ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

GENOMICS

Genomics 82 (2003) 153–161

www.elsevier.com/locate/ygeno

Cloning and characterization of a novel gene, SHPRH, encoding a conserved putative protein with SNF2/helicase and PHD-finger domains from the 6q24 region☆

Raman Sood,[a,*] Izabela Makalowska,[b] Michal Galdzicki,[a] Ping Hu,[a] Erica Eddings,[a] Christiane M. Robbins,[a] Tracy Moses,[a] Jin Namkoong,[c] Suzie Chen,[c] and Jeffrey M. Trent[a,1]

[a] *Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA*
[b] *Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA*
[c] *Susan Lehman Cullman Laboratory for Cancer Research, Department of Chemical Biology, Ernest Mario School of Pharmacy, Rutgers, the State University of New Jersey, Piscataway, NJ 08854, USA*

Received 21 November 2002; accepted 8 April 2003

Abstract

Here we report the identification of a novel transcript containing SNF2, PHD-finger, RING-finger, helicase, and linker histone domains mapping to the q24 band region of human chromosome 6. These domains are characteristic of several DNA repair proteins, transcription factors, and helicases. We have cloned both human and mouse homologs of this novel gene using interexon PCR and RACE technologies. The human cDNA, termed *SHPRH*, is 6018 bp and codes for a putative protein of 1683 amino acids. The mouse cDNA, termed *Shprh*, is 7225 bp and codes for a putative protein of 1616 amino acids. The deduced amino acid sequences of the two proteins share 86% identity. Both genes are expressed ubiquitously, with a transcript size of ~7.5 kb. Mapping of this gene to 6q24, a region reported to contain a tumor suppressor locus, prompted us to evaluate SHPRH by mutation analysis in tumor cell lines. We have identified one truncating and three missense mutations, thus suggesting *SHPRH* as a possible candidate for the tumor suppressor gene.

Keywords: SHPRH; 6q24; Tumor suppressor; SNF2 domain; PHD-finger domain; RING-finger domain; Linker histone domain; DNA/RNA helicases

Several studies have shown that the long arm of human chromosome 6 harbors a tumor suppressor gene involved in melanoma [1], ovarian cancer [2], breast cancer [3], pancreatic cancer [4], prostate cancer [5], cervical cancer [6], childhood endodermal sinus tumors [7], and hepatocellular carcinoma [8]. Allelic loss or deletions of the chromosomal band region 6q24–q27 seem to be most prevalent in these cancers [9–13]. Despite the availability of Human Genome Project resources, including ESTs, full-length cDNA clones, and predicted genes, the identity of the tumor suppressor gene(s) from the 6q24–q27 region affecting so many different types of cancer remains elusive. Here we report the cloning and characterization of a novel gene with domains characteristic of DNA repair proteins and transcription factors. This gene, referred to as *SHPRH*, was identified during our efforts to characterize the region encompassing the transgene integration site of a mouse model displaying a hereditary melanoma phenotype [14,15]. Here we present data on cloning and characterization of both the murine and the human homologs of this gene, including the complete coding sequence, alternative splice variants, expression pattern in various tissues, and genomic organization. Analysis of the putative protein for homologies and functional motifs to gain insight into its function shows the

☆ Sequence data from this article have been deposited with the GenBank Data Library under Accession Nos. AY161136, AY162264, AY162265, and AY162266.

* Corresponding author: NHGRI, NIH, Building 50, Room 5343, 9000 Rockville Pike, Bethesda, MD 20892, USA. Fax: +1-301-594-0023.
E-mail address: rsood@nhgri.nih.gov (R. Sood).
[1] Current address: Translational Genomics Research Institute, 400 North Fifth Street, Suite 1600, Phoenix, AZ 85004, USA.

0888-7543/03/$ – see front matter
doi:10.1016/S0888-7543(03)00121-6

presence of five recognizable domains: SNF2, histone linker, PHD-finger, RING-finger, and helicase, hence the name SHPRH. Preliminary mutation analysis and protein homology data suggest that *SHPRH* could be a candidate tumor suppressor gene.

Results

Cloning of murine Shprh

We assembled a mouse BAC contig of ~1 Mb encompassing the transgene integration site in TG3 mice [14,15] and sequenced several clones. Subsequent BLAST analysis of the sequence contigs from one of these BAC clones (13P17) identified several regions of homology with DNA repair proteins, such as Rad8, indicating the presence of a related gene in this BAC. We designed an interexon PCR strategy based on the alignment of two mouse ESTs, AU024614 and AV116169, and Rad8 protein with our BAC sequence contigs. We assembled 6.1 kb of cDNA sequence by PCR between primers from AV116169 and AU024614. Further extension of 1.1 kb was obtained using 5′ rapid amplification of cDNA ends (RACE) on brain and testis cDNA templates. The final 7225-bp nucleotide sequence of full-length *Shprh* revealed an open reading frame of 4848-bp coding for a putative protein of 1616 amino acids. The presence of an in-frame upstream stop codon in the 5′UTR sequence indicates that we have identified the complete coding sequence for this gene (isoform A, GenBank Accession No. AY162264). In addition, we identified two splice variants, isoforms B and C, with alternative amino and carboxyl termini, respectively. Isoform B (GenBank Accession No. AY162265) lacks the first 285 amino acids of isoform A, replacing them with an alternative 19 amino acid sequence. Isoform C (GenBank Accession No. AY162266) has an extra 31 amino acids at the carboxyl terminus due to splicing of a 545-bp intron near the stop codon of isoform A. These extra 31 amino acids are unique to isoform C, showing no homology to the protein database or conserved domain database of GenBank.

BLAST analysis of the three isoforms identified several ESTs and two partial cDNA sequences. These include a 4044-bp sequence annotated as similar to SWI/SNF-related matrix-associated, actin-dependent regulator of chromatin (LOC215776), submitted recently by the Genome Annotation Project, and a 3491-bp sequence derived from an IMAGE clone from a mammary tumor library (GenBank Accession No. BC006883). Both sequences from GenBank have a different carboxyl terminus. Alignment of our cDNA sequence with mouse genomic sequences by BLAST analysis indicates at least 27 coding exons for isoform A.

