ELSEVIER

Gene 288 (2002) 119–127

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

# Gene structure, tissue expression, and linkage mapping of the mouse DLC-1 gene (*Arhgap7*)☆

Marian E. Durkin*, Bao-Zhu Yuan[1], Snorri S. Thorgeirsson, Nicholas C. Popescu

*Laboratory of Experimental Carcinogenesis, Building 37, Room 3C28, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA*

Received 12 November 2001; received in revised form 15 January 2002; accepted 29 January 2002
Received by J.L. Slightom

**Abstract**

DLC-1 (deleted in liver cancer 1) is a candidate tumor suppressor gene for hepatocellular carcinoma and other cancers. It is the human homologue of rat p122, which has been shown to function as a GTPase activating protein for RhoA, and it may be involved in signal transduction pathways regulating cell proliferation and adhesion. To establish an animal model for studying the regulation and function of DLC-1, we have undertaken the characterization of the mouse DLC-1 gene. Northern blot analysis shows that the mouse DLC-1 mRNA is widely expressed, with the highest levels in heart, liver, and lung. Mouse genomic clones that contain the entire DLC-1 gene of 47 kb were isolated. The mouse gene consists of 14 exons, and the structural organization is highly similar to that of the human gene. The promoter region of the mouse gene was GC-rich and contained potential binding sites for transcription factors SP1, GCF, and AP-2. A polymorphic microsatellite marker in intron 8 was used for mapping the gene (*Arhgap7*) to 20 cM on mouse chromosome 8 and for allelotyping of mouse liver tumor DNAs. Published by Elsevier Science B.V.

*Keywords*: Chromosome 8; Promoter; Rho GAP; Tumor suppressor gene

## 1. Introduction

Alterations in signal transduction pathways are a hallmark of cancer cells. The Rho proteins are small GTPases of the ras superfamily that regulate actin cytoskeleton organization, gene expression, and cell proliferation (Van Aelst and D'Souza-Schorey, 1997; Zohn et al., 1998). Rho GTPases have been implicated in tumorigenesis, as Rho activity is increased in transformed cells and is necessary for the transformation of fibroblasts by oncogenic Ras (Zohn et al., 1998; Sahai et al., 2001). The Rho proteins cycle between an active, GTP-bound form and an inactive GDP-bound form, and these transitions are regulated primarily by three classes of effector proteins: guanine nucleotide exchange factors, guanine nucleotide dissociation inhibitors, and GTPase activating proteins (GAPs) (Van Aelst and D'Souza-Schorey, 1997; Zohn et al., 1998). The Rho GAPs are a diverse group of proteins that possess a common 140-amino-acid GAP domain and inactivate by Rho by stimulating the hydrolysis of GTP (Lamarche and Hall, 1994).

The human DLC-1 (deleted in liver cancer 1) gene was cloned by subtractive hybridization as a gene homozygously deleted in a human hepatocellular carcinoma (HCC) DNA sample (Yuan et al., 1998). Determination of the DLC-1 cDNA sequence showed that it is the human homologue of rat p122, which has been found to act as a GAP for RhoA and to stimulate the hydrolysis of phosphatidylinositol 4,5-bisphosphate by phospholipase C-δ1 (Homma and Emori, 1995). The DLC-1/p122 aa sequence has a Rho GAP domain (Homma and Emori, 1995) and at least two other potential functional motifs, a sterile alpha motif (SAM) domain and a StAR-related lipid-transfer (START) domain (Ponting and Aravind, 1999). Overexpression of p122 in cultured cells resulted in the loss of actin stress fibers and the detachment of cells from the substratum (Sekimata et al., 1999), implicating DLC-1/p122 in the signal transduction pathways that regulate cell morphology and adhesion.

Abbreviations: aa, amino acid(s); BAC, bacterial artificial chromosome; bp, base pair(s); cM, centimorgans; EST, expressed sequence tag; GAP, GTPase activating protein; HCC, hepatocellular carcinoma; kb, kilobase pair(s); nt, nucleotide(s); ORF, open reading frame; PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends; RT, reverse transcriptase; SAM, sterile alpha motif; START, StAR-related lipid-transfer; UTR, untranslated region

☆ Sequence data from this article have been deposited with the GenBank/EMBL data libraries under accession numbers AF411435–AF411442.

