

Identification and Characterization of a Novel Human Testis-Specific Kinase Substrate Gene Which Is Downregulated in Testicular Tumors

Andreas Scorilas,^{*,†} George M. Yousef,^{*,†} Klaus Jung,[‡] Ewa Rajpert-De Meyts,[§] Stephan Carsten,[‡] and Eleftherios P. Diamandis^{*,†,1}

^{*}Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada;

[†]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario M5G 1L5, Canada;

[‡]Department of Urology, University Hospital Charite, Humboldt University, Berlin, Germany; and [§]Department of Growth and Reproduction, Juliane Marie Centre, National University Hospital, Copenhagen, Denmark

Received June 8, 2001

By using the positional candidate gene approach, we identified a novel putative serine/threonine kinase substrate gene that maps to chromosome 19q13.3. Screening of expressed sequence tags and reverse transcription-polymerase chain reaction of total RNA from human tissues allowed us to establish the expression of the gene and delineate its genomic organization (GenBank Accession No. AF200923). This gene (TSKS, for testis-specific kinase substrate) is composed of 11 exons and 10 intervening introns and is likely the human homolog of the mouse testis-specific serine kinase substrate gene. The predicted protein-coding region of the gene is 1779 bp, encoding for a 592-amino-acid polypeptide with a calculated molecular mass of 65.1 kDa. Genomic analysis of the region 19q13.3 placed the TSKS gene close to the known genes IRF3, RRAS, and α -Adaptin A. TSKS exhibits high levels of expression exclusively in human testicular tissue. The expression of TSKS is downregulated in cancerous testicular tissue, in comparison to adjacent normal tissue. TSKS expression was very low or undetectable in seminoma, teratocarcinoma, embryonal, and Leydig cell tumors, while high expression was observed in testicular tissue adjacent to tumors which contain premalignant carcinoma *in situ*. The expression of the TSKS gene was very low in two human embryonal carcinoma cell lines, 2102Ep and NTERA-2. These observations suggest a role of TSKS in testicular

physiology, most probably in the process of spermatogenesis or spermiogenesis. © 2001 Academic Press

Key Words: kinase substrates; testis-specific genes; gene mapping; gene characterization; TSKS; testis-specific kinase substrate; RRAS; IRF3; testicular cancer.

Abbreviations used: TSKS, testis-specific kinase substrate, IRF3, interferon regulatory factor 3; PSA, prostate-specific antigen; tssk, testis-specific serine kinase; tssks, testis-specific serine kinase substrate; LLNL, Lawrence Livermore National Laboratory; EST, expressed sequence tag; RT-PCR, reverse transcription-polymerase chain reaction.

¹ To whom correspondence and reprint requests should be addressed at Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada. Fax: (416) 586-8628. E-mail: ediamandis@mtsina.on.ca.

Protein phosphorylation is the most common post-translational protein modification in eukaryotes and a fundamental mechanism for the direct or indirect control of all cellular processes. Considering the integral role that protein kinases play in the control of cellular mechanisms and signal transduction, it is not surprising that several protein kinases have been shown to be involved in spermatogenesis (1, 2). Nevertheless, only a few kinases have been characterized, whose expression is restricted to either germ cells or to the testis. The phosphoglycerate kinase-2 and the serine/threonine kinase MAK are predominantly expressed in spermatocytes during meiosis (3). The serine/threonine kinase c-MOS and the kinase NEK-1 have been implicated in the control of meiosis in both spermatocytes and oocytes (4, 5). Mouse studies demonstrate that spermiogenesis, comprising cytodifferentiation and detachment, involves a complex pattern of interaction between members of a novel family of serine/threonine kinases and as yet uncharacterized substrates (2). Other reports describe a possible role of serine/threonine kinases and their substrates in the progression of testicular tumors of germ cell origin (6, 7).

Recently, Kueng *et al.* employed the yeast two-hybrid system with testis-specific serine kinase (tssk) 1 and tssk2 cDNA as baits and a prey cDNA library from mouse testis and isolated a novel cDNA which encoded for a 65-kDa protein, interacting specifically with both

tssk1 and tssk2 (2). This protein was phosphorylated by both kinases (see GenBank Accession No. AF 025310). The expression of the interacting clone was also testis-specific and paralleled the developmental expression observed for the kinases themselves (2). In the same study, tssk2 was found to be the orthologue of the human DGS-G gene, whose deletion is suspected to be involved in the pathogenesis of Di George and velocardiofacial syndromes (8).

In this report we describe analysis of an area spanning approximately 100 kb of contiguous DNA sequence on human chromosome 19 (19q13.3) for the purpose of identifying new genes by the method of positional candidate cloning (9–11). Our approach allowed us to clone a gene, tentatively named TSKS (for testis-specific kinase substrate; GenBank Accession No. AF200923). This gene is located closed to RRAS and the human IRF3 gene. On homology analysis, the human TSKS gene was found to be similar to the *Mus musculus* tssk1 and tssk2 substrate (tssks) gene, described by Kueng *et al.* (2). We here describe the identification of the new gene, its mapping and its localization, in relation to other genes clustering in the same region. In addition, we describe its genomic, mRNA and protein structure. We further present extensive tissue expression studies and demonstrate that, in addition to testis, which shows the highest expression, the gene is expressed at low levels in prostate, female breast, placenta, ovary, and thymus. Moreover, we examined TSKS expression in human testicular tumors and teratocarcinoma-derived cell lines and found gene downregulation in comparison to normal testicular tissue.

MATERIALS AND METHODS

Gene mapping. Large DNA sequencing data for chromosome 19 is available at the Lawrence Livermore National Laboratory (LLNL) database. We have screened approximately 100 kb of genomic sequence, encompassing a region on chromosome 19q13.3–13.4, where several cancer-associated genes are localized. These sequences were represented by 10 contigs of variable lengths. We performed a restriction analysis study of the available sequences using the “Web-Cutter 2” computer program and with the aid of the *EcoRI* restriction map of this area (also available from the LLNL), we were able to construct an almost contiguous area of genomic sequences (Fig. 1). By using the published sequences of the RRAS oncogene and IRF3 gene, and the alignment program BLAST 2 (12), we identified the relative positions of these genes on the contiguous map.

