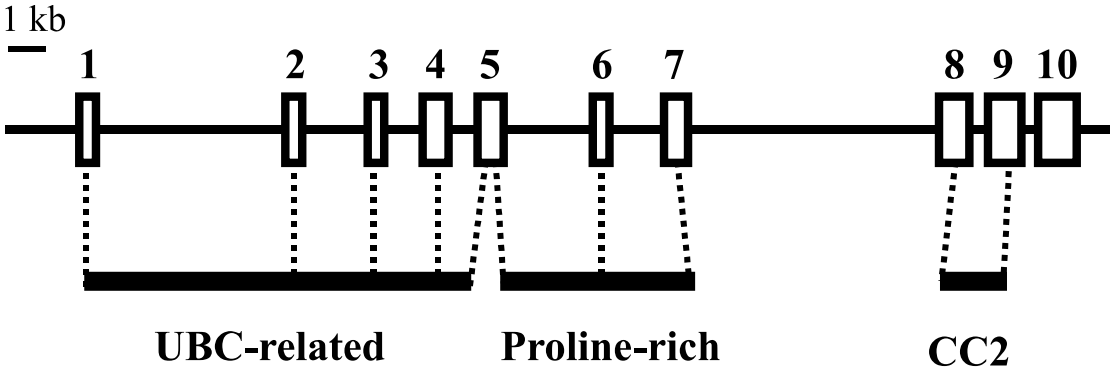

Figure 1 Genomic structure of the mouse *tsg101* gene. The numbered boxes illustrating exons 1–10 are not in scale. Solid bars represent known encoded protein domains

Table 1 Nucleotide sequence of intron-exon boundaries of the mouse *tsg101* gene

3′ Acceptor Sequence				*5′ Donor Sequence*			
Intron	*Exon*	*Exon No.*	*Size (bp)*	*Exon*	*Intron*	*Intron No.*	*Intron Size (kb)*
	CCCTCTGCCT	1	72	GATGTCCAAG	gtgagcccgc	1	6.3
tgcttttcag	TACAAATACA	2	84	GATTCATATG	gtaagtttat	2	2.6
tccctttag	TTTTTAATGA	3	65	CGTTATCCGAG	gtaaataata	3	1.7
tctctcttag	GTAATATATA	4	163	CTGGAAACAT	gtaagtatta	4	1.8
tccttttag	CCACGGTCAG	5	126	CCACCAAATA	gtaagtacac	5	3.5
ttatttttag	CCTCCTACAT	6	66	CCAACCCCAG	gtaatgaaaa	6	2.4
ttttctatag	TGGTTATCCT	7	91	ACCACTGTTG	gtaagtataa	7	8.6
tctcctgcag	GTCCAGCAG	8	202	TCAAGAAGTA	gtaagtaagc	8	1.2
ctgttttcag	GCTGAAGTTG	9	239	GTTCCTGAAA	gtaagtaccc	9	1.2
tccttcctag	CACGTCCGCT	10	329	TGTTGCATAA			

1995). This particular domain is encoded by two of the largest coding exons of *tsg101*, exon 8 and 9 (Figure 1). The proline rich region, which is known to be characteristic for transcription factors, is entirely encoded by exons 5–7. Both groups of exons encoding the α-helix domain and proline rich region are separated by the largest intron of more than 8 kb. Recently, it was reported that the N-terminal region of TSG101 has a sequence similarity to the catalytic domain (UBC) of ubiquitin-conjungating enzymes E2 and UBC-related DNA-binding proteins (Koonin and Abagyan, 1997; Ponting *et al*., 1997). Nucleotide sequences encoding conserved amino acid residues of TSG101 and other E2 derivatives are scattered within the first five exons.

A 2.9-kb fragment spanning the promoter/upstream was sequenced. Sequences surrounding the putative transcriptional start site (Li and Cohen, 1996) have features characteristic of a house-keeping gene promoter. It lacks both TATA and CAAT boxes, has a high GC content, and contains several potential Sp1 and AP2 consensus sites (Figure 2). The *tsg101* 5′ region also contains putative sites for nuclear factor GAPBF2 known to exist in promoters of other ubiquitously expressed genes (Aki *et al*., 1997).

The mouse genome carries a tsg101 *pseudogene*

Initially, we probed a lambda phage library containing 129SVJ mouse genomic DNA with a cDNA probe (see Materials and methods) and isolated ten phage clones. The presence of the *tsg101* gene within the recombinant phage DNA was tested by PCR using primers from the 3′-untranslated region as described by Li *et al*. (1997). Inserts of two PCR positive lambda clones were sequenced. The resulting sequence (GenBank AF060867) was 98% identical to the mouse *tsg101* cDNA and did not contain any introns suggesting the presence of a pseudogene. The processed pseudogene differs in a few significant mutations from the murine cDNA. The AT-deletion at position 153 results in a frame shift and a precocious termination of the open reading frame. The presence of the pseudogene in other lambda phages could be easily detected by PCR and subsequent *Nde*I restriction digest. The AT-deletion in the pseudogene eliminates the *Nde*I recognition site