* Corresponding author. Tel.: +1-301-496-5688; fax: +1-301-496-0734.
*E-mail address:* marian_durkin@nih.gov (M.E. Durkin).

[1] Present address: Toxicology and Molecular Biology Branch, Health Effects Laboratory Division, National Institute for Occupational Safety and Health, Morgantown, WV 26505, USA.

0378-1119/02/$ - see front matter. Published by Elsevier Science B.V.
PII: S0378-1119(02)00462-6

The human DLC-1 gene was mapped to human chromosome 8p22-p21.3 (Yuan et al., 1998), a region of recurrent deletions in HCC and in other cancers such as colorectal, lung, breast, ovarian, and prostate carcinomas (Emi et al., 1992; Pineau et al., 1999; Yokota et al., 1999; Lassus et al., 2001; Arbieva et al., 2000). Deletion of the DLC-1 gene was found in many primary human liver and breast tumors and in HCC cell lines (Yuan et al., 1998, 2000; Ng et al., 2000). In normal adult tissues the DLC-1 gene is widely expressed (Yuan et al., 1998; Wilson et al., 2000), but DLC-1 mRNA levels are down-regulated in liver tumors and in HCC, breast, colorectal, and prostate carcinoma cell lines (Yuan et al., 1998; Yuan et al., 2000; Ng et al., 2000). Transfection of the DLC-1 cDNA caused growth inhibition in HCC cell lines (Ng et al., 2000) and suppressed tumor formation in breast carcinoma cell lines in nude mice (Yuan et al., unpublished results). These observations strongly suggest that DLC-1 is a bona fide tumor suppressor gene.

The mouse DLC-1 cDNA has been cloned, and the sequence is highly homologous to that of the human (Yuan et al., 1999). To aid in the development of a mouse model system for analyzing the normal physiological function of DLC-1 and its role in tumorigenesis, we have determined the expression pattern, structural organization, and promoter sequence of the mouse DLC-1 gene. We have also identified a polymorphic microsatellite marker in an intron of the gene that was used to refine the localization of the gene by linkage analysis and to test for allelic loss at the DLC-1 locus in mouse liver tumors.

## 2. Materials and methods

### 2.1. Northern blot analysis of mouse DLC-1 expression

A mouse 12-tissue poly(A)$^+$ RNA Northern blot was purchased from Origene Technologies (Rockville, MD). The blot was hybridized to a 630 bp mouse DLC-1 cDNA probe generated by RT–PCR. Total RNA was extracted from adult mouse heart and reverse transcribed into cDNA as detailed earlier (Durkin et al., 2001). The cDNA template was amplified with the forward primer AGGAAGTCAGGTGTCAAA (nt 2316–2333 of the mouse DLC-1 cDNA sequence; GenBank AF178078) and the reverse primer GTGTCTGTAGTCAGCAGG (complementary to nt 2928–2945). The PCR product was labeled with [$\alpha$-$^{32}$P]dCTP by oligolabeling and hybridized to the blot in QuikHyb solution (Stratagene), as previously described (Durkin et al., 2001). As a control for loading, the blot was re-hybridized to a mouse $\beta$-actin cDNA probe.