Gene prediction analysis. For exon prediction analysis of the genomic area of interest, we used a number of different computer programs as previously described (13, 14).

Expressed sequence tag (EST) searching. Sequence homology searching was performed using the BLASTN algorithm (12) against the human EST database (dbEST). Clones with >95% identity were obtained from the IMAGE consortium (15) through Research Genetics Inc. (Huntsville, AL) and from the Institute for Genomic Research (TIGR). Clones were propagated and purified and then sequenced from both directions with an automated sequencer, using insert-flanking vector primers. All clones tested are shown in Table 1.

Protein homology searching. Putative exons of the newly identified gene were first translated to the corresponding amino acid

sequences. BLAST homology searching for the proteins encoded by the exons were performed using the BLASTP program and the GenBank databases (12).

Structure analysis studies. Multiple alignment was performed using the Clustal X multiple alignment program (16). Phylogenetic studies were performed using the Phylip software package (17). Distance matrix analysis was performed using the “Neighbor-Joining/UPGMA” program and parsimony analysis was done using the “Protpars” program (18). Hydrophobicity study was performed using the Baylor College of Medicine search launcher program. Signal peptide and transmembrane regions were predicted using the SignalP (19) and Tmpred (20) software. Protein structure analysis was performed by SAMP program (21). Sequence analysis tools were utilized to detect the presence of possible sites of posttranslational modification on our putative protein. We used the analysis program PROSITE (22) and NetOGlyc 2.0 (23) to detect N- and O-glycosylation, as well as the presence of kinase phosphorylation motifs.

Tissue expression of TSKS. Total RNA isolated from 28 different human tissues was purchased from Clontech (Palo Alto, CA). We prepared cDNA as described below and used it for PCRs with the primers TSKS-F2 and TSKS-R1, described in Table 2.

Expression of TSKS in testis cancer. Included in this study tissue samples from 15 patients who had undergone radical orchiectomy for testicular, tumors at the Charite University Hospital, Germany and the National University Hospital, Denmark. Patient age ranged from 23 to 60 years with a median of 36. Matched testicular tissue samples were obtained from the cancerous and non-cancerous parts of the same testis, from 10 of the patients. All patients had a histologically-confirmed diagnosis of primary testicular cancer and received no treatment before surgery. Of the 10 matched samples, 5 tumors were seminomas, 3 were categorized as embryonal carcinomas and 2 were teratocarcinomas. Other tissue samples included one Leydig cell tumor and two samples of testicular tissue containing pre-malignant carcinoma *in situ* tubules within morphologically normal testicular parenchyma. In addition, the expression of the TSKS gene was examined in two embryonal carcinoma cell lines derived from a human teratocarcinoma; pluripotent (NTERA-2) and nullipotent (2102Ep) (24). The normal and tumor tissues were immediately frozen in liquid nitrogen after surgical resection and stored at -80°C until extraction. We prepared cDNA as described below and used it for PCRs with the primers TSKS-F2 and TSKS-R1 (Table 2). Tissue cDNAs were amplified at a 1:10 dilution.

Reverse transcription-polymerase chain reaction (RT-PCR). Total RNA was extracted from the tissues or cell lines using Trizol reagent (Gibco BRL, Gaithersburg, MD) following the manufacturer's instructions. RNA concentration was determined spectrophotometrically. Two micrograms of total RNA was reverse transcribed into first strand cDNA using the Superscript preamplification system (Gibco BRL). The final volume was 20 μl . Based on the combined information obtained from the predicted genomic structure of the new gene and the EST sequences, two gene-specific primers were designed (TSKS-F2 and TSKS-R1; Table 2) and PCR was carried out in a reaction mixture containing 1 μl of cDNA, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl_2 , 200 μM dNTPs (deoxynucleoside triphosphates), 150 ng of primers and 2.5 units of AmpliTaq Gold DNA polymerase (Roche Molecular Systems, Branchburg, NJ) on a Perkin-Elmer 9600 thermal cycler. The cycling conditions were 95°C for 15 min to activate the *Taq* Gold DNA polymerase, followed by 42 cycles of 94°C for 30 s, 67°C for 1 min and a final extension at 67°C for 10 min. Equal amounts of PCR products were electrophoresed on 2% agarose gels and visualized by ethidium bromide staining. The primers for RT-PCR spanned at least 2 exons to avoid contamination by genomic DNA.

Cloning and sequencing of the PCR products. To verify the identity of the PCR products, they were cloned into the pCR 2.1-TOPO vector (Invitrogen, Carlsbad, CA) according to the manufacturer's

TABLE 1
EST Clones with >95% Identity to Exons
of the TSKS Gene

GenBank accession no.	Tissue	IMAGE ID	Homologous exons
AL041339	Testis	—	6–11
AI587014	Pancreatic adenocarcinoma	2222191	9–11
H02568	Placenta	150825	10, 11
AW236904	Pooled germ cell tumors	268766	11
AW003686	Prostate	2477783	11

TABLE 2
Primers Used for Reverse Transcription–Polymerase Chain
Reaction (RT-PCR) Analysis of the TSKS Gene

Gene	Primer name	Sequence ^a	Product sizes (base pairs)
TSKS	TSKS-F1	AAGACGATCTGGCAGTCCAA	1748
	TSKS-F2	GAGCTGGAGCGCCAGGCCTT	269
	TSKS-R1	GCTGAGCCCCCTGTTTGG	
PSA	PSAS	TGCGCAAGTTCACCCTCA	754
	PSAAS	CCCTCTCTTACTTCATCC	
Actin	ACTINS	ATCTGGCACCACACCTTCTA	838
	ACTINAS	CGTCATACTCCTGCTTGCTG	

^a All nucleotide sequences are given in the 5' → 3' orientation.

instructions. The inserts were sequenced from both directions using vector-specific primers, by an automated DNA sequencer.