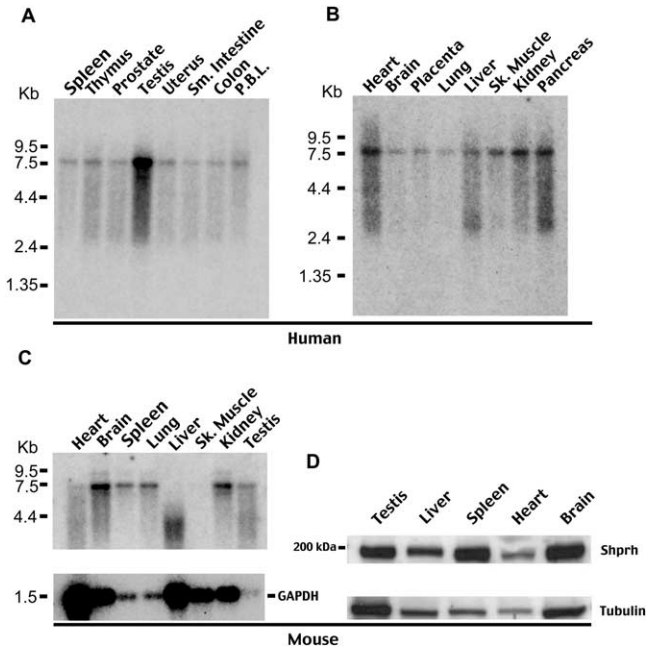


Fig. 1. Northern and Western blot analyses to determine the transcript size and tissue distribution. Northern blot analysis in (A, B) 16 adult human tissues and (C) 8 adult mouse tissues. The tissues are marked on top of each lane and size markers in kb are indicated on the left side of each blot. A GAPDH probe was used on the same blot as a control. Mouse heart, liver, and skeletal muscle RNA appears to be degraded, thereby not exhibiting a 7.5-kb transcript for the Shprh probe. (D) Western blot analysis of Shprh in 5 adult mouse tissues using our polyclonal antibody. Tubulin antibody was used on the same blot as a loading control.

Cloning of human homolog (SHPRH)

The murine *Shprh* cDNA sequence was used to identify two Unigene clusters, Hs.231907 and Hs.213335, representing 5′ and 3′ ends of the gene. In addition, a single EST, D81883, was found to represent the middle of the gene. We used PCR to cover the gaps between the available EST sequences, followed by 5′ RACE to obtain the amino-terminal sequence not covered by the ESTs. The final 6018-bp full-length cDNA sequence contains an open reading frame of 5049-bp coding for a protein of 1683 amino acids (GenBank Accession No. AY161136). There is an in-frame upstream stop codon, indicating identification of the complete coding sequence for this gene. Comparison of our cDNA sequence with the nonredundant nucleotide database identified a predicted mRNA annotated as similar to helicase (LOC257218), recently submitted by the Genome Annotation Project. The prediction differs from our cloned version in the complete absence of a 93-bp exon (exon 26), different 3′ splice acceptor sites for exons 14 and 19, and different 5′ and 3′ ends.

Expression analysis

We examined the size and tissue distribution patterns of both human and mouse transcripts by Northern blot analy-