### 2.2. Isolation and characterization of the mouse DLC-1 gene

A mouse DLC-1 genomic clone, $\lambda$G2, was isolated from a 129/SvJ DNA library in lambda FIXII (Stratagene) as previously described (Yuan et al., 1999). Sequence analysis showed that the $\lambda$G2 insert contained exons 4–9 of the DLC-1 gene. To obtain the remainder of the gene, the Genome Systems 129/SvJ ES BAC library was screened by PCR using primers derived from the sequence of exon 5. DNA from two BAC clones positive for the DLC-1 gene, 135o23 and 194h12, was purified according to the vendor's instructions, partially digested with *Sau*3A1, and ligated to lambda FIXII vector arms according to the manufacturer's protocol (Stratagene). The ligation products were packaged with Gigapack extracts (Stratagene), and the phage were screened by plaque hybridization to $^{32}$P-labeled DNA probes amplified by PCR from the BAC clone DNAs using primers derived from sequences at the 5′ and 3′ ends of the mouse DLC-1 cDNA, and from the sequence at the 5′ end of the $\lambda$G2 insert. Clones containing the rest of the gene, plus 10 kb of the 5′ flanking DNA and 5 kb of the 3′ flanking DNA, were isolated from the lambda sublibrary.

### 2.3. DNA sequence analysis

Restriction fragments of the lambda clone inserts were cloned into pBluescript (Stratagene) and sequenced using a PRISM BigDye Terminator Cycle Sequencing kit, and the products were analyzed on an ABI 377 automated sequencer (Perkin Elmer Applied Biosystems). Sequence analysis was performed using the programs of the Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, WI. Searches of the GenBank/EMBL Nucleotide Sequence databases were performed using the BLAST network of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/). The mouse genomic DNA sequence was also used to search the Mouse Gene Index of The Institute for Genome Research (located at http://www.tigr.org/tdb/mgi). Analysis of the promoter region of the mouse gene was accomplished using the PROSCAN Version 1.7 software, available online at http://bimas.dcrt.nih.gov/molbio/proscan/. The structure of the human DLC-1 gene was elucidated by performing a BLAST search of the GenBank High-Throughput Genomic Sequence (HTGS) database of unfinished BAC clone sequences, using the human DLC-1 cDNA sequence (AF035119) as the query. Exons 2–14 and the final 6528 bp of intron 1 are located in nt 1–38822 of BAC RP11-92E1 (AC068412). Exon 1 and all of intron 1 are present in BAC clone RP11-465E23 (AC084085).

### 2.4. Mapping of the transcription start point of the mouse DLC-1 gene

The 5′ end of the DLC-1 mRNA was determined by means of RNA ligase-mediated 5′ RACE, using the GeneRacer kit (Invitrogen). Total RNA from adult mouse heart (2.5 $\mu$g) was treated with calf intestine alkaline phosphatase and tobacco acid phosphatase, and the decapped RNA was ligated to the GeneRacer RNA oligo, according to the manufacturer's protocol. The RNA was reverse transcribed with AMV reverse transcriptase and random hexamer primers, and the

cDNA was amplified first with the GeneRacer 5′ Primer and a DLC-1 primer located in exon 1, CGGGATCCCACGTCGGGACCCACGGTGGC (complementary to nt 198–226 of the cDNA). The products were subjected to a second round of amplification using the GeneRacer 5′-nested primer and the same DLC-1 primer. A single PCR product of 0.3 kb was observed, which was gel-purified and cloned into the pCR4-TOPO vector (Invitrogen). The cloned 5′ RACE products were sequenced as described above. Amplification was also performed using a DLC-1 primer spanning the junction of exons 2 and 3 (GAACAGTAGATCTTCGTACAGCTGTG, complementary to nt 379–404 of the cDNA), and the 0.45 kb product was gel-purified and sequenced directly.