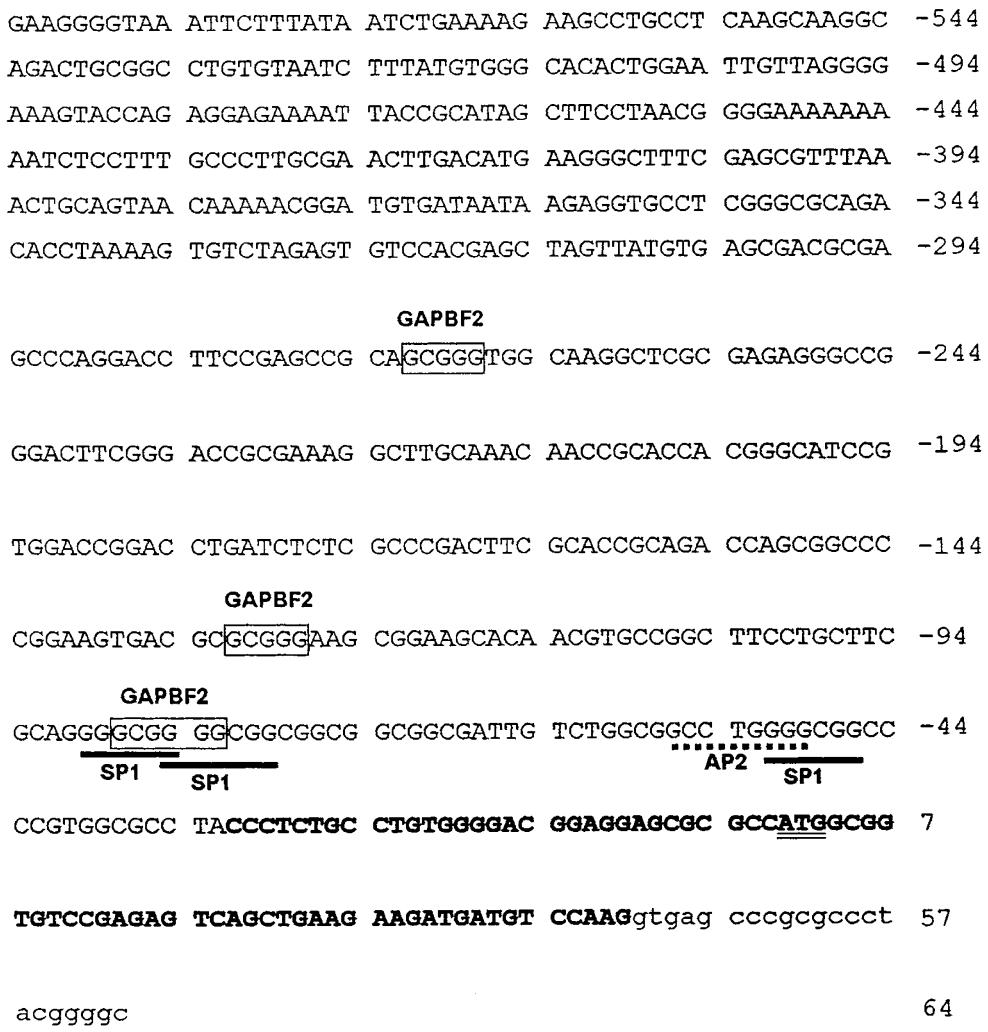

Figure 2 Nucleotide sequence of the mouse *tsg101* 5′-flanking region. The transcriptional start site (Li and Cohen, 1996) and exon 1 is indicated in bold and upper case; the following intron sequence is shown in normal lower case. The newly suggested translation start codon (ATG) is double underlined. Nucleotides numbering on the right refers to the distance in bp to the translation start codon. Locations of putative responses elements in the 5′-flanking region are boxed (GAPBF2) and underlined (SP1 solid line; AP2 dashed line)

present in the mouse *tsg101* cDNA. Interestingly, a downstream ACCC-deletion at position 269 shifts an ATG codon at position 246 into frame with the rest of the original *tsg101* reading frame. Since the 5′-flanking region of the pseudogene shows no similarity to the promoter region of the *tsg101* gene and secondly, the ATG does not contain any consensus sequence for initiation of translation (Kozak, 1997), it is very unlikely that the pseudogene is expressed. To identify potential *tsg101* pseudogene transcripts, we screened 222 mouse *tsg101* ESTs using BLAST 2.0. None of the ESTs carried any characteristic mutations for the pseudogene. Moreover, we performed a RT–PCR amplification of a 300 bp region flanking the *Nde*I recognition site that is specific for the actual *tsg101* gene. The RT–PCR product was completely digested with *Nde*I restriction enzyme into expected smaller fragments (data not shown). Taken together, there is no experimental evidence that the pseudogene is expressed.

Determination of the structure of the human TSG101 *gene*

We have cloned and sequenced the entire mouse *tsg101* gene. The mouse *tsg101* gene differs not only in size, but also in the number of exons and introns, and intron sizes from the architecture of the human gene, which had been mapped by restriction analysis (Li *et al*., 1997). To compare the location of intron sites between the mouse and human *TSG101* gene, we aligned the coding sequence of the mouse gene to the human *TSG101* cDNA (Li *et al*., 1997). Sequences flanking the exon junctions of the mouse *tsg101* gene and the corresponding nucleotides of human cDNA are illustrated in Table 2. The results show a very high similarity on the nucleotide level in this particular region and the first nucleotide of subsequent exons is entirely conserved. Based on these data we could predict the location of introns within the human *TSG101* gene at a level of a single nucleotide in areas that were predicted earlier to contain intron sequences (Li *et al*., 1997). In addition, we determined four additional introns in the sequence considered being the first exon. We experimentally tested the presence of those four additional introns by PCR amplification of human genomic DNA using primers corresponding to the human coding sequence, which flank the anticipated intron sites. The PCR amplification products are shown in Figure 3. As predicted, the human gene also contains four additional introns. We also amplified previously predicted intron sequences (Li *et al*., 1997) to accurately determine their location. The PCR fragments were gel purified and sequenced, and the sequences of all intron-exon junctions within the human *TSG101* gene are summarized in Table 3. As predicted from the alignment of the mouse and human cDNAs, the intron-exon boundaries are conserved between both species on the level of a single nucleotide (compare Tables 2 and 3). We were not able to amplify intron 7 from human genomic DNA probably because of its larger size similar to the mouse counterpart (Figure 1). Taken together, not only the mRNA and amino acid sequences of the mouse and human TSG101 protein are highly conserved, but also the entire structure of the *TSG101* gene is similar between both species. This supports earlier suggestions (Li *et al*., 1997) that the mouse and human *TSG101* gene are true gene homologues and that they must have fundamental biological functions in common.