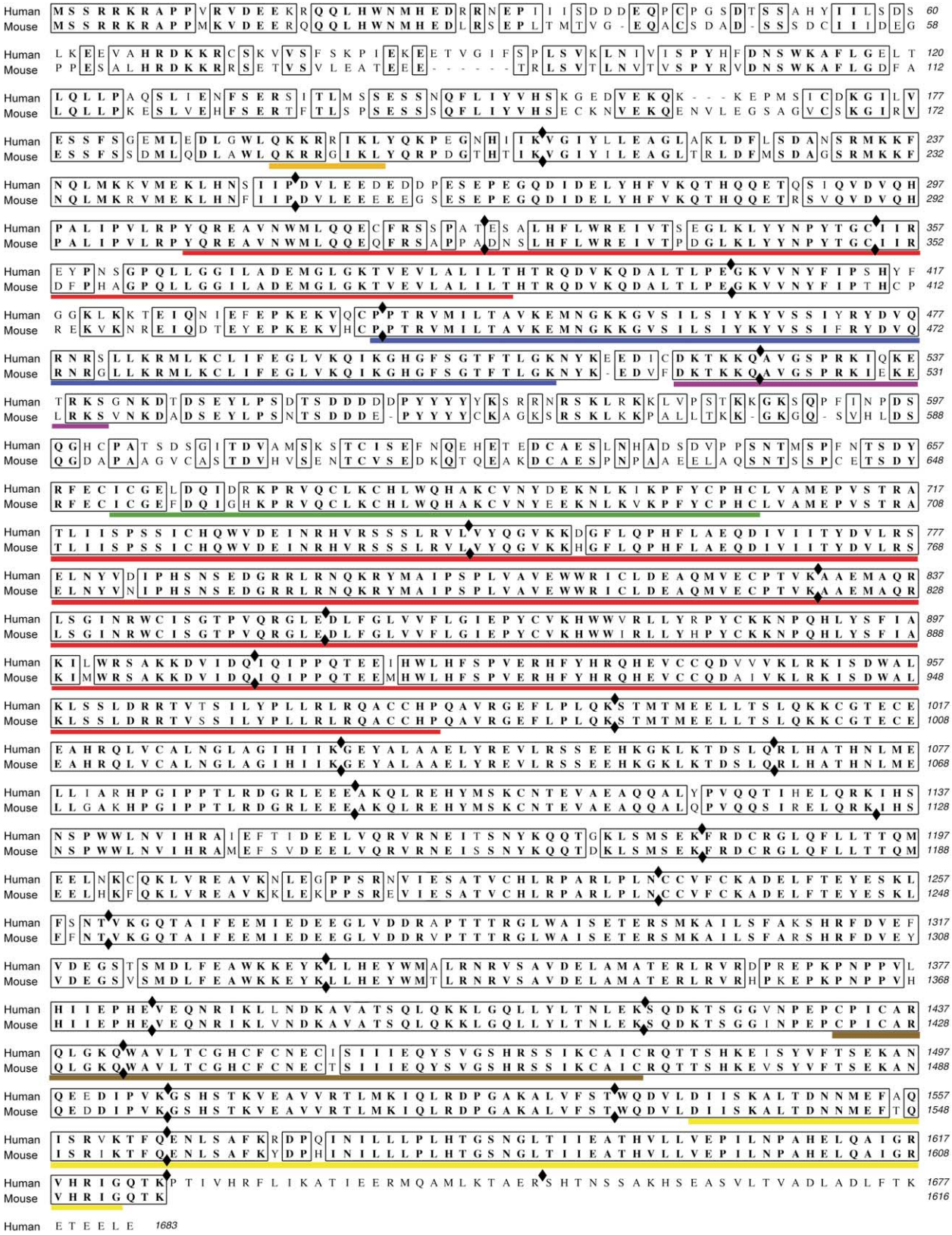
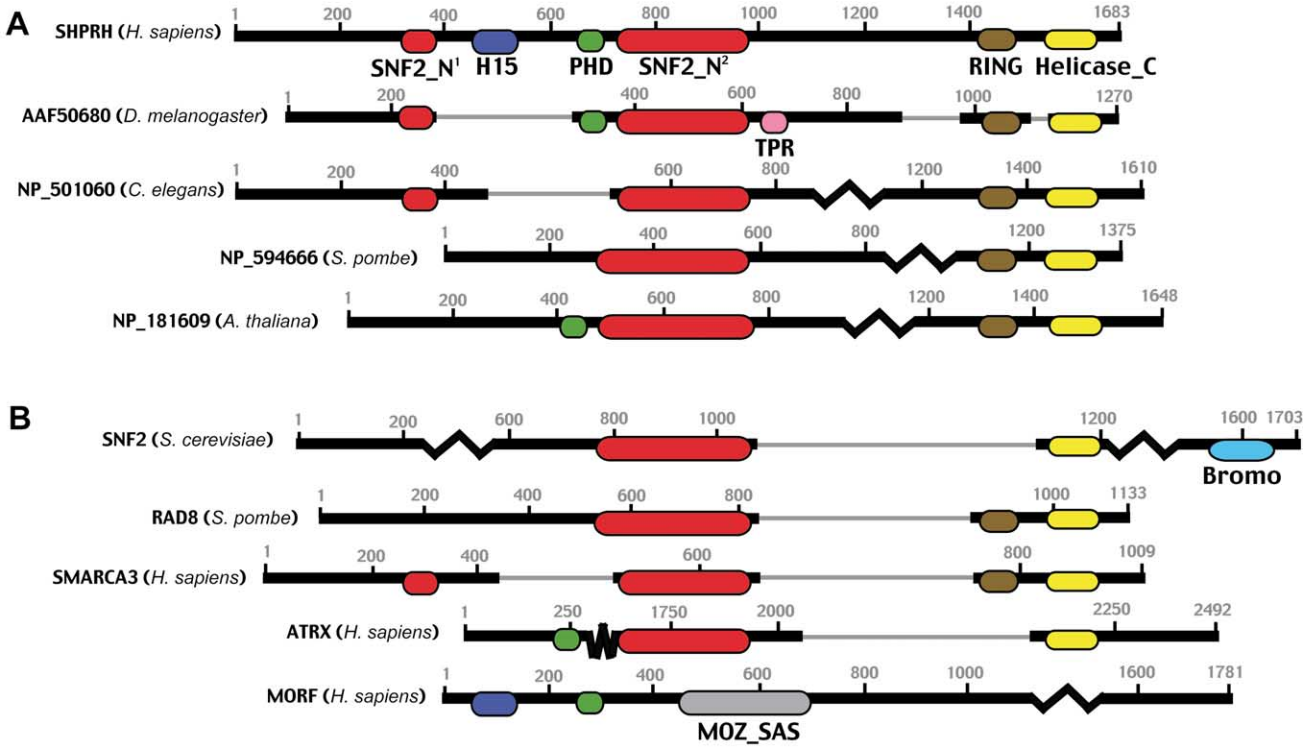
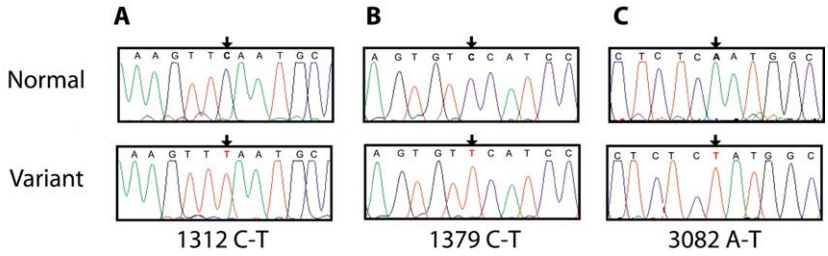
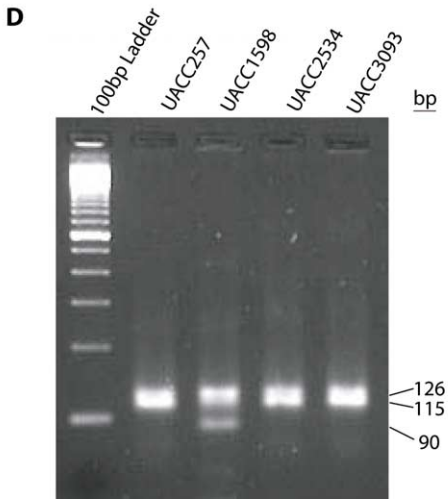


Fig. 2. Alignment of the deduced amino acid sequences of the human and mouse SHPRH proteins. The alignment was performed using the ClustalW program of the Macvector software package (Genetics Computer Group, Madison, WI, USA). Identical (boldface) and similar amino acids are shown in boxes. Various domains and motifs are underlined as follows: nuclear localization signal, orange; SNF2 domains, red; linker histone domain, blue; PHD-finger domain, green; RING-finger domain, brown; helicase domain, yellow. The peptide used in antibody generation is underlined in purple. Exons are indicated by filled diamonds above the alignment (human) and below the alignment (mouse).