### 2.5. Linkage mapping of the mouse DLC-1 gene (Arhgap7)

A $(GT)_{16}$ repeat was observed in the sequence of intron 8 of the DLC-1 gene, and PCR primers flanking the repeat were designed: the forward primer 5′-ACCTGCATGCTGATCTTCTCG (nt 611–631 of GenBank AF411439) and the reverse primer 5′-GCTACACACAATCCCTCTGCC (complementary to nt 738–758 of GenBank AF411439). To test whether the repeat was polymorphic, PCR was performed using mouse genomic DNA from the C57BL/6J and *Mus spretus* (SPRET/Ei) strains, purchased from The Jackson Laboratory Mouse DNA Resource. The reactions (25 μl) employed *Taq* DNA polymerase and buffer from Promega, 50 ng of mouse genomic DNA, and primers at a final concentration of 0.4 μM each. After an initial denaturation for 2 min at 95 °C, 35 cycles of 40 s at 95 °C, 40 s at 55 °C, and 60 s at 72 °C were performed, followed by 5 min at 72 °C. The PCR products were separated on 2.5% agarose gels and visualized by ethidium bromide staining. The primers amplified an ~200 bp band in DNA from C57BL/6J mice and an ~150 bp band in DNA from *M. spretus* mice. The PCR assay was used to genotype the 94 progeny of The Jackson Laboratory interspecific backcross panel (C57BL/6JEi × SPRET/Ei) × SPRET/Ei (called Jackson BSS; Rowe et al., 1994). Data analysis and chromosomal assignment of the mouse DLC-1 gene (*Arhgap7*) were performed by The Jackson Laboratory Backcross DNA Panel Mapping Resource.

### 2.6. Genotyping of mouse liver tumors with the DLC-1 microsatellite repeat

Liver tumors were obtained from double transgenic mice generated by crossing mice homozygous for a c-*myc* transgene (on a C57BL/6J × CBA/J (B6 × CBA) genetic background) with mice homozygous for a TGF-α transgene (on a CD-1 genetic background), as previously described (Murakami et al., 1993). Genomic DNA samples from five tumors, the control B6 × CBA and CD-1 strains, and the spleen of a c-*myc*/TGF-α mouse were amplified with the DLC-1 intron 8 repeat primers as described above, except that the annealing temperature was increased to 60 °C. The products were resolved on a 10% polyacrylamide/TBE gel (Novex) and visualized with SYBR Gold nucleic acid stain, according to the manufacturer's protocol (Molecular Probes). The blue-excited fluorescent images were recorded on a Storm 860 optical scanner (Molecular Dynamics).

## 3. Results

### 3.1. Expression of the mouse DLC-1 mRNA

The tissue distribution of mouse DLC-1 mRNA was studied by Northern blot analysis. The mouse DLC-1 gene was found to be widely expressed (Fig. 1), with the highest levels in heart, liver, and lung, and the lowest in small intestine, stomach, and thymus. In mouse tissues two principal transcripts of ~6.5 and 5.5 kb were observed; the larger mRNA was more abundant in most tissues, although thymus and testis appeared to contain approximately equal amounts of both transcripts. A minor transcript of ~7.5 kb was also seen in heart. Although a similar expression pattern was observed for the human DLC-1 gene, the sizes of the DLC-1 mRNAs differed in humans, with a major transcript of 7.5 kb and a minor transcript of 4.5 kb observed on Northern blots (Yuan et al., 1998; Wilson et al., 2000).

### 3.2. Exon/intron organization of the mouse DLC-1 gene

Genomic clones containing the entire mouse DLC-1 gene were isolated from 129/Sv libraries in lambda and BAC vectors, and the structure of the gene was determined by restriction enzyme mapping and partial sequencing of the inserts. The gene was found to span ~47 kb and to consist of 14 exons (Fig. 2 and Table 1). The 13 introns range in size from 14 kb (intron 1) to 124 bp (intron 7). Mapping of the 5′ end of the mouse DLC-1 mRNA by RNA ligase-mediated 5′ RACE indicated that there was a single transcription start point. The 349 bp first exon contains a 336 bp 5′ UTR and the putative ATG translation initiation codon. Three additional ATGs are present upstream of the start of the open reading frame; the first two are followed by in-frame stop codons, and the third is in-frame with the presumptive start codon but in a less favorable context for translation initiation (Kozak, 2000).