Table 2 Comparison of mouse (m) and human (h) *TSG101* cDNA sequences encompassing exon junctions that where determined by sequencing of the murine *tsg101* gene. Boxed nucleotides correspond to the first base of subsequent exons

Exon-junction	*Species*	*Sequence*	*Location*
1 – 2	m	AAG[T]AC	73
	h	AAG**T**AC	133
2 – 3	m	ATG[T]TT	158
	h	ATG**T**TT	218
3 – 4	m	GAG[G]TA	224
	h	GAG**G**TA	284
4 – 5	m	CAT[C]CA	388
	h	CAc**C**CA	448
5 – 6	m	ATA[C]CT	515
	h	ATA**C**tT	572
6 – 7	m	CAG[T]GG	582
	h	CAG**T**GG	639
7 – 8	m	TTG[G]TC	674
	h	TTG**G**TC	731
8 – 9	m	GTA[G]CT	877
	h	GTA**G**Cc	934
9 – 10	m	AAA[C]AC	1117
	h	AAg**C**At	1174

Expression of the mouse tsg101 *gene in various organs and the developing mammary gland*

Sequencing data of the 5′ flanking region of the mouse *tsg101* gene exhibited several features that indicate that this gene might be constitutively expressed in many tissues. To test this hypothesis we analysed the expression of *tsg101* in several organs of adult mice (Figure 4a). Total RNA was analysed using a ^{32}P-labeled cDNA. *Tsg101* was highly expressed in all organs examined. The highest expression was observed in brain and mammary tissue and the lowest expression was detected in liver. These Northern blot data support our hypothesis that *tsg101* might be ubiquitously expressed in all mouse organs and confirm earlier studies in various human tissues (Li *et al*., 1997).

It was suggested that the *tsg101* gene is involved in cell growth, differentiation and mammary neoplasia (Li *et al*., 1997; Xie *et al*., 1998). To explore the role of *tsg101* in normal breast development, we established its expression profile at different stages of mammogenesis (Figure 4b). *Tsg101* was expressed highly at all stages of mammary gland development. A slight downregulation at the transcriptional level was detected at early and mid pregnancy. This phase of mammary gland development is characterized by massive epithelial cell proliferation.

Tsg101 *transcription in primary breast tumors of transgenic mouse models*

A reduced amount of full-length *TSG101* transcripts and the presence of truncated mRNA forms have been

correlated to various stages of tumorigenesis in human mammary tissue (Gaythe *et al*., 1997; Lee and Feinberg, 1997; Li *et al*., 1997) and other malignancies (Sun *et al*., 1997). We have now analysed *tsg101* transcription in established transgenic mouse models for breast cancer to address the question, whether *tsg101* expression is altered in primary tumors. Biopsies of mammary tumors were taken from transgenic mice expressing different oncogenes under the control of mammary specific regulatory elements of the *whey acidic protein* (*WAP*) gene. Specifically, tissue was harvested from mice expressing transgenic Int3, SV40-TAg and TGFα. Total RNA was isolated and analysed on Northern blots (Figure 5). Full-length *tsg101* trancripts were found in all tumor samples. The level of *tsg101* transcription was not altered in tumors from WAP-TAg and WAP-TGFα transgenic mice. In contrast, a sharp reduction of *tsg101* expression was observed in WAP-int3 mammary tumors. A comparison of *tsg101* expression in solid tumors and adjacent mammary tissue of the same animal (Figure 5, lanes 2–5) demonstrated reduced levels of *tsg101* mRNA in Int3 tumors (64 and 36%, respectively). We were not able to detect smaller truncated *tsg101* transcripts (data not shown), which suggests that the decreased amount of full length mRNA is not the result of aberrant or alternative splicing.

Discussion

The human *TSG101* gene has been mapped on chromosome 11 band p15 (Li *et al*., 1997), a region that is known to be associated with a loss of heterozygosity (LOH) in human breast cancers (Ali *et al*., 1987) and Wilms' tumors (Reeve *et al*., 1989). Aberrant *TSG101* transcripts have been detected in a number of primary human breast cancers and other tumors (Li *et al*., 1997; Gaythe *et al*., 1997; Lee and Feinberg, 1997; Sun *et al*., 1997), but intragenic deletions of the *TSG101* gene are rare events in human malignancies (Steiner *et al*., 1997; Lee and

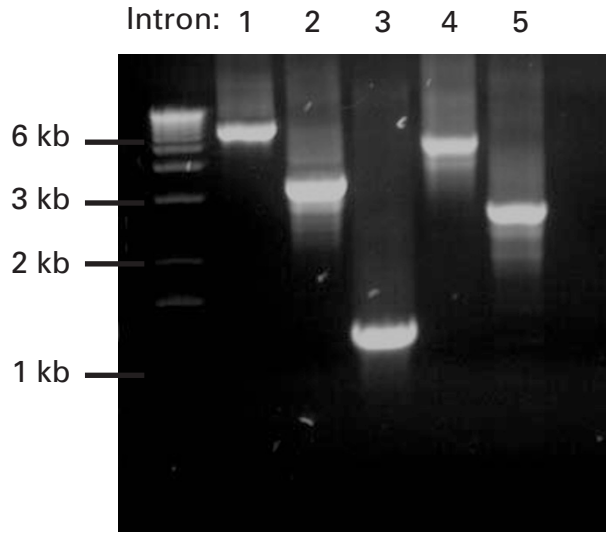

Figure 3 PCR amplification of introns 1–5 of the human *TSG101* gene. The analysis was performed on human genomic DNA using exon specific primers that encompass individual