Statistical analysis. Statistical analysis was performed with SAS software (SAS Institute, Cary, NC). The analyses of differences between TSKS expression in noncancerous and cancerous tissues were performed with the nonparametric McNemar test. The binomial distribution was used to compute the significance level.

RESULTS

Identification of the TSKS Gene

Exon prediction analysis of the 100-kb DNA sequence around chromosome 19q13.3 identified a novel gene with a structure reminiscent of a serine/threonine kinase substrate gene. Since our gene was found to have 73% and 81% identity at the protein and cDNA level, respectively, to the mouse *tssks* gene reported by Kueng *et al.* (2), we assume that the newly identified TSKS gene (GenBank Accession No. AF200923) is the human homolog of the mouse *tssks* gene.

Cloning of the TSKS Gene

Sequence homology search of the putative exons identified by the gene prediction programs against the human EST database (dbEST) revealed five EST clones with >95% identity to the putative exons of our gene (Table 1). The IMAGE clones were obtained and the inserts were sequenced from both directions. Alignment was used to compare the EST sequences and the exons predicted by the computer programs, and final selection of the exon–intron splice sites was done according to the EST sequences. Furthermore, many of the ESTs were overlapping, further ensuring the accu-

racy of the data. Moreover, two ESTs were found to have a poly(A) tail which was not found in the genomic sequence, verifying the 3' end of the gene.

To verify the accuracy of the cDNA sequence, PCRs were performed using gene-specific primers for the first and last exons of the predicted structure of the gene (TSKS-F1 and TSKS-R1; Table 2) with cDNA isolated from different human tissues as putative templates. The PCRs were performed under different optimization conditions using the EST clones as positive controls. A positive band of the expected size was isolated from testis cDNA and fully sequenced. Its sequence was aligned by BLAST against the genomic sequence to further define and confirm the exon/intron boundaries. The genomic and cDNA sequence of the gene is now deposited in GenBank (Accession No. AF200923).

Mapping and Chromosomal Localization of the TSKS Gene

Alignment of the sequences of TSKS and other known genes that we found to be located within the 100-kb area of interest, on chromosome 19q13.3, enabled us to precisely localize these genes and determine the distances between them, as shown in Fig. 1. The TSKS gene maps between two known genes, namely RRAS and the human α -Adaptin A gene, and transcribes in the opposite direction. Human α -Adaptin A is 3.7 kb more telomeric to TSKS while RRAS is 5.5 kb more centromeric to it. The IRF3 gene maps in the same region, and transcribes in the same

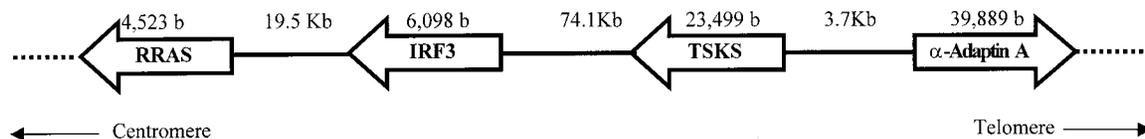


FIG. 1. Localization of the TSKS gene within an approximately 100-kb region of contiguous genomic sequence around chromosome 19q13.3. Gene lengths are presented above each arrow, and distances between genes are also shown. Arrows denote the directions of transcription. Figure is not drawn to scale. kb, kilobase; b, base. For full gene names, see abbreviations used footnote.