Exon 14 contains the final 118 bp of the open reading frame, followed by the TGA stop codon. The previously reported mouse DLC-1 cDNA sequence did not contain a consensus polyadenylation signal and lacked the complete 3′ UTR. Starting 534 bp from the 3′ end of the cDNA sequence, the genomic DNA sequence matches that of a 1.26 kb EST contig (TIGR Mouse Gene Index TC202949; UniGene Cluster Mm.89698). Four ESTs in the contig (GenBank entries BE946576, AW228989, BE447323, and BE199599) have three poly(A) addition sites in an 8 nt segment downstream of an AATAAA polyadenylation signal. The length of the 3′ UTR would be 2542 bp in DLC-1 transcripts utilizing the most distal poly(A) addition site, and the predicted size of the full-length cDNA (6133

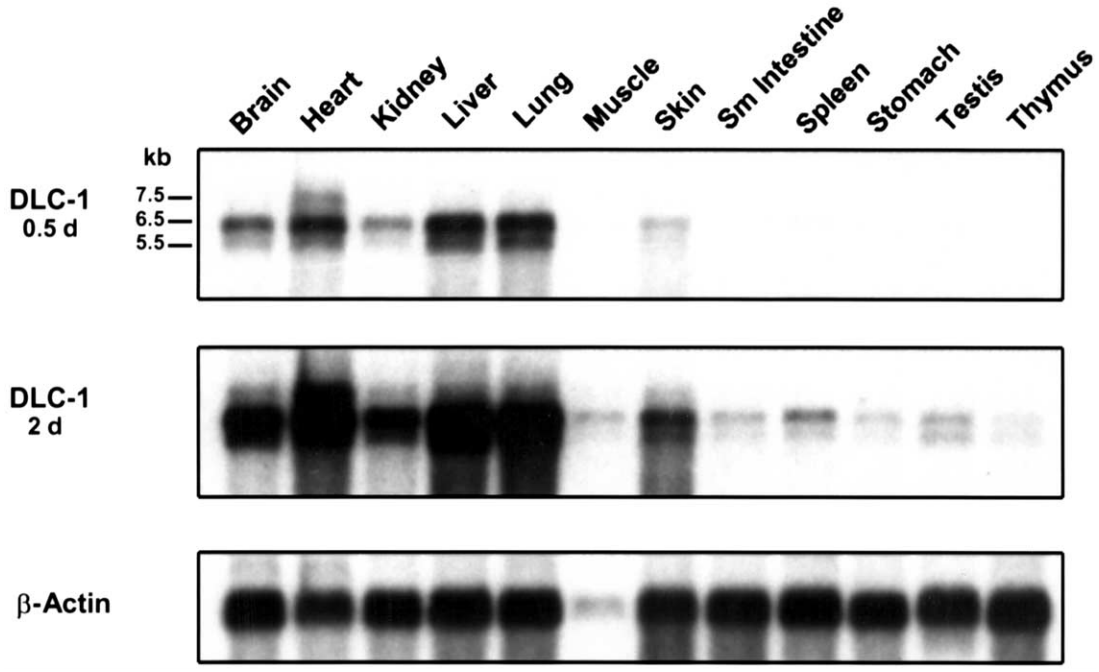


Fig. 1. Expression of DLC-1 mRNA in adult mouse tissues. The top two panels show a Northern blot of poly(A)$^+$ RNA from the indicated tissues that was hybridized to a $^{32}$P-labeled mouse DLC-1 cDNA probe. The blot was exposed to film for 0.5 and 2 days (d). The sizes of the DLC-1 transcripts, in kb, are indicated. In the lower panel, the same blot was re-hybridized to a cDNA probe derived from the 3′ UTR of mouse β-actin.

bp) is consistent with the size of the larger DLC-1 mRNA observed on Northern blots. Polyadenylation at an additional AATAAA site located 635 bp upstream could account for the size of the smaller transcript.

The mouse gene has a 3252 nt open reading frame that encodes a 1084 aa polypeptide that is 92.3% identical to the human DLC-1 and 97.5% identical to the rat p122 aa sequences (not shown). Exons 2–4 encode the SAM domain (aa 9–68), a protein-binding module present in a number of proteins involved in development and regulation, including tyrosine and serine/threonine kinases (Schultz et al., 1997) and transcription factors (Kim et al., 2001). At residue 23 of