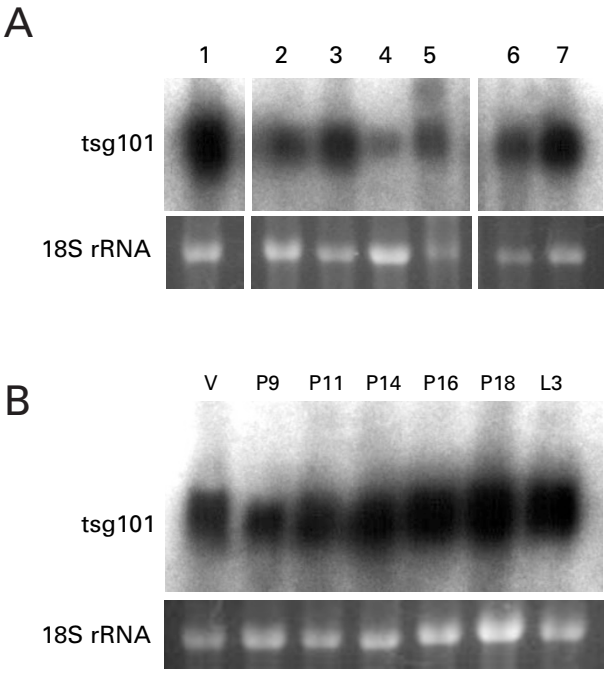

Figure 4 Expression of the mouse *tsg101* gene in various organs (**a**) and the developing mammary gland (**b**). Twenty μg of total RNA was separated in 1.5% agarose gels containing 18% formaldehyde, transferred to a membrane and hybridized with a ^{32}P-labeled *tsg101* cDNA probe. (**a**) The *tsg101* mRNA was present in brain (lane 1), muscle (lane 2), heart (lane 3), liver (lane 4), salivary gland (lane 5), kidney (lane 6), and mammary gland (lane 7). (**b**) The abbreviations are v, virgin; P, days of gestation; L, days of lactation

Table 3 Nucleotide sequence of intron-exon boundaries of the human *TSG101* gene

3′ Acceptor Sequence			*5′ Donor Sequence*			
Intron	*Exon*	*Exon No.*	*Exon*	*Intron*	*Intron No.*	*Intron Size (kb)*
		1	GGTGTCCAAG	gtgaggctgc	1	~7.0
tgcttttcag	TACAAATACA	2	GATTCATATG	gtgagtttat	2	~3.3
ttcctttcag	TTTTTAACGA	3	CCTTATAGAG	gtaaatgtct	3	~1.4
tctttttcag	GTAATACATA	4	ATGGAAACAC	gtaagtattc	4	~5.5
tcatttttag	CCACAGTCAG	5	CCACCAAATA	gtaagtagaa	5	~2.7
atgtttttag	CTTCCTACAT	6	CCAATCCCAG	gtaatgaaaa	6	~4.5
ctgtctttag	TGGTTACCCA	7	ACCACTGTTG	ND[a]	7	ND[a]
ND[a]	GTCCCAGTAG	8	TCAAGAAGTA	gtaagtgact	8	~2.3
ctgttttcag	GCCGAGGTTG	9	CTTCCTGAAG	gtatttcttc	9	~1.2
cccctcccag	CATGTACGTC	10				

[a]ND, not determined

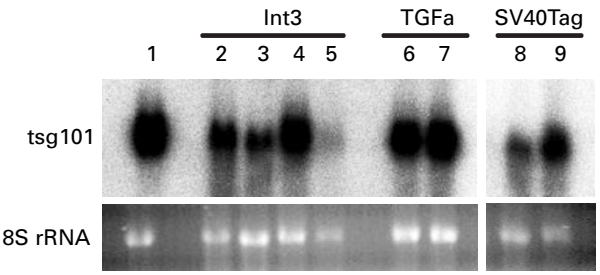

Figure 5 *Tsg101* expression levels in tumors from transgenic mouse models for breast cancer. Mammary tissue of a non-transgenic mouse (lane 1), tumors (lanes 3 and 5) and adjacent mammary tissue (lanes 2 and 4) of WAP-int3 transgenic mice and tumors from WAP-TGFα (lanes 6 and 7) and WAP-SV40Tag (lanes 8 and 9) were analysed. Twenty mg of total RNA was separated in 1.5% agarose gels containing 18% formaldehyde, transferred to a membrane and hybridized with a ^{32}P-labeled *tsg101* cDNA probe

Feinberg, 1997; Wang *et al*., 1998). It was proposed that shorter transcripts are caused by abnormal splicing events recognizing cryptic splicing donor and acceptor sites within exons (Gayther *et al*., 1997; Lee and Feinberg, 1997). Based on our analysis of the mouse and the human *TSG101* genes we suggest that many of the reported aberrant transcripts are in fact alternative splice forms. We show that part of the *TSG101* coding sequences, which had been described as the first exon, actually spans four exons, and that exon skipping results in different RNA forms, which had been described in tumors and normal tissues of different origin.

How similar are the human and mouse TSG101 *genes?*

We have cloned and sequenced the entire murine *tsg101* gene. A comparison of the mouse genomic sequence with the map of the human counterpart (Li *et al*., 1997) revealed differences in size and in the number of exons. The mouse gene consists of ten exons and spans about 33.6 kb. In contrast, the human gene cloned from a PAC library was estimated to be much smaller in size and to contain six exons (Li *et al*., 1997). The first exon of the human *TSG101* gene was represented by five exons in the mouse. A comparison of the mouse *tsg101* coding region with the corresponding human cDNA illustrated that the junctions of all exons in the mouse gene were preserved in the human cDNA, suggesting that the human gene may also contain additional exons. Indeed, we cloned and verified the existence of four additional introns in the human gene. In addition to those four introns, we amplified earlier predicted introns (Li *et al*., 1997), determined their sizes and demonstrated their accurate location.

A reexamination of the human *TSG101* cDNA (Li *et al*., 1997) with the revised mouse sequence revealed an ATG start codon 30 nucleotides 5′ of the previously published translational initiation site. Our data suggest that the human TSG101 protein contains the same ten additional N-terminal amino acids. The expansion of the open reading frame of the mouse and human *TSG101* gene was found independently by Li and coworkers and the database entries in GenBank were updated (U52945 and U82130) and match with our genomic sequence.

Taken together, not only the mRNA and amino acid sequences of the mouse and human TSG101 protein are highly conserved but also the entire structure of the *TSG101* gene is identical in both species.