cctgtgcccccactcactggggccagtggttggtcctttgtgagcagcagctgaggggacagtgatcctgagggtctgtgtcacag
taggttaggaagtctcagaggaaggagccccctctggagccagACCCCCACACC(ATG)GCG AGC GTG GTG GTG AAG ACG
M A S V V V K T
ATC TGG CAG TCC AAA GAG ATC CAT GAG GCC GGG GAC ACC CCC ACG GGG GTG GAG AGC TGC TCC CAG
I W Q S K E I H E A G D T P T G V E S C S Q
CTA GTC CCA GAG GCT CCC CGG AGG GTG ACC AGC CGG GCC AAG GGG ATC CCG AAG AAA AAG AAG GC
L V P E A P R R V T S R A K G I P K K K K A
GTG TCG TTC CAC GG[gtgggtt]gt.....intron-1.....gttcccat[ag]G GTG GAG CCC CAG ATG TCC CAT CAG CC
V S F H G V E P Q M S H Q P
ATG CAC TGG TGC CTG AAC CTC AAA CGG TCC TCG GCC TGC ACC AAC GTG TCA CTG CTC AAC CTG GCC
M H W C L N L K R S S A C T N V S L L N L A
GCC ATG GAG CCC ACT GAC TCC ACG GGG ACA GAC TCC ACA GTG GAA GAC CTC AGC GGC CAA CTC ACA
A M E P T D S T G T D S T V E D L S G Q L T
CTG GCT GGG CCC CCT GCC CCT ACC CTA CCC TGG GAT CCG GAT GAC GCA GAC ATC ACG GAA ATC
L A G P P A S P T L P W D P D D A D I T E I
CTG[gtgagg]gccc.....intron-2.....tcttcccc[ag]AGT GGG GTC AAC AGT GGA TTG GTC CGC GCC AAA GAC TCC
L S G V N S G L V R A K D S
ATC ACC AGC TTG AAG GAA AAG ACC AAC CGG GTT AAC CAG CAC GTG CAG TCT CTG CAG[gtgtgagg]ac.....
I T S L K E K T N R V N Q H V Q S L Q
intron-3.....catgctcc[ag]AGC GAG TGT TCT GTG CTG AGC GAG AAT CTG GAG AGA AGG CGG CAA GAG GCA
S E C S V L S E N L E R R R Q E A
GAA GAG TTG GAG GGG TAC TGC ATT CAA CTC AAG[gtgagccag]tg.....intron-4.....tggatggga[ag]GAG AAC
E E L E G Y C I Q L K E N
TGC TGG AAG GTG ACC CGG TCT GTG GAA GAT GCT GAA ATC AAA ACC AAC GTC TTG AAG CAG AAT
C W K V T R S V E D A E I K T N V L K Q N
TCT GCC CTG CTG GAG[gtlaagg]gcta.....intron-5.....taccacc[ag]GAG AAG CTG CGC TAC CTC CAG CAG
S A L L E E K L R Y L Q Q
CAG CTG CAG GAT GAG ACG CCG CGA CGG CAG GAG GCC GAG CTG CAG GAG CCG GAG GAG AAG CAG GAG
Q L Q D E T P R R Q E A E L Q E P E E K Q E
CCG GAG GAG AAG CAG GAG CCG GAG GAG AAG CAG AAG CCG GAG GCT GGC CTC TCC TGG AAC AGC
P E E K Q E P E E K Q K P E A G L S W N S
CTG GGC CCC GCC ACC CCG TCC CAG GGC TGC CCC GGC CCG CCA GGG AGT CCC GAC AAA CCC TCG CGG
L G P A A T S Q G C P P G P P G S P D K P S R
CCA CAC GGC CTG GTC CCC GCA GGC TGG GGA ATG GGG CCT CGG GCT GGC GAG GGC CCC TAC GTG AGC
P H G L V P A G W G M G P R A G E G P Y V S
GAG CAG GAA TTG CAG AAG CTG TTC ACC GGC ATC GAA GAG CTG AG[gtgggg]cgga.....intron-6..... gccc
E Q E L Q K L F T G I E E L R
ctcc[ag]G AGA GAG GTG TCC TCA CTG ACC GCC CGG TGG CAT CAG GAG GAG GGG GCG GTG CAG GAA GCC
R E V S S L T A R W H Q E E G A V Q E A
CTG CGG CTG CTC GGG GGC CTG GGC GGC AGG GTC GAC GGC TTC CTA GGC CAG TGG GAG CCG GCA CAG
L R L L G G L G G R V D G F L G Q W E R A Q
CGC GAA CAG GCA CAG ACG GCG CGG GAC TTG CAG GAG CTG CGA GGT CCG GCG GAT GAG CTG TGC ACC
R E Q A Q T A R D L Q E L R G R A D E L C T
AT[gtgagg]attg.....intron-7.....catgcctc[ag]G GTG GAG CCG TCA GCA GTG TCT DTG GCT TCA CTG AGG
M V E R S A V S V A S L R
AGC GAA CTG GAG GGG CTG GGC CCA CTG AAA CCC ATT CTG GAG GAG TTC GGG CCG CAA TTT CAG AAC TCT
S E L E G L G P L K P I L E E F G R Q F Q N S
CGA AGA GGG CCT GAC CTC TCC ATT AAC CTG GAT CCG TCC CAC CAA GGC AAC TGT GCC CGC TGT GCC AG
R R G P D L S M N L D R S H Q G N C A R C A S
[gaacgg]gggc.....intron-8.....ctctcccc[ag]C CAG GGG TCG CAG TTG TCT ACG GAG TCC CTG CAG CAG CTG CTG
Q G S Q L S T E S L Q Q L L
GAC CGA GCA CTG ACC TCA CTA GTG GAC GAG GTG AAG CAG AGG GGC CTG ACT CCT GCC TGT CCC AG
D R A L T S L V D E V K Q R G L T P A C P S
TGT CAG AGG CTA CAC AAG AAG ATT CTG[glaccag]gga.....intron-9.....ctctcccc[ag]GAG CTG GAG CGC CAG
C Q R L H K K I L E L E R Q
GCC TTA GCC AAA CAC GTC AGG GCA GAG GCC CTG AGC TCC ACC CTT CGG CTG GCC CAA GAC GAG GCC CTG
A L A K H V R A E A L S S T L R L A Q D E A L
CGG GCC AAG AAC CTA CTG CTG ACA GAC AAG ATG AAG CCA GA[gtgagga]agg.....intron-10.....ctcatgac[ag]G
R A K N L L L T D K M K P E
GAG AAG ATG GCC ACT CTG GAC CAT CTA CAC TTG AAG ATG TGC TCC CTC CAC GAT CAT CTC AGC AAC
E K M A T L D H L H L K M C S L H D H L S N
CTG CCA TTT GAG GGG TCC ACG GGA ACA ATG GGG GGA GGC AGC AGT GCA GGA ACC CCC CCA AAA CAG
L P L E G S T G T M G G G S S A G T P P K Q
GGG GGC TCA GCC CCT GAA CAA(TAA)ATGGCCTCTCATGCTA[gcata]gaattctgtctgcttttgggaccagagtcctagtgaattg
G G S A P E Q

FIG. 2. Genomic structure of the TSKS gene. Shown are the exon/intron boundaries, as well as the predicted protein sequence. The underlined region denotes 5' genomic sequences and the double underlined region 3' genomic sequences. The possibility for 5'-untranslated exons could not be excluded. Introns are shown with lower case letters and exons with capital letters. For full sequence, see GenBank Accession No. AF200923. The start and stop codons are encircled and the exon-intron junctions are boxed. The translated amino acids of the coding region are shown underneath by a single-letter abbreviation. A putative polyadenylation signal (AATAAA) is present within the stop codon.