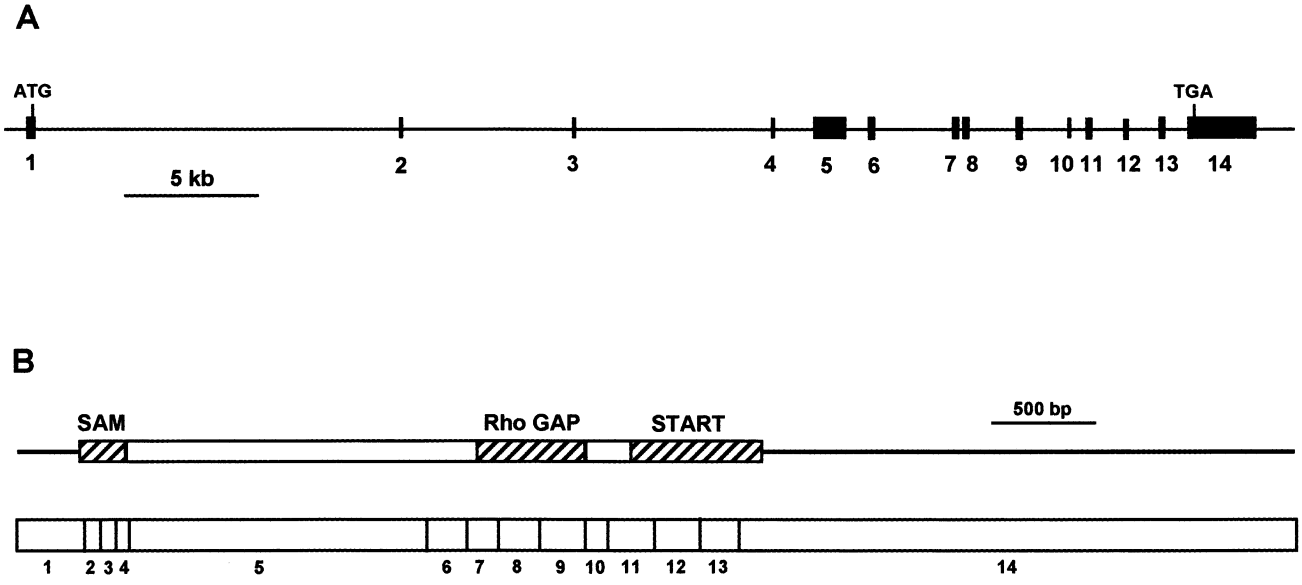


Fig. 2. Structure of the mouse DLC-1 gene. (A) Diagram of the mouse DLC-1 locus, in which the exons are represented by numbered boxes and the non-coding regions of the genomic DNA are indicated by lines. The locations of the ATG translation start codon and the TGA termination codon are shown. (B) Alignment of the coding features of the mouse DLC-1 cDNA with the exon boundaries of the gene. On top, the 5′ and 3′ UTRs of the cDNA are indicated by lines, while the open reading frame is represented by boxes. The segments of the cDNA encoding the SAM, Rho GAP, and START domains of the polypeptide are depicted as hatched boxes. Below, the positions of the exon/intron boundaries in the cDNA are indicated, and the exons are numbered.

Table 1
Exon/intron structure of the mouse DLC-1 gene

| Exon no. | Exon size (bp) | cDNA position (nt) | Exon/intron junction[a] | Intron length |
|---|---|---|---|---|
| 1 | 349 | 1–321 | AACACgtaag | 14 kb |
| 2 | 72 | 322–393 | tctagAAATT.....CGAAGgtaag | 6.4 kb |
| 3 | 82 | 394–475 | tacagATCTA.....TGCAGgtaac | 7.4 kb |
| 4 | 64 | 476–539 | tgcagGCGCT.....AGCGAgtgag | 1.7 kb |
| 5 | 1427 | 540–1966 | ttcagAGTGA.....AATAGgtaag | 663 bp |
| 6 | 177 | 1967–2143 | tgcagGCACA.....AGCTGgtaag | 3.0 kb |
| 7 | 160 | 2144–2303 | aacagGGCTG.....ACCAGgtgag | 124 bp |
| 8 | 199 | 2304–2502 | aaaagGTCGG.....CCAGTgtaag | 1.8 kb |
| 9 | 214 | 2503–2716 | tgcagATGTG.....CCAAGgtaca | 1.9 kb |
| 10 | 115 | 2717–2831 | ctcagGGTAA.....TCCAGgtaac | 613 bp |
| 11 | 219 | 2832–3050 | cccagGTCCC.....AGAAGgtcag | 1.2 kb |
| 12 | 218 | 3051–3268 | ttcagGTCAG.....TTGAGgtgag | 1.2 kb |
| 13 | 174 | 3269–3442 | cgcagGACCT.....TTAAGgtaca | 922 bp |
| 14 | 2663 | 3443–4320[b] | gacagGGGCC | |