Aberrant vs *alternative splicing*

Several reports demonstrated the presence of aberrantly spliced *TSG101* transcripts in tumor specimen (Gayther *et al*., 1997; Lee and Feinberg, 1997). Based on our findings that the human *TSG101* gene contains additional introns, we re-analysed published data about aberrant splice products. We suggest that many truncated transcripts frequently observed in tumors and nonmalignant tissues are true alternative splice products, rather than aberrant transcripts. For instance, type B (133–731), C (133–639) and D (133–448) deletions observed in primary breast tumors (Lee and Feinberg, 1997) are true alternative splice products (Figure 6). All three alternative splice variants were also reported as aberrant transcripts to occur in human malignancies other than primary breast tumors (Gayther *et al*., 1997) and normal tissues of different origin (Gayther *et al*., 1997; Lee and Feinberg, 1997). The most frequently observed alternative splice variant in normal human tissues (about 50% of all abnormal variants) is the 133–448 deletion (Gayther *et al*., 1997; Lee and Feinberg, 1997). Interestingly, this is the only type of alternative splice product that would result in an in-frame deletion. If translated, this particular mRNA would create a short TSG101 protein lacking domains encoded by exons 2, 3 and 4, which represent a considerable portion of conserved amino acid residues to the catalytic domain (UBC) of ubiquitin-conjungating enzymes E2 and UBC-related DNA-binding proteins. It needs to be determined whether such a smaller TSG101 protein variant can interfere with the potential role of TSG101 as a regulator in ubiquitin-mediated protein degradation or other specific functions.

Despite the presence of alternative splice variants of *TSG101* transcripts, some reported shorter transcripts are probably real aberrant splice forms. For instance, a frequently observed aberrant *TSG101* mRNA contains a 153–1053 deletion (Gayther *et al*., 1997; Lee and Feinberg, 1997; Sun *et al*., 1997). We confirmed by sequencing and by PCR using primers flanking both cryptic splicing donor and acceptor sites within exons that no extra introns are located at those positons in exon 2 and exon 9 of the human *TSG101* gene (data not shown). Therefore, the 153–1053 deletion in the human *TSG101* mRNA appears to be a real aberrant splice product. It does not encode any known functional domain of the TSG101 protein and, therefore, its biological role during tumorigenesis is not clear. Moreover, this particular mutation is not restricted to neoplastic tissues since it is also found in normal breast tissue and various other organs (Lee and Feinberg, 1997; Gayther *et al*., 1997).

Taken together, our findings suggest that shorter transcripts are caused by splicing events at either existing intron–exon junctions (alternative splicing) or cryptic splice donor and acceptor sites within exons (aberrant splicing). Therefore, our data invalidate the

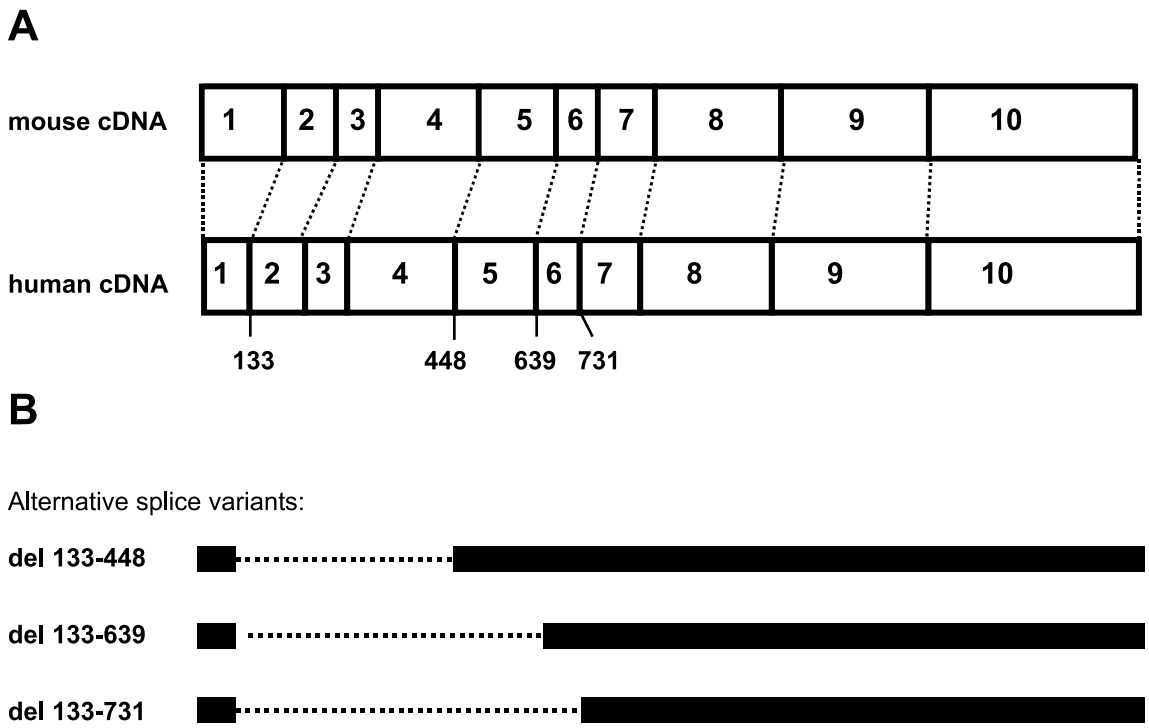

Figure 6 Alignment of the mouse and human *TSG101* cDNAs (**a**) and alternative splice variants of human *TSG101* mRNA (**b**) detected in various malignancies. (**a**) Numbered boxes indicate individual exons of the coding sequence that were determined by sequencing. (**b**) Alternative splice variants of the human tsg101 mRNA frequently observed in normal tissues and tumors were summarized from earlier reports

hypothesis that anomalous transcripts are pure PCR artifacts as proposed by Hampl *et al*. (1998). It remains to be established if alternative splice variants of *TSG101* encode functional proteins, and if such shorter proteins can interfere with the function of the wildtype TSG101 protein.