3





4

Fig. 3. Diagrammatic representation of (A) the domain organization of SHPRH and its comparison with putative orthologs and (B) related proteins with similar domain organizations. Horizontal black lines represent different proteins, with the positions of amino acids marked at the top. The protein sequence is stretched (with a thin gray line) or compressed (with wavy line) to align the identical domains with each other.

Fig. 4. SHPRH mutations in tumor cell lines. Representative chromatograms showing the homozygous changes (A) 1312C–T, (B) 1379C–T, and (C) 3082A–T, with normal sequence at the top and variant at the bottom. Only one of the observed changes, 1313A–G, led to the creation of a new restriction site. (D) *Taq*I digest of cell line UACC1598 and three other cell lines without the change. The 241-bp PCR product for exon 7 has a *Taq*I site at position 126, thereby generating two products of 126 and 115 bp, observed as a double band. In the variant form, another *Taq*I site is created, leading to 90- and 25-bp fragments instead of a 115-bp fragment.

sis. A transcript of ~7.5 kb was detected for *SHPRH* in all tissues examined, albeit at varying levels (Figs. 1A and 1B). Similarly a major band at 7.5 kb is detected in brain, spleen, lung, kidney, and testis with the *Shprh* probe. In addition, *Shprh* shows a minor band of ~9.5 kb in brain, spleen, and kidney (Fig. 1C). These sizes are slightly larger than the cloned cDNA sizes, especially for *SHPRH*, indicating that there may be more sequence at either the 5′UTR or the 3′UTR.

Immunoblot analysis of Shprh in several mouse tissues revealed a protein band of ~200 kDa in all tissues examined (Fig. 1D). The predicted molecular weight based on the primary sequence in the absence of any posttranslational modifications is 185 kDa for mouse Shprh protein and 193 kDa for human SHPRH protein. Analysis of tumor ears from TG3 animals by RT-PCR and immunoblot analysis did not reveal any evidence for aberrant expression of Shprh (data not shown).

Protein characterization of conserved domains

Comparison of the predicted amino acid sequences of mouse and human *SHPRH* genes shows 86% amino acid identity and 92% similarity. The nucleotide sequence identity between the two cDNA sequences is 85%, which is an average degree of identity between human and mouse orthologs [16]. The alignment of two proteins shows that apart from 58 extra amino acids at the carboxyl terminus of SHPRH eight small gaps of 1 to 6 amino acids are inserted between the two sequences (Fig. 2). Most of the divergence between the two proteins is seen at the amino termini. Protein database analysis revealed the strongest overall homology with a putative SWI2/SNF2 family transcription factor from *Arabidopsis thaliana* (NP_181609), indicating that this novel gene is evolutionarily conserved. There is roughly 50% similarity with hypothetical proteins deduced from putative open reading frames of several other species, such as *Drosophila melanogaster* (CG7376 gene product), *Schizosaccharomyces pombe* (NP_594666 putative DNA repair protein), *Neurospora crassa* (AL451013 conserved hypothetical protein), and *Caenorhabditis elegans* (NP_501060-helicase). Most likely these putative proteins represent the orthologs of human and mouse SHPRH, thus demonstrating conservation throughout evolution. However, no biological function has been described for any of these proteins.

In the absence of homology to any known proteins that can give us clues to the function of this gene, we performed a search for functional domains against the conserved domain database. This analysis showed that both proteins contain two partial SNF2 domains and one each of PHD-finger, RING-finger, linker histone, and DEAD helicase domains (Figs. 2 and 3). The SNF2 domain (CDD7500), consisting of 294 amino acids, is found in proteins involved in transcription regulation, DNA repair, DNA recombination, and chromatin unwinding [17]. SHPRH contains two partial SNF2 domains, shown in red in Figs. 2 and 3. Proteins containing an SNF2 domain also contain a helicase domain (CDD7372), although no helicase activity has been demonstrated for these proteins. This 78-amino-acid domain is shown in yellow in Figs. 2 and 3. The PHD-finger domain (CDD7700) is characterized by a C4HC3 arrangement of cysteines and histidine with intervening sequences of varying length and composition, whereas the RING-finger domain (CDD7314) has a C3HC4 arrangement. Both domains are capable of binding zinc ions and occur in a variety of functionally distinct proteins [18,19]. The linker histone domain (CDD475), shown in blue (Figs. 2 and 3), is characteristic of histones of the H1 and H5 families, hence termed H15. In addition, amino acids 193–200 of human SHPRH protein represent a nuclear localization signal, underlined in orange in Fig. 2. This motif is involved in transport of nuclear proteins to the nucleus.

Chromosomal localization and genomic organization

We assembled a contig of BAC and PAC clones spanning roughly 1 Mb of the syntenic region in human (6q24) in an effort to identify all transcript units as candidates for a melanoma susceptibility gene in TG3 mice. The BAC/PAC clones representing the minimal tiling path were submitted to the Sanger Center for sequencing. The genomic structure of *SHPRH* was determined by the alignment of the cDNA sequence to the finished sequence of two overlapping BAC clones (GenBank Accession Nos. AL451145 and AL356599). Our data show that the *SHPRH* gene has 30 exons and spans across a genomic region of ~78 kb. Table 1 lists the positions of exons and adjacent intronic sequences. Exons range in size from 63 to 665 bp, and introns range from 0.1 to 15.5 kb. All introns have splice donor and acceptor sites that conform to the general GT-AG consensus motif (Table 1). The putative translation initiation codon is located in exon 2 and the stop codon is in exon 30. Genomic organization of *SHPRH* is very similar in human and mouse, as indicated by filled diamonds above the alignment for human and below the alignment for mouse exons in Fig. 2. There is 1 extra intron in the mouse that splits exon 16 into 2 exons.