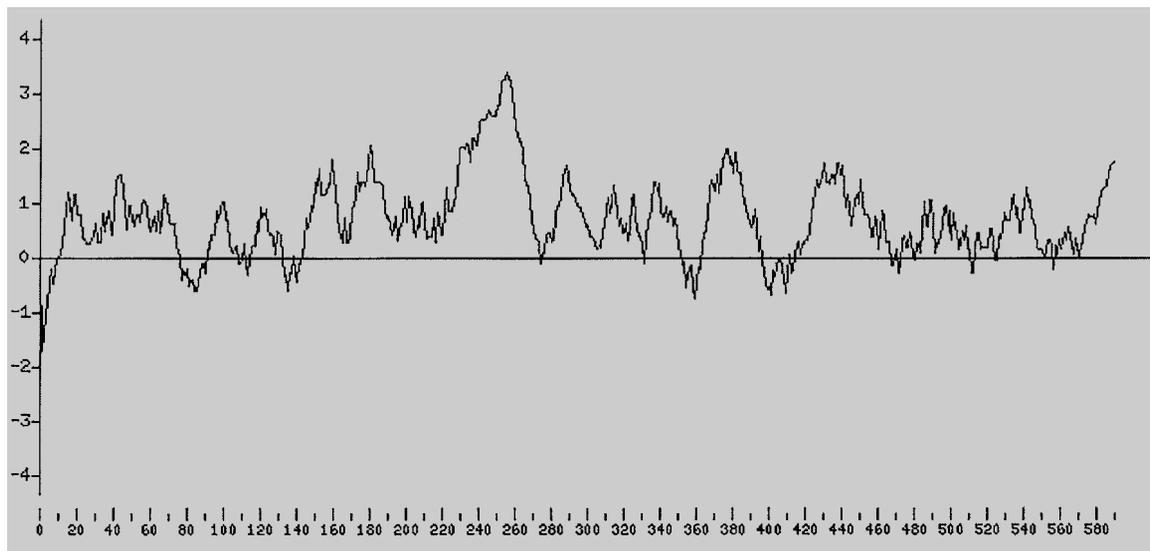


FIG. 3. Hydrophobicity plot of the TSKS protein. There is one hydrophobic region around residues 1–19, corresponding to the putative signal peptide.

direction with RRAS. The distance between IRF3 and RRAS was calculated to be 19.5 kb (Fig. 1).

Structural Characterization of the TSKS Gene and Its Protein Product

As shown in Fig. 2, the TSKS gene is formed of eleven exons and ten intervening introns, spanning an area of 23.6 kbp of genomic sequence on chromosome 19q13.3. The lengths of the coding exons are 181, 229, 96, 84, 84, 329, 195, 174, 136, 125, and 173 bp, respectively of which there are 11 bp of 5' and 16 bp of 3' untranslated regions. All of the exon/intron splice sites conform to the consensus sequence (-mGT AGm-) for eukaryotic splice sites (25). A potential translation initiation codon (ATG) is present at nucleotides 12–14 of the predicted first exon (numbers refer to our GenBank Submission No. AF200923). The flanking sequence of that codon (CACCATGG) matches closely with the Kozak consensus sequence (GCC A/G CCATGG) for initiation of translation (26). The stop codon (TAA) is located 16 bp upstream of the poly(A) tail.

Protein Analysis

The predicted protein-coding region of the gene is formed of 1779 bp (including the stop codon), encoding a deduced amino acid polypeptide of 592 residues, with a predicted molecular mass of 65.1 kDa and a theoretical isoelectric point of 5.7.

A hydrophobicity study of the TSKS protein shows a hydrophobic area in the N-terminal region of the protein (Fig. 3); thus, a presumed signal peptide is present. By software analysis, a possible 19-amino-acid

signal peptide is predicted with a cleavage site at the carboxyl end of Gly¹⁹. TSKS protein is predicted to have no transmembrane region, in agreement with the mouse homolog protein. The mouse tssks protein was found to be phosphorylated on serine residues by the tssk1 and tssk2 kinase enzymes (2). We predicted 22 serine and 9 threonine potential phosphorylation sites, as well as 3 serine and 2 threonine glycosylation sites (Fig. 4).

Expression of the TSKS Gene in Human Tissues

We have assessed by RT-PCR the tissues that express the TSKS gene. The experiments were performed at various dilutions of the cDNA to obtain preliminary information about the relative levels of expression. RT-PCR for actin was used as a positive control and RT-PCR for PSA cDNA was used as another positive control with tissue restricted specificity. Positive ESTs for TSKS were also used as controls. The PSA gene was found to be highly expressed in the prostate, as expected, and to a much lower extent in mammary, salivary glands, thyroid gland, and trachea, as also reported recently (27, 28).

The tissue expression of TSKS is summarized in Fig. 5. This kinase substrate is highly expressed in the testis. Much lower levels of expression are seen in prostate, placenta, fetal liver, thymus, and mammary gland. No expression was seen for the other tested tissues. To verify the RT-PCR specificity, representative PCR products were cloned and sequenced.

TSKS Expression in Testicular Cancer

Expression of the TSKS gene in testicular cancer, at the mRNA level, was examined by RT-PCR, with the