Does the human genome possess TSG101 *pseudogenes?*

We have cloned and sequenced a processed *tsg101* pseudogene from a mouse genomic library that is 98% identical to the murine *tsg101* cDNA. A Southern blot examination of human genome DNA with various short cDNA probes revealed an unusual fragment pattern (Steiner *et al*., 1997) that was different from the map provided earlier (Li *et al*., 1997). These results do not only indicate that the genomic map is incomplete because of missing introns, but it could also suggest that there are pseudogenes in the human genome. However, based on a PCR assay multiple exon-specific primers to amplify intron sequences (Figure 3) we did not identify a human *TSG101*-derived processed pseudogene, which would result in very short PCR fragments. Whereas pseudogenes do not interfere with the analysis of genomic sequences in human samples, the existence of pseudogenes in the mouse genome has profound implications on the design of assays to detect *tsg101* gene deletions and gene expression patterns by PCR, the cloning of new genes with sequence similarities to *tsg101*, and mutational analyses of the *tsg101* locus to study the gene function *in vivo*.

Transcriptional activation of the tsg101 *gene in various tissues and the developing mammary gland*

The 5′ flanking region of the *tsg101* gene exhibited features of housekeeping gene promoters, such as a high GC content and potential Sp1, AP2 and GAPBF2 consensus sites. Initial expression studies on mRNA from human tissues revealed that *TSG101* is expressed in a variety of organs (Li *et al*., 1997). All mouse tissues we examined contained a *tsg101* transcript of about 1.2 kb in size, which support the hypothesis that *tsg101* is expressed ubiquitously. While brain and mammary gland show the highest expression levels, less *tsg101* mRNA was detected in liver. Interestingly, the high amount of *tsg101* transcripts in brain correlates with the expression profile of stathmin mRNA. Stathmin, which is suggested to interact with the α-helical (cc2) domain of the TSG101 protein, is most abundant in brain and neurons (Ozon *et al*., 1997). The hypothesis that *tsg101* might be essential for common functions in a living cell is supported by the fact that this gene is expressed throughout development. A search for sequence similarities of the full length mouse *tsg101* cDNA in the mouse ESTs database (BLAST 2.0) indicate that *tsg101* transcripts are already present in 1-cell and 2-cell stage embryos.

From *in vitro* data it has been suggested that *tsg101* is involved in cell growth and differentiation (Xie *et al*., 1998). *Tsg101* is expressed at all stages of mammogenesis regardless of their proliferative and differentiation status. A slightly reduced transcriptional level was detected at early to mid-pregnancy, a time of massive

cell proliferation. A downregulation of *tsg101* at the peak of alveolar proliferation would support the hypothesis that the TSG101 protein might be a negative growth regulator (Watanabe *et al*., 1998). Indeed, a strong overexpression of this molecule *in vitro* leads to the inhibition of cell division and, eventually, cell death (Xie *et al*., 1998; Wagner, unpublished). Nevertheless, the reduction of *tsg101* transcripts *in vivo* in the developing mammary gland is moderate. Our observations support earlier findings *in vitro* (Xie *et al*., 1998) that the level of *tsg101* expression is independent from the stage of the cell cycle. Therefore, it can be assumed that differences in the activity of *TSG101* are regulated primarily at the point of protein synthesis and modification such as phosphorylation, and/or heterodimerization with other binding partners, such as stathmin.

Regulation of tsg101 *expression in primary tumors from transgenic models for breast cancer*

The overexpression of oncogenes and growth regulators in mammary epithelial cells can lead to cellular transformation and tumor progression. Transgenic mouse models have been generated to decipher the molecular pathways leading to tumorigenesis. We investigated the transcriptional activation of the *tsg101* gene by Northern blot analysis in three well-established models that develop breast cancer through distinct signaling cascades. First, pathways effecting involution, programmed cell death and remodeling are impaired in transgenic mice expressing TGFα (Sandgren *et al*., 1995). Secondly, we investigated *tsg101* expression in WAP-SV40 large T antigen (SV40-Tag) transgenic mice (Tzeng *et al*., 1993). SV40-Tag is known to bind and inactivate molecules that are linked to the cell cycle such as Rb and p53, but p53-independent pathways of programmed cell death are not affected (Li *et al*., 1996). Thirdly, we studied a mouse model expressing a tumor specific truncated form of the murine *int3* gene, a Drosophila Notch4-related cell fate protein. Int3 inhibits the growth and differentiation of the mammary lobulo-alveolar compartment and leads to mammary dysplasia, tumorigenesis, and lung metastases.

Transgenic mice expressing TGFα and SV40-Tag did not show any reduction in *tsg101* expression. These data suggest that *tsg101* expression is not altered when apoptosis is inhibited (WAP-TGFα) or induced (WAP-SV40-Tag). Moreover, these data support the hypothesis that *tsg101* trancription is not linked to the cell cycle since the functional inactivation of Rb and p53 has no influence on the mRNA level of *tsg101*. Therefore, WAP-TGFa and WAP-SV40-Tag transgenic mice do not serve as an appropriate model to study the transcriptional inactivation of *tsg101* during the progression of tumor formation. A reduction of *tsg101* mRNA was, however, observed in mice expressing *int3* under the control of WAP regulatory elements. Total RNA from tumors was compared to adjacent mammary tissues of the same transgenic animals to determine whether the downregulation of *tsg101* is a direct effect of the transgene expression or a secondary event in the progression towards cancer. Full length *tsg101* transcripts were present in both fractions, however, the tumors contained less *tsg101* mRNA. The reduction was not the result of aberrant or alternative splice forms. The mechanisms of Notch4/Int3 signaling in the mammary gland is still unresolved. The only *int3* mutation that leads to the formation of tumors is caused by MMTV proviral integration and the subsequent expression of a truncated protein, which consists of the intracellular domain of *int3* (Gallahan and Callahan, 1997). From studies in *Drosophila* and *C. elegans* it is suggested that the intracellular domain of Notch4 localizes to the nucleus and that it interacts with proteins that regulate gene transcription (for references see Greenwald, 1998). It needs further investigation if *tsg101* expression is directly regulated by mechanisms downstream of the *Notch4*/*int3* cascade, or if the downregulation of *tsg101* is a secondary event during tumorprogression.