Mutation analysis of SHPRH in melanoma and breast and ovarian tumor cell lines

Loss of heterozygosity (LOH) spanning 6q24 is seen in several cancers, particularly melanoma, breast, and ovarian cancers. We screened the entire coding sequence of *SHPRH* for mutations in 44 tumor cell lines, including 37 melanoma, 2 prostate, 4 breast, and 1 ovarian cancer, available in our laboratory. Each exon was amplified using intronic primers and bidirectionally sequenced to screen both the coding sequence and the splice sites for nucleotide changes among the cell lines. All changes believed to cause a significant effect on the protein function were evaluated further by their frequency of occurrence in controls (95 Caucasian

Table 1
Exon–intron organization of the human *SHPRH* gene

Exon	Exon size (bp)	5′-splice donor site	Intron size (kb)	3′-splice acceptor site
1	319	AACTGAgtgggt	8.4	ctctagAGTTCA
2	665	ATCAAGgtactc	2.2	ttttagGTTGGA
3	130	TTCCAGgtaata	1.9	gtacagATGTGT
4	219	CTACTGgtaaga	1.9	ctttagAAAGTG
5	79	AGGCTGgtaagg	0.6	cctcagCATCAT
6	151	CCCGAGgtagga	1.1	tatcagGGAAAA
7	109	GCCCTCgtaagt	0.6	atttagCTACAC
8	257	AAGCAGgtaaca	1.6	tcatagGCAGTA
9	660	GTCTTGgtaagg	1.3	tcctagGTATAT
10	252	GTAAAAgtgaga	0.8	catcagGCTGCA
11	79	TAGAGGgtaatg	5.3	tttcagATCTTT
12	164	GACCAAgtcagt	0.1	tcacagATCCAA
13	257	AAAAAGgtttta	1.7	ttttagCACCAT
14	121	TTAAAGgtagaa	5.8	ctttagGTGAGT
15	90	CTTCAAgtaagt	0.9	ttgtagAGACTT
16	93	GAAGAGgtaatt	1.3	acacagGCCAAA
17	108	AGAAAGgtatgc	0.9	ttacagATTCAT
18	143	AGAGAAgtaaga	1.8	aatcagGTTCCG
19	174	CAACTGgtaggt	0.3	tttcagCTGTGT
20	63	CAACACgtaagt	0.9	tttcagAGTCAA
21	226	TATAAGgtatag	1.7	cttcagTTGCTT
22	144	CATGAGgtaatt	1.1	ctttagGTAGAA
23	102	GAGAAGgtattg	4.6	ggatagTCTCAA
24	72	AAACAGgtaaaa	2.9	ttttagTGGGCG
25	189	GTGAAGgtaaag	15.5	tcaaagGGCAGC
26	93	TCAACGgtataa	0.6	ttctagTGGCAA
27	87	TTTCAGgtatgt	0.8	ttttagGAGAAC
28	179	GACAAAgtaagg	5.1	ttttagACCTAC
29	81	GAGAAGgtagggt	1.2	ttgcagTCACAC
30	705			

samples from Corriell, 73 CEPH parents, and 47 Australians). A total of six nucleotide changes were observed, four of which are listed in Table 2 and are most likely to have an effect on the protein. The other two changes had the hallmarks of a polymorphism, one being intronic, the other being a silent change, and both occurring in several cell lines. Three of the four variants were observed as homozygous changes, indicating hemizygosity with LOH affecting the other allele. For two of these cell lines, MM253 and MM370, Walker and colleagues [20] have shown reasonable evidence for LOH using several markers spanning the interval 6q23–q26. Fig. 4 shows representative chromatograms for the homozygous changes and a restriction digest of the PCR product with *Taq*I for the heterozygous change 1313A–G. No change in restriction pattern was observed for the three homozygous changes. Apart from the truncating mutation, it is difficult to say how the other changes might affect the protein function. Two of them, serine at codon 460 and asparagine at codon 1028, affect amino acids that are conserved in the mouse protein and only the Ser460Phe variant affects a known domain, the linker histone domain. All other variants fall outside of the known domains in the protein.

Discussion

In this study we have cloned and characterized a novel human gene with five functional domains and, therefore, designated it as the *SHPRH* gene (SNF2, histone linker, PHD, RING, helicase). We have also identified the mouse homolog, termed *Shprh*, which revealed a high degree of conservation at nucleotide and amino acid sequence levels, suggesting similar functions. Both genes have identical genomic organizations, except for one extra intron in the mouse that splits coding exon 16 into two exons. Both genes are ubiquitously expressed as shown by Northern and Western blot analyses. The Genome Annotation Project utilizing available human and mouse genomic sequences has predicted both transcripts with a high degree of similarity to our cloned cDNAs. However, the prediction for the mouse transcript is only partial, lacking 2168 bp (723 amino acids) of the amino terminus and the entire 5′ and 3′ untranslated regions. The prediction for human *SHPRH* has several differences from our cloned version, such as different amino and carboxyl termini, absence of exon 26, different 3′ splice acceptor sites for exons 14 and 19, adding 27 bp at the end of exon 14, and missing 15 bp at the end of exon 19. Gene prediction programs will definitely help us in identification of the complete transcriptome upon completion of the Human Genome Project. However, validation using wet lab techniques would be required prior to initiation of functional studies, such as choosing a peptide to generate antibodies, as is clear from this example. In addition, cDNA clones for both human and mouse homologs of this novel gene would provide a tangible source for further functional studies.