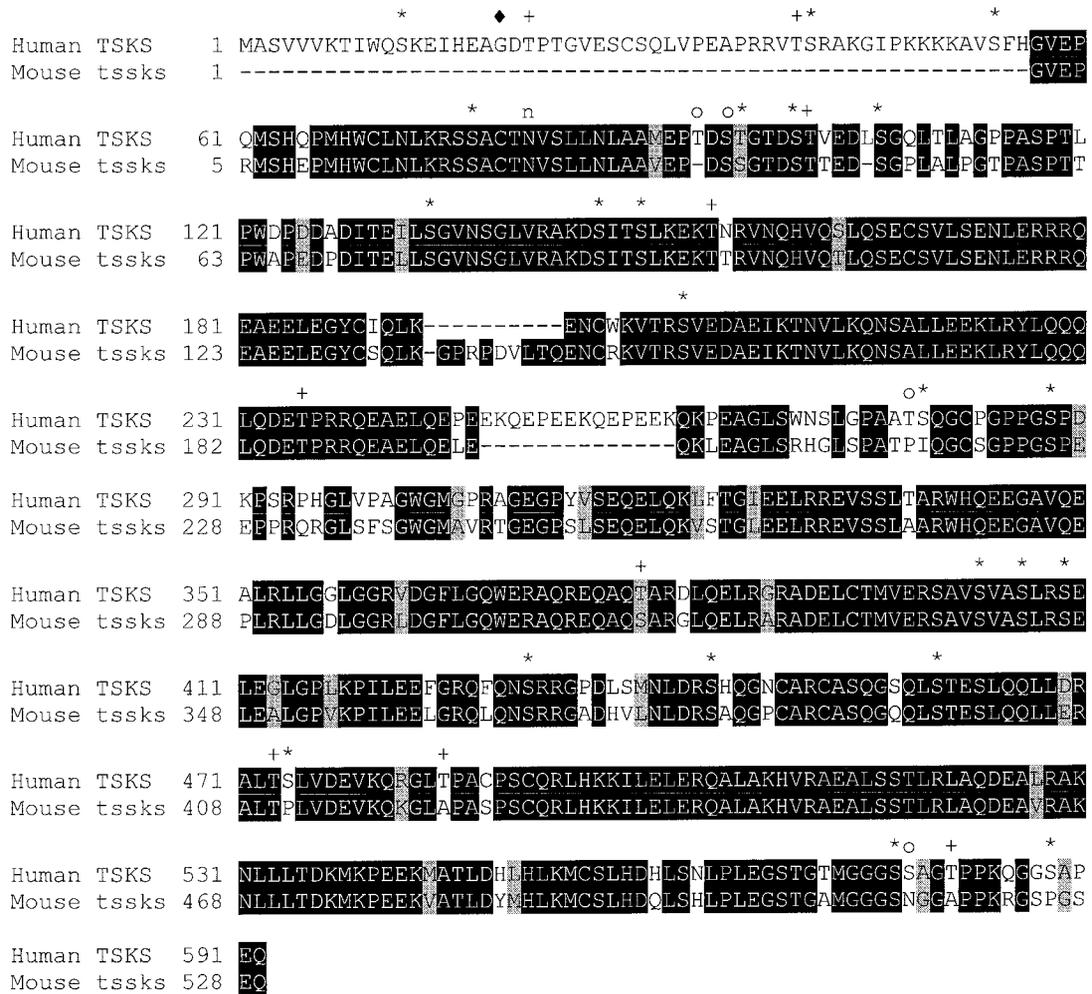


FIG. 4. Alignment of the deduced amino acid sequence of TSKS with the mouse testis specific serine kinase substrate (tssks) protein. Serine and threonine kinase potential phosphorylation motifs are represented by (*) and (+) respectively. Putative N- and O-glycosylation sites are indicated by (n) and (o), respectively. The predicted cleavage site of the signal peptide is indicated by (◆). Conservative amino acid substitutions are highlighted in gray.

use of actin, as a positive control. We first examined the expression of this gene in three samples of morphologically normal testicular parenchyma adjacent to overt tumors; two with carcinoma one *in situ* tubules, one without carcinoma *in situ*, one Leydig-cell tumor as

well as in pluripotent and nullipotent human embryonal carcinoma cell lines. High levels of expression of TSKS in both samples of tissue with carcinoma *in situ* were observed. TSKS was not expressed at all in the Leydig-cell tumor, while the nullipotent and pluripo-

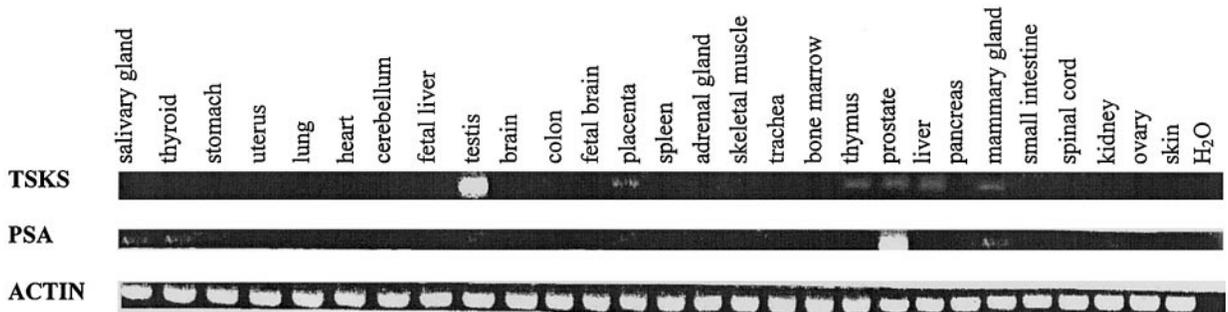


FIG. 5. Tissue expression profile of the TSKS gene as assessed by RT-PCR. Actin and PSA are control genes. For discussion, see text.

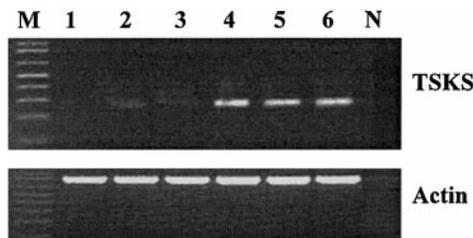


FIG. 6. Expression of the TSKS gene in testicular tumors and embryonal carcinoma (EC-like) cell lines. 1, Leydig-cell tumor; 2, pluripotent EC cell line NT-2; 3, nullipotent EC cell line 2102Ep; 4, 5, testicular tissue containing carcinoma *in situ*; 6, normal testicular tissue; M and N, molecular weight markers and negative control, respectively. Actin was used as a control gene. For discussion, see text.

tent embryonal carcinoma-cell lines exhibited very low levels of expression (Fig. 6).

To examine the expression of the TSKS gene in benign and malignant tissues we analyzed 10 pairs of testicular tissues (cancer/normal) and the results are summarized in Fig. 7. Nine of 10 patients had significantly higher TSKS expression in the non-cancerous tissue and only one had about the same levels of expression in both tissues. Expression of the TSKS gene was not detectable in tumor from 7 of 10 patients. Analysis by the McNemar test indicated that the changes between non-cancer vs cancer tissues are statistically significant ($P = 0.004$). Because of the small number of cases, binomial distribution was used to compute the significance level.

DISCUSSION

Using linear genomic sequences of considerable length, gene prediction programs and the available EST database, we were able to identify a new gene, tentatively named TSKS (for testis specific kinase substrate). This gene is composed of 11 exons and ten introns and is likely to be the human homolog of the mouse testis specific serine kinase substrate gene (*tssks*) (2). This is based on the considerable homologies at the amino acid and nucleotide sequence level, and the tissue-restricted expression to the testis. The predicted protein-coding region of the gene is

1779 bp and encodes for a 592-amino-acid polypeptide with a predicted molecular weight of 65.1 kDa. The 3' end of the gene was verified by the presence of a poly A tail in the sequenced ESTs which was not found in the genomic sequence. The start codon was identified by the presence of a consensus Kozak sequence. The exon-intron splice junctions were identified by comparing the genomic sequence with the EST or cDNA sequence obtained by RT-PCR and were further verified by the full conservation of the consensus splice sequences (25).