The functional role of TSG101 during cell proliferation, differentiation and tumor formation remains to be determined. It has been suggested that TSG101 acts primarily as a 'care taker' rather than a 'gate keeper' (Xie *et al*., 1998). This includes the regulation of ubiquitin-mediated protein degradation of other important tumor suppressors, or TSG101 could act as a 'care taker' for chromosomal stability during cell division. In future studies it will be important to investigate the loss of TSG101 function *in vivo* in conjunction with tumor progression. Since a functional knockout using an antisense strategy could lead simultaneously to an inactivation of other genes with partial sequence similarities to *tsg101*, it remains to be determined if a *tsg101* gene deletion approach could also result in a neoplastic transformation and tumor progression. However, a conventional gene targeting and knockout approach could lead to an early embryonic lethality since *tsg101* is ubiquitously expressed at all stages of development. A more defined tissue-specific deletion of the *tsg101* gene in the mammary gland using the Cre-lox recombination system (Wagner *et al*., 1997) could be a practicable solution to study the loss-of-function during proliferation and differentiation *in vivo*.

Materials and methods

Cloning of a mouse tsg101 *cDNA probe*

A cDNA corresponding to the mouse *tsg101* mRNA was cloned by RT–PCR. Total RNA was isolated from mouse mammary tissue 24 h post-partum according to Chomczynski and Sacchi (1987). One microgram of RNA was reverse transcribed at 37°C for 1 h in a total volume of 20 μl using MLV reverse transcriptase (Gibco–BRL) and a gene-specific primer (5′-TTC GTT TCA AGG CAT TAA GCT C-3′). Subsequently, the RT reaction mix was heated at 95°C for 10 min and an aliquot of 10 μl was used for amplification in a 100 μl PCR reaction using *tsg101*-specific forward and reverse primers (5′-CAT GGC TGT CCG AGA GTC AGC-3′ and 5′-CTG TGA GCT TGT TTG GGC AGG G-3′). The PCR conditions were 3 min 94°C, 35 cycles (45 s 94°C, 45 s 65°C, 60 s 72°C), 5 min 72°C, 4°C. The amplification products were separated on a 2% agarose gel and the expected band of about 600 bp was purified using the QIAquick gel extraction kit (Qiagen). The *tsg101* cDNA probe, after sequencing analysis, was subcloned into the pZERO-1 vector (Invitrogen) and used for hybridization and cloning of *tsg101* genomic DNA and Northern blot analysis.

Cloning of the mouse tsg101 *gene and a processed pseudogene*

Initially, we screened a lambda phage mouse genomic library (129/SVJ Lambda FIX II, Stratagene) following the manufacture's protocol with ^{32}P-labeled *tsg101* cDNA (Random Priming Kit, Stratagene). The plaques were further screened with PCR using mouse-specific primers for *tsg101* (Li *et al*., 1997). After sequencing of subcloned fragments we confirmed that the PCR positive plaques contained a processed pseudogene but not the *tsg101* gene. Since pseudogenes show a much stronger signal of several magnitudes during the plaque-screening assay that makes it more difficult to find the actual gene, we switched from a lambda phage to a BAC library (FBAC-4431, Genome Systems, Inc). Filters containing arrayed BAC clones were hybridized again with *tsg101* cDNA according to the manufacture's protocol. Six positive BAC clones were evaluated further by PCR to confirm the presence of the *tsg101* gene and to eliminate clones that contained pseudogenes. Therefore, we had to set up a PCR assay to detect specifically the actual *tsg101* gene. We assumed that mouse *tsg101* gene would have the same overall structure as the human gene that was published previously (Li *et al*., 1997). We designed various primers within exons to amplify intron sequences. Initially, we tested those primers with genomic DNA derived from murine embryonic stem cells. Only one set of primers (5′-CCT TGG AGA AGC TTT GCG GCG-3′ and 5′-TAG CCC AGT CAG TCC CAG CAC AGC ACA G-3′) that amplified the last intron showed expected PCR fragments of about 1.6 kb for the *TSG101* gene and 350 bp for the *tsg101* pseudogene. The 1.6 kb PCR fragment was subcloned and the presence of both intron-exon junctions were confirmed by sequencing. We used the same primer set to screen individual BAC clones that hybridized the *tsg101* cDNA probe. Five out of six BAC clones contained the pseudogene and only one contained the actual gene for *tsg101*. None of the clones had both sequences. The BAC clone containing the *tsg101* gene was mapped using Southern hybridization. Multiple fragments were subcloned into pZERO-1 vector (Invitrogen) and sequenced using a standard protocol for cycle sequencing (Perking Elmer). The Sequencher software (Gene Codes, Corp.) was used to align and assemble individual sequences. Since the BAC clone did not contain the first exon and the 5′ flanking region, we isolated a 500 bp fragment of the most five prime part of the *tsg101* clone and used this fragment as a probe to rehybridize the BAC library. We isolated a second BAC clone containing the missing first exon and more than 100 kb of 5′ flanking region of *tsg101*. This clone had only 5 kb overlapping sequence in common with the first clone.

Isolation of total RNA and Northern blot hybridization

RNA of normal mammary gland tissues and tumors was isolated following a procedure by Chomczynski and Sacchi (1987). The total RNA (20 μg per lane) was separated in 1.5% agarose gels containing 18% (v/v) formaldehyde, transferred to a GeneScreen Plus membrane (Dupont) and hybridized with *tsg101* cDNA as described above.

Transgenic mice

Transgenic mice expressing *int3* (Gallahan *et al*., 1996), *SV40-TAg* (Tzeng *et al*., 1993), and *TGF*α (Sandgren *et al*., 1995) cDNAs under regulatory elements of the WAP gene have been characterized previously. All mouse models developed mammary tumors after several pregnancies. The tumor samples were taken *post mortem*. RNA was isolated and Northern blot hybridization was performed as described above. Radioactive signals were measured using a phosphorimager (Fuji) and the quantity of *tsg101* transcripts was corrected by the amount of total RNA present in each lane (18S rRNA).