We performed homology searches against the conserved domain database to gain preliminary insight into the function of this newly identified gene. The protein contains five functional domains, SNF2-N, linker histone, PHD finger, RING finger, and helicase, and probably belongs to the

Table 2
Mutations in *SHPRH* gene in tumor cell lines

Cell line	Tumor type	Nucleotide change	Coding change	Mutated domain	Control data
UACC1940	Melanoma	1312C–T	Gln438X	None	0/336 chromosomes
UACC1598	Ovarian	1313A–G	Gln438Arg	None	0/336 chromosomes
MM370	Melanoma	1379C–T	Ser460Phe	Linker histone	0/268 chromosomes
MM253	Melanoma	3082A–T	Asn1028Tyr	None	0/274 chromosomes

superfamily of SWI2/SNF2-related helicases. However, none of the previously characterized proteins have the exact same domain arrangement as SHPRH. A homology search indicates that this gene is evolutionarily conserved and the predicted proteins from *D. melanogaster* (AAF50680), *A. thaliana* (NP_181609), *C. elegans* (NP_501060), *S. pombe* (NP_594666), and *N. crassa* (CAC18163) are most likely orthologs of SHPRH (Fig. 3).

Due to lack of homology to any well-characterized protein, we can only infer the function of SHPRH from its domains. The largest domain seen in SHPRH is SNF2. This domain was initially identified in yeast SNF2 protein, which is a member of the SWI/SNF complex involved in regulation of transcription by chromatin remodeling [17,21]. This domain is observed in several mammalian proteins involved in a variety of cellular processes, such as DNA repair (ERCC6, Rad8), transcription activation or repression (brahma, Mot1), DNA recombination (Rad54), and maintenance of chromosome stability during mitosis (lodestar). Most proteins with SNF2 homology also contain a putative helicase domain that uses ATP to relocate nucleosomes along the DNA, thus facilitating transcription [22]. Mutations in ERCC6, a member of the SNF2 family of helicases, are found in Cockayne's syndrome [23].

The PHD (plant homology domain)-finger domain is a unique zinc finger defined by C4HC3 spatial arrangement of seven cysteines and one histidine, as opposed to the C3HC4 pattern of the RING-finger domain. Although function of the PHD-finger domain is not known, mutations in this domain of the *ATRX* gene are associated with α-thalassemia, mental retardation, and developmental defects [24]. It is interesting to note that the *ATRX* gene is also a SNF2/helicase (Fig. 3). Mutations in ATRX also affect the pattern of methylation of several highly repeated sequences [25], thus suggesting a role in regulation of gene expression via DNA methylation. One of the well-characterized examples of a protein with a RING-finger domain is BRCA1, a tumor suppressor gene involved in familial breast and ovarian cancers [26]. Although a RING-finger domain is found in several functionally unrelated proteins, recent evidence points to a general function of the RING-finger domain in ubiquitin-mediated proteolysis, thus maintaining cellular protein levels [19]. H15 is a domain found in linker histone families 1 and 5. Linker histones are involved in transcriptional regulation by blocking or providing access to the regulatory regions by transcription factors [27,28]. An example of a protein with an H15 domain is MORF, which is a histone acetyltransferase involved in myelogenous leukemia by producing a MORF–CBP chimera [29,30]. Based on our knowledge of various domains discussed above and the presence of a nuclear localization signal, we would predict that SHPRH plays a role in chromatin-mediated transcriptional regulation. However, functional studies are required to determine the exact biological role of this gene, as the domains may have an altered function due to their interaction with each other and the remainder of the protein.

LOH at 6q, particularly band region q24–q27, has been observed in multiple human malignancies [2,3,6,9,11]. *SHPRH*, by virtue of its localization to this region and its possible role in transcription regulation, becomes a candidate tumor suppressor gene. Critical to this argument, previous studies have shown that subunits of the SWI/SNF complex act as tumor suppressors in human and mice [31]. In our preliminary mutation analysis of tumor cell lines, we observed one truncating mutation and three other variants not seen in >250 normal chromosomes. Three of the four variants, including truncating mutation, were observed in the hemizygous or homozygous state, thus indicating a complete lack of normal protein in these cell lines. Given the low frequency of mutations observed in this study, other mechanisms for inactivation of SHPRH, such as large deletions, rearrangements, and regulatory mutations affecting expression, need to be assessed to confirm the role of SHPRH as a tumor suppressor. It would be particularly helpful to screen SHPRH in cell lines and primary tumors with evidence for LOH of the 6q24 region.

In conclusion, we have described the cloning and characterization of both human and mouse homologs of a novel gene with a unique domain organization as a possible candidate for a tumor suppressor gene on human chromosome 6. Furthermore, we have generated polyclonal antibodies to analyze the protein expression and localization in different tissues as a first step toward understanding the function of this gene.

Materials and methods

Rapid amplification of cDNA ends and interexon PCR

5′ RACE was performed using the Marathon-Ready RACE System (Clontech, Palo Alto, CA, USA) according to the specifications of the manufacturer. Marathon-Ready cDNA libraries derived from human and mouse brain and testis were used as templates for both RACE and interexon PCRs. Antisense gene-specific primers were designed from the partial cDNA sequence. Primary RACE reactions were purified from unincorporated nucleotides and primers using the QIAquick PCR purification kit (Qiagen, Valencia, CA, USA) and then used as template for secondary RACE reactions using the AP2 adaptor primer and nested gene-specific primers. Products were cloned using TOPO-TA cloning (Invitrogen, Carlsbad, CA, USA). Interexon PCRs were performed in 50-μl reaction volumes containing 5 μl of the appropriate cDNA as template using Advantage cDNA polymerase (Clontech) and their recommended protocol.

Clone characterization

IMAGE clones obtained from Research Genetics (Invitrogen), RACE clones, and interexon PCR clones were

processed as follows: DNA was extracted from a 3- to 5-ml overnight bacterial culture using the Wizard Plus Minipreps DNA Purification System (Promega, Madison, WI, USA). Clones were sequenced using primers from both ends of the vector followed by walking primers designed from the initial sequence data. DNA sequencing reactions were performed using ~0.5 μg DNA per reaction and PE Biosystems' FS+ and Big Dye Terminator chemistry [32] and were analyzed on a PE Biosystems 377 XL DNA sequencer.