The human TSKS gene is located on chromosome 19q13.3, close to *RRAS* and *IRF3*. In this study, we also identified the location of the *IRF3* gene in relation to the *RRAS* gene. The distance between *IRF3* and *RRAS* was calculated to be 19.5 kb and both of them transcribe in the same direction (Fig. 1). The connection of the RAS genes to human malignancy is well-known (29). *IRF3* is a member of a growing family of transcription factors. It is becoming increasingly apparent that *IRF3* plays a pivotal role in the cell's response to viral infection. Phosphorylation and activation of *IRF3* in response to stimuli other than viral infection may result in transcription of genes, in addition to the early inflammatory genes (30, 31). Activation of *IRF3* by growth factors can lead to a negative feedback control of cell growth (31).

Kueng *et al.* recently found that mouse *tssk2* is the likely ortholog of the human DGS-G gene (2). The DGS-G gene has been characterized as one of eleven putative transcription units encoded in the minimal DiGeorge critical region of 250 kb, located on the proximal arm of human chromosome 22, whose deletion is suspected to be involved in the pathogenesis of the DiGeorge and velocardiofacial syndromes. Both syndromes represent developmental disorders associated with a spectrum of malformations including hypoplasia of the thymus and parathyroid glands, cardiovascular anomalies, and mild craniofacial dysmorphism. In light of the fact that the mouse *tssks* protein, a substrate for mouse *tssk2*, is homologous to human TSKS, we can speculate that the TSKS protein may be a substrate of the DGS-G protein. This proposal merits further investigation.

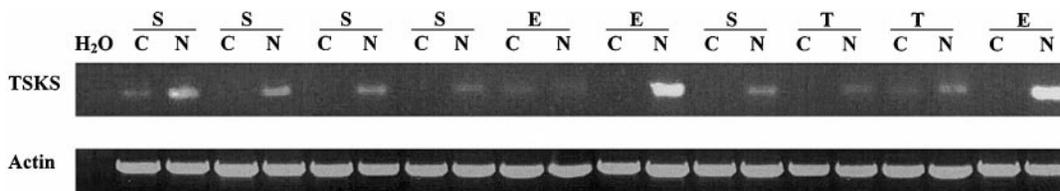


FIG. 7. Expression of the TSKS gene in paired (from the same patient) cancerous (C) and noncancerous (N) testicular tissue. S, seminoma; E, embryonal carcinoma; T, teratocarcinoma. Actin was used as a control gene. The TSKS gene is down-regulated in tumor vs adjacent testicular tissue in 9 of the 10 pairs. For more comments, see text.

Our results indicate that the expression of TSKS is down-regulated in testicular cancer. TSKS expression was very low or undetectable in a Ledig cell tumor and in germ-cell tumors, including seminoma, teratocarcinoma, embryonal carcinoma as well as in an embryonal carcinoma cell lines. The high TSKS expression in tissue with carcinoma *in situ* is most probably due to the presence of normal seminiferous tubules with preserved spermatogenesis. Such tubules were seen in other fragments of tissue isolated from the same patients.

The neighboring RRAS and IRF3 genes are also involved in tumorigenesis pathways and progression (29–32). The observation that TSKS is located in the region of chromosome 19 that is often deleted in gliomas, suggests that it should be further investigated as a putative tumor suppressor (30, 31). The presence of many GC clusters in the promoter area (data not shown) of the TSKS suggests that transcription of this gene, similar to that of other tumor suppressor genes, may be inactivated through methylation.

The mouse homolog of TSSK, *tssks*, was found with immunohistochemistry to be expressed in the cytoplasm of spermatids located in the luminal cell layer of the seminiferous tubules (2). Consistent with these findings, we found that TSKS, aside from a putative 19-amino-acid signal peptide, does not have any predicted transmembrane regions. Interestingly, *tssks* was expressed in the cytoplasm of spermatids undergoing nuclear condensation as well as in apparently unnuclear cytoplasmic remnants, which may later give rise to the smaller residual bodies. Mouse *tssks* protein is not involved in the process of chromatin condensation, but rather, participates in the reconstruction of the sperm cytoplasm and reflects the different stages of sperm maturation in the individual tubular section. The human homolog may play similar roles but this proposal needs further investigation when antibodies have become available. Nevertheless, our observation of the high expression of the human TSKS gene in normal tissue and morphologically normal tissue adjacent to overt tumors (with or without carcinoma *in situ*) is consistent with the putative localization of the transcript to the haploid germ cells, most probably spermatids.

In conclusion, we have cloned a new gene, TSKS, which appears to be the human homolog of the mouse *tssks* gene. This gene appears to be a substrate for testis-specific kinases and is specifically expressed in the testis. We speculate that this gene is involved in sperm formation and may play a role in the pathogenesis of DiGeorge and velocardiofacial syndromes. Due to its tissue specificity, and differential expression in testicular cancers, TSKS protein may find clinical applications in testicular pathologies and may be a target for future gene therapy treatments.