PCR amplification of intron sequences within the human TSG101 *gene*

Human DNA was purchased (Promega) and the presence of additional introns within the human *TSG101* gene was tested by PCR using exon-specific primers that encompass individual introns. The primers are 5′-GGT GTC GGA GAG CCA GCT CAAG-3′ and 5′-CAG TTT CAC GTA CAG TTA GGT CTC TG-3′ for intron 1, 5′-GAG ACC TAA CTG TAC GTG AAA CTG-3′ and 5′-TTA GTT CCC TGG AAC TGC CAT CG-3′ for intron 2, 5′-TAA CGA TGG CAG TTC CAG GGA AC-3′ and 5′-GTA TGT GTC CAG TAG CCA TAG GC-3′ for intron 3, 5′-AAT ATG CCT ATG GCT ACT GGA CAC-3′ and 5′-CCA CAA TCA TGA CCT GAA TAA GCC-3′ for intron 4, 5′-GGC TTA TTC AGG TCA TGA TTG TGG-3′ and 5′-ACC TGG CAT GCC TGG CAT GTA G-3′ for intron 5, 5′-TAC ATG CCA GGC ATG CCA GGT G-3′ and 5′-GTA CTG AGA ACT TGT TGT GGC AGG-3′ for intron 6, 5′-GTC CTT ACC CAC CTG GTG GTC C-3′ and 5′-GCC ATC TCA GTT TGT CAC TGA CC-3′ for intron 7, 5′-GGT CAG TGA CAA ACT GAG ATG GC-3′ and 5′-CCC TTC TCA AGG CTT CTC CCA AG-3′ for intron 8, 5′-ATG AAG AAC TCA GTT CTG CTC TGG-3′ and 5′-GAA GTC AGT AGA GGT CAC TGA GAC-3′ for intron 9. PCR was performed using a proof-reading polymerase (ELONGase, Gibco – BRL) specifically designed to amplify long DNA templates. The PCR conditions were 3 min 94°C, 30 cycles (45 s 94°C, 45 s 65°C, 9 min 68°C), 10 min 68°C, 4°C. The amplification products were separated on a 1% agarose gel. The DNA fragments representing individual introns were gel purified and the presence of intron-exon junctions was confirmed by sequencing.

Acknowledgements

This work was supported in part by a German Research Society (Wa 1119/1-1) fellowship and through funding by NIH grant 369VFDK013712. The authors would like to thank Dr Stanley Cohen (Stanford University School of Medicine) for providing us with the full-length mouse and human *TSG101* cDNAs and for helpful suggestions while preparing the manuscript.

References

Aki T, Yanagisawa S and Akanuma H. (1997). *J. Biochem.,* **122,** 271 – 278.

Ali IU, Lidereau R, Theillet C and Callahan R. (1987). *Science,* **238,** 185 – 188.

Chomczynski P and Sacchi N. (1987). *Anal. Biochem.,* **162,** 156 – 159.

Gallahan D and Callahan R. (1997). *Oncogene,* **14,** 1883 – 1890.

Gallahan D, Jhappan C, Robinson G, Hennighausen L, Sharp R, Kordon E, Callahan R, Merlino G and Smith GH. (1996). *Cancer Res.,* **56,** 1775 – 1785.

Gayther SA, Barski P, Batley SJ, Li L, de Foy, KA, Cohen SN, Ponder BA and Caldas C. (1997). *Oncogene,* **15,** 2119 – 2126.

Greenwald I. (1998). *Genes Dev.,* **12,** 1751 – 1762.

Hampl M, Hampl J, Plaschke J, Fitze G, Schackert G, Saeger HD and Schackert HK (1998). *Biochem. Biophys. Res. Commun.,* **248,** 753–760.

Koonin EV and Abagyan RA. (1997). *Nat. Genet.,* **16,** 330–331.

Kozak M. (1997). *EMBO,* **16,** 2482–2492.

Lee MP and Feinberg AP. (1997). *Cancer Res.,* **57,** 3131–3134.

Li L and Cohen SN. (1996). *Cell,* **85,** 319–329.

Li L, Li X, Francke U and Cohen SN. (1997). *Cell,* **88,** 143–154.

Li M, Hu J, Heermeier K, Hennighausen L and Furth PA. (1996). *Cell Growth Differ.,* **7,** 3–11.

Maucuer A, Camonis JH and Sobel A. (1995). *Proc. Natl. Acad. Sci. USA,* **92,** 3100–3104.

Ozon S, Maucuer A and Sobel A. (1997). *Eur. J. Biochem.,* **248,** 794–806.

Ponting CP, Cai YD and Bork P. (1997). *J. Mol. Med.,* **75,** 467–469.

Reeve AE, Sih SA, Raizis AM and Feinberg AP. (1989). *Mol. Cell. Biol.,* **9,** 1799–1803.

Sandgren EP, Schroeder JA, Qui TH, Palmiter RD, Brinster RL and Lee DC. (1995). *Cancer Res.,* **55,** 3915–3927.

Steiner P, Barnes DM, Harris WH and Weinberg RA. (1997). *Nat. Genet.,* **16,** 332–333.

Sun Z, Pan J, Bubley G and Balk SP. (1997). *Oncogene,* **15,** 3121–3125.

Tzeng YJ, Guhl E, Graessmann M and Graessmann A. (1993). *Oncogene,* **8,** 1965–1971.

Wagner KU, Wall RJ, St-Onge L, Gruss P, Wynshaw-Boris A, Garrett L, Li M, Furth PA and Hennighausen L. (1997). *Nucleic. Acids. Res.,* **25,** 4323–4330.

Wang Q, Driouch K, Courtois S, Champeme MH, Bieche I, Treilleux I, Briffod M, Rimokh R, Magaud JP, Curmi P, Lidereau R and Puisieux A. (1998). *Oncogene,* **16,** 677–679.

Watanabe M, Yanagi Y, Masuhiro Y, Yano T, Yoshikawa H, Yanagisawa J and Kato S. (1998). *Biochem. Biophys. Res. Commun.,* **245,** 900–905.

Xie W, Li L and Cohen SN. (1998). *Proc. Natl. Acad. Sci. USA,* **95,** 1595–1600.

Zhong Q, Chen CF, Chen Y, Chen PL and Lee WH. (1997). *Cancer Res.,* **57,** 4225–4228.