Mutation analysis

The genomic structure of *SHPRH* was used to design primers from intronic sequences flanking the exons. Primers were modified by the addition of M13 tails for subsequent sequence analysis. PCRs were carried out using 20 ng of template DNA with 2.25 mM Mg^{2+}, 250 nM dNTPs, 10 pmol forward and reverse primer mix, PCR buffer II (Perkin–Elmer, Norwalk, CT, USA), and 0.6 units of AmpliTaq Gold polymerase (Perkin–Elmer) in a 50-μl total PCR volume. PCR was carried out in a Model 9700 thermocycler (Perkin–Elmer) using the following cycling protocol: initial denaturation at 94°C for 10 min and then 94°C for 20 s, 57°C for 20 s, 72°C for 20 s for 35 cycles followed by a final extension at 72°C for 10 min. Annealing temperatures were different from 57°C for a few primer sets. Primer sequences and annealing temperatures are available upon request. All PCRs were analyzed for robustness of amplification by running a 5-μl aliquot on 2% agarose gels. Fluorescence sequencing reactions were performed using the Big Dye Terminator cycle sequencing kit (Applied Biosystems, Foster City, CA, USA) at 1/4 strength reaction. All samples were cleaned up using Sephadex plates (Millipore, Burlington, MA, USA) and loaded on an ABI 3700 DNA analyzer (Applied Biosystems). For validation of observed sequence changes, restriction digest of 5 μl of PCR product was performed using the manufacturer's recommendations and analyzed on a 2% agarose gel.

Northern blot analysis

Northern blots with mRNA from multiple human and mouse tissues were purchased from Clontech. The probes (PCR products covering part of the coding region of both genes) were labeled with [α-^{32}P]dCTP by random priming (Stratagene, La Jolla, CA, USA) following the manufacturer's directions. Hybridization was carried out at 42°C overnight in Hybrisol 1 hybridization buffer (Intergen, Purchase, NY, USA) followed by stringent washing in 0.1% SDS, 0.1× SSC at 65°C. Autoradiography was performed for 1–2 days at −80°C with a BioMax intensifying screen and film.

Antibody generation and Western blots

Polyclonal antibodies were generated by immunization of rabbits with a synthetic peptide, CDKTKKQAVGSPRKIEKELRKS, representing amino acids 515–535 of the deduced protein sequence of Shprh and an additional cysteine at the N-terminus for coupling by Genosys (Sigma, St. Louis, MO, USA). There are only two amino acid substitutions in this region between human and mouse homologs. Tissue extracts were prepared at 4°C (from frozen samples) using a Polytron (Brinkmann Instruments, Westbury, NY, USA) in lysis buffer 1 (50 mM Tris–HCl, 150 mM NaCl, 1 mM EDTA, 1% NP-40, 5% glycerol, and 1 mM dithiothreitol, pH 7.4) with protease inhibitors (Complete protease inhibitor cocktail tablets; Roche Molecular Biochemicals, Indianapolis, IN, USA). Protein concentrations were measured using the Bio-Rad DC Protein Assay Kit (Bio-Rad, Hercules, CA, USA). Equal amounts of lysates (25 μg) were loaded in each lane in 7.5% polyacrylamide gels (Bio-Rad). The gels were transferred to nitrocellulose membranes (Osmonics, Minnetonka, MN, USA). Membranes were probed with Shprh rabbit polyclonal antibody (2 μg/ml). Blots were visualized using the Amersham ECL Plus System (Amersham/Pharmacia, Piscataway, NJ, USA).

Computer analysis

BAC 13P17 shotgun sequencing reads were analyzed using PHRED, PHRAP, Cross_match, and Consed components of semiautomated sequence assembly and analysis suite [33,34]. PHRED was used to give quality scores followed by trimming of low quality sequence and Cross_match was used for vector screening. Trimmed sequences were assembled using PHRAP and contigs were analyzed using Consed. Contigs were analyzed using GeneMachine [35], a software suite comprising model-based and similarity-based exon/gene prediction programs.

cDNA sequences were assembled into contigs and analyzed both for open reading frames and for mutations using Sequencher, v4.1 (Gene Codes, Ann Arbor, MI, USA). Sequence homology to previously characterized DNA and protein sequences in the nr and dbEST databases of GenBank was determined using the BLAST algorithm [36] accessible from the NCBI home page (http://www.ncbi.nlm.nih.gov/). Protein sequences predicted from the cDNAs were analyzed for functional domains using the NCBI Conserved Domain Database (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Nuclear localization signal was detected using Columbia University's predictNLS program [37] accessible online at http://cubic.bioc.columbia.edu/predictNLS/.

Acknowledgments

We thank Peter Parsons and Nick Hayward from Queensland Institute of Medical Research, Australia, for providing some of the melanoma cell lines and John Lueders from NHGRI, NIH, for extraction of DNA from tumor cell lines.

References

[1] J.M. Trent, et al., Tumorigenicity in human melanoma cell lines controlled by introduction of human chromosome 6, Science 247 (1990) 568–571.

[2] J.H. Lee, J.J. Kavanagh, D.M. Wildrick, J.T. Wharton, M. Blick, Frequent loss of heterozygosity on chromosomes 6q, 11, and 17 in human ovarian carcinomas, Cancer Res. 50 (1990) 2724–2728.

[3] P. Devilee, et al., Allelotype of human breast carcinoma: a second major site for loss of heterozygosity is on chromosome 6q, Oncogene 6 (1991) 1705–1711.

[4] C.A. Griffin, et al., Consistent chromosome abnormalities in adenocarcinoma of the pancreas, Cancer Res. 55 (1995) 2394–2399.

[5] E.R. Hyytinen, et al., Genetic changes associated with the acquisition of androgen-independent growth, tumorigenicity and metastatic potential in a prostate cancer model, Br. J. Cancer 75 (1997) 190–195.