REFERENCES

- Hunter, T. (1996) Tyrosine phosphorylation: Past, present and future. *Biochem. Soc. Trans.* **24**, 307–327.
- Kueng, P., Nikolova, Z., Djonov, V., Hemphill, A., Rohrbach, V., Boehlen, D., Zuercher, G., Andres, A. C., and Ziemiecki, A. (1997) A novel family of serine/threonine kinases participating in spermiogenesis. *J. Cell Biol.* **139**, 1851–1859.
- Jinno, A., Tanaka, K., Matsushime, H., Haneji, T., and Shibuya, M. (1993) Testis-specific mak protein kinase is expressed specifically in the meiotic phase in spermatogenesis and is associated with a 210-kilodalton cellular phosphoprotein. *Mol. Cell Biol.* **13**, 4146–4156.
- Letwin, K., Mizzen, L., Motro, B., Ben-David, Y., Bernstein, A., and Pawson, T. (1992) A mammalian dual specificity protein kinase, Nek1, is related to the NIMA cell cycle regulator and highly expressed in meiotic germ cells. *EMBO J.* **11**, 3521–3523.
- Goldman, D. S., Kiessling, A. A., Millette, C. F., and Cooper, G. M. (1987) Expression of *c-mos* RNA in germ cells of male and female mice. *Proc. Natl. Acad. Sci. USA* **84**, 4509–4513.
- Shuin, T., Misaki, H., Kubota, Y., Yao, M., and Hosaka, M. (1994) Differential expression of protooncogenes in human germ cell tumors of the testis. *Cancer* **73**, 1721–1727.
- Avizienyte, E., Roth, S., Loukola, A., Hemminki, A., Lothe, R. A., Stenwig, A. E., Fossa, S. D., Salovaara, R., and Aaltonen, L. A. (1998) Somatic mutations in LKB1 are rare in sporadic colorectal and testicular tumors. *Cancer Res.* **58**, 2087–2090.
- Gong, W., Emanuel, B. S., Collins, J., Kim, D. H., Wang, Z., Chen, F., Zhang, G., Roe, B., and Budarf, M. L. (1996) A transcription map of the DiGeorge and velo-cardio-facial syndrome minimal critical region on 22q11. *Hum. Mol. Genet.* **5**, 789–800.
- Birren, B., Green, E. D., and Klaphoz, S. (1998) *Genome Analysis, a Laboratory Manual*, Cold Spring Harbor Laboratory Press, Plainview, NY.
- Ballabio, A. (1993) The rise and fall of positional cloning? *Nat. Genet.* **7**, 513–520.
- Collins, F. S. (1995) Positional cloning moves from perdictional to traditional. *Nat. Genet.* **9**, 347–350.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1998) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
- Murakami, K., and Takagi, T. (1998) Gene recognition by combination of several gene-finding programs. *Bioinformatics* **14**, 665–675.
- Yousef, G. M., Luo, L-Y., and Diamandis, E. P. (1999) Identification of novel human kallikrein-like genes on chromosome 19q13.3-q13.4. *Anticancer Res.* **19**, 2843–2852.
- Lennon, G., Auffray, C., Polymeropoulos, M., and Soares, M. B. (1996) The IMAGE Consortium: An integrated molecular analysis of genomes and their expression. *Genomics* **33**, 151–152.
- Aiyar, A. (2000) The use of CLUSTAL W and CLUSTAL X for multiple sequence alignment. *Methods Mol. Biol.* **132**, 221–241.
- Retief, J. D. (2000) Phylogenetic analysis using PHYLIP. *Methods Mol. Biol.* **132**, 243–258.
- Takezaki, N. (1998) Tie trees generated by distance methods of phylogenetic reconstruction. *Mol. Biol. Evol.* **15**, 727–737.
- Nielsen, H., Brunak, S., and von Heijne, G. (1999) Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Eng.* **12**, 3–9.
- Hofmann, K., and Stoffel, W. (1993) TMbase—A database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler* **347**, 166.
- Brendel, V., Bucher, P., Nourbakhsh, I., Blaisdell, B. E., and

- Karlin, S. (1992) Methods and algorithms for statistical analysis of protein sequences. *Proc. Natl. Acad. Sci. USA* **89**, 2002–2006.
22. Bairoch, A., Bucher, P., and Hofmann, K. (1997) The PROSITE database, its status in 1997. *Nucleic Acids Res.* **25**, 217–221.
23. Hansen, J. E., Lund, O., Tolstrup, N., Gooley, A. A., Williams, K. L., and Brunak, S. (1998) NetOglyc: Prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj. J.* **15**, 115–130.
24. Andrews, P. W., Bronson, D. L., Benham, F., Strickland, S., and Knowles, B. B. (1982) A comparative study of eight cell lines derived from human testicular teratocarcinoma. *Int. J. Cancer* **29**, 523–531.
25. Iida, Y. (1990) Quantification analysis of 5'-splice signal sequence in mRNA precursors. Mutations in 5'-splice signal sequence of human β -globin gene and β thalassemia. *J. Theor. Biol.* **145**, 523–533.
26. Kozak, M. (1991) An analysis of vertebrate mRNA sequences: Intimations of translational control. *J. Cell Biol.* **115**, 887–892.
27. Diamandis, E. P., Yu, H., and Sutherland, D. J. (1994) Detection of prostate-specific antigen immunoreactivity in breast tumors. *Breast Cancer Res. Treat.* **32**, 301–310.
28. Ishikawa, T., Kashiwagi, H., Iwakami, Y., Hirai, M., Kawamura, T., Aiyoshi, Y., Yashiro, T., Ami, Y., Uchida, K., and Miwa, M. (1998) Expression of alpha-fetoprotein and prostate specific antigen genes in several tissues and detection of mRNAs in normal circulating blood by reverse transcriptase-polymerase chain reaction. *Jpn. J. Oncol.* **28**, 723–728.
29. Webb, C. P., Van Aelst, L., Wigler, M. H., and Woude, G. F. (1998) Signaling pathways in Ras-mediated tumorigenicity and metastasis. *Proc. Natl. Acad. Sci. USA* **95**, 8773–8778.
30. Lowther, W. J., Moore, P. A., Carter, K. C., and Pitha, P. M. (1999) Cloning and functional analysis of the human IRF3 promoter. *DNA Cell Biol.* **18**, 685–692.
31. Au, W. C., Moore, P. A., Lowther, W., Juang, Y. T., and Pitha, P. M. (1995) Identification of a member of the interferon regulatory factor family that binds to the interferon-stimulated response element and activates expression of interferon-induced genes. *Proc. Natl. Acad. Sci. USA* **92**, 11657–11661.
32. Movilla, N., Crespo, P., and Bustelo, X. R. (1999) Signal transduction elements of TC21, an oncogenic member of the R-Ras subfamily of GTP-binding proteins. *Oncogene* **18**, 5860–5869.