[6] N. Mazurenko, et al., High resolution mapping of chromosome 6 deletions in cervical cancer, Oncol. Rep. 6 (1999) 859–863.

[7] E.J. Perlman, J. Hu, D. Ho, B. Cushing, S. Lauer, R.P. Castleberry, Genetic analysis of childhood endodermal sinus tumors by comparative genomic hybridization, J. Pediatr. Hematol. Oncol. 22 (2000) 100–105.

[8] G. Wang, et al., Allelic loss and gain, but not genomic instability, as the major somatic mutation in primary hepatocellular carcinoma, Genes Chromosomes Cancer 31 (2001) 221–227.

[9] D. Millikin, E. Meese, B. Vogelstein, C. Witkowski, J. Trent, Loss of heterozygosity for loci on the long arm of chromosome 6 in human malignant melanoma, Cancer Res. 51 (1991) 5449–5453.

[10] G.P. Robertson, A.B. Coleman, T.G. Lugo, Mechanisms of human melanoma cell growth and tumor suppression by chromosome 6, Cancer Res. 56 (1996) 1635–1641.

[11] C. Noviello, F. Courjal, C. Theillet, Loss of heterozygosity on the long arm of chromosome 6 in breast cancer: possibly four regions of deletion, Clin. Cancer Res. 2 (1996) 1601–1606.

[12] M. Wan, T. Sun, R. Vyas, J. Zheng, E. Granada, L. Dubeau, Suppression of tumorigenicity in human ovarian cancer cell lines is controlled by a 2 cM fragment in chromosomal region 6q24–q25, Oncogene 18 (1999) 1545–1551.

[13] M.G. Tibiletti, et al., Chromosome 6 abnormalities in ovarian surface epithelial tumors of borderline malignancy suggest a genetic continuum in the progression model of ovarian neoplasms, Clin. Cancer Res. 7 (2001) 3404–3409.

[14] S. Chen, H. Zhu, W.J. Wetzel, M.A. Philbert, Spontaneous melanocytosis in transgenic mice, J. Invest. Dermatol. 106 (1996) 1145–1150.

[15] H. Zhu, et al., Development of heritable melanoma in transgenic mice, J. Invest. Dermatol. 110 (1998) 247–252.

[16] W. Makalowski, J. Zhang, M.S. Boguski, Comparative analysis of 1196 orthologous mouse and human full-length mRNA and protein sequences, Genome Res. 6 (1996) 846–857.

[17] J.A. Eisen, K.S. Sweder, P.C. Hanawalt, Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions, Nucleic Acids Res. 23 (1995) 2715–2723.

[18] R. Aasland, T.J. Gibson, A.F. Stewart, The PHD finger: implications for chromatin-mediated transcriptional regulation, Trends Biochem. Sci. 20 (1995) 56–59.

[19] P.S. Freemont, Ubiquitination: RING for destruction? Curr. Biol. 10 (2000) R84–R87.

[20] G.J. Walker, J.M. Palmer, M.K. Walters, D.J. Nancarrow, P.G. Parsons, N.K. Hayward, Simple tandem repeat allelic deletions confirm the preferential loss of distal chromosome 6q in melanoma, Int. J. Cancer 58 (1994) 203–206.

[21] C. Muchardt, M. Yaniv, When the Swi/SNF complex remodels . . . the cell cycle, Oncogene 20 (2001) 3067–3075.

[22] M.J. Pazin, J.T. Kadonaga, SWI2/SNF3 and related proteins: ATP-driven motors that disrupt protein–DNA interactions? Cell 88 (1997) 737–740.

[23] C. Troelstra, A. van Gool, J. de Wit, W. Vermeulen, D. Bootsma, J.H. Hoeijmakers, ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes, Cell 71 (1992) 939–953.

[24] R.J. Gibbons, et al., Mutations in transcriptional regulator ATRX establish the functional significance of a PHD-like domain, Nat. Genet. 17 (1997) 146–148.

[25] R.J. Gibbons, et al., Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation, Nat. Genet. 24 (2000) 368–371.

[26] Y. Miki, et al., A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1, Science 266 (1994) 66–71.

[27] J. Zlatanova, P. Caiafa, K. Van Holde, Linker histone binding and displacement: versatile mechanism for transcriptional regulation, FASEB J. 14 (2000) 1697–1704.

[28] P.T. Georgel, J.C. Hansen, Linker histone function in chromatin: dual mechanisms of action, Biochem. Cell Biol. 79 (2001) 313–316.

[29] N. Champagne, et al., Identification of a human histone acetyltransferase related to monocytic leukemia zinc finger protein, J. Biol. Chem. 274 (1999) 28528–28536.

[30] I. Panagopoulos, et al., Fusion of the MORF and CBP genes in acute myeloid leukemia with the t (10;16) (q22;p13), Hum. Mol. Genet. 10 (2001) 395–404.

[31] A. Klochendler-Yeivin, C. Muchardt, M. Yaniv, SWI/SNF chromatin remodeling and cancer, Curr. Opin. Genet. Dev. 12 (2002) 73–79.

[32] C.M. Robbins, E. Hsu, P.M. Gillevet, Sequencing homopolymer tracts and repetitive elements, Biotechniques 20 (1996) 862–868.

[33] B. Ewing, L. Hillier, M.C. Wendl, P. Green, Base-calling of automated sequencer traces using Phred, I. Accuracy assessment. Genome Res. 8 (1998) 175–185.

[34] D. Gordon, C. Abajian, P. Green, Consed: a graphical tool for sequence finishing, Genome Res. 8 (1998) 195–202.

[35] I. Makalowska, J.F. Ryan, A.D. Baxevanis, GeneMachine: gene prediction and sequence annotation, Bioinformatics 17 (2001) 843–844.

[36] S.F. Altschul, W. Gish, W. Miller, E. W. Myers, D. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.

[37] M. Cokol, R. Nair, B. Rost, Finding nuclear localization signals, EMBO Rep. 1 (2000) 411–415.