[a] Exon and intron sequences are in upper and lower case letters, respectively.
[b] The cDNA does not contain the complete 3′ UTR.

the mouse DLC-1 sequence there is a conserved tyrosine that is phosphorylated in some SAM domain proteins (Schultz et al., 1997), suggesting that DLC-1 activity may be regulated by tyrosine kinases. Exon 5 is an unusually long coding exon (1427 bp) that, along with exon 6, encodes a serine-rich segment that separates the SAM domain from the Rho GAP domain encoded by exons 7–9 (aa 631–818). At the carboxy terminus, aa 879–1079 are homologous to the START lipid binding domain (Ponting and Aravind, 1999) and could be involved in the subcellular localization of the DLC-1 protein or in regulating the activity of the adjacent Rho GAP domain by binding to a ligand.

### 3.3. Comparison of the mouse and human DLC-1 genes

BLAST searches of the GenBank databases showed that the complete sequence of the human DLC-1 gene is available in partially sequenced BAC clones, and the structure of the gene was determined by alignment of the cDNA and genomic DNA sequences. The length of the human gene is about 50 kb and the exon/intron organization is highly similar to that of the mouse gene, except that exon 5 is 3 bp shorter in humans (Table 2). The exon boundaries of the human DLC-1 gene had been reported previously (Wilson et al., 2000), but there are several discrepancies with the gene structure deduced from the draft human genome sequence. The human DLC-1 cDNA, along with a number of DLC-1 ESTs and cDNAs, is missing the complete 3′ UTR and appears to have been primed at an $(A)_{16}$ repeat located 219 bp downstream of the stop codon. A contig of EST and cDNA sequences (Unigene Hs.8700; TIGR THC696006) contains the remainder of the 3′ UTR. The sequence of cDNA FLJ13066 fis (GenBank AK023128) ends 19 bp downstream of an AATAAA element, and polyadenylation at this site in the human DLC-1 mRNA would yield a 3′ UTR of 2447 bp.

Table 2
Exon/intron structure of the human DLC-1 gene

| Exon no. | Exon size (bp) | cDNA position (nt) | Exon/intron junction[a] | Intron length |
|---|---|---|---|---|
| 1 | >333 | 29–361 | AACACgtaag | 17.3 kb |
| 2 | 72 | 362–433 | tctagAAATT.....TGAAGgtaag | 4.8 kb |
| 3 | 82 | 434–515 | catagATTTC.....TGCAGgtaaa | 7.9 kb |
| 4 | 64 | 516–579 | tatagGCGTC.....AACGAgtgag | 2.0 kb |
| 5 | 1424 | 580–2003 | tgcagAGTGA.....AACAGgtaag | 771 bp |
| 6 | 177 | 2004–2180 | cgcagGCACC.....AGCTGgtaag | 3.2 kb |
| 7 | 160 | 2181–2340 | aacagGGCCG.....ATCAGgtgag | 128 bp |
| 8 | 199 | 2341–2539 | gacagGTTGG.....CCAATgtgag | 1.9 kb |
| 9 | 214 | 2540–2753 | tgcagATGTG.....CCCAGgtacg | 1.2 kb |
| 10 | 115 | 2754–2868 | cctagGGTAA.....TCCAGgtaag | 848 bp |
| 11 | 219 | 2869–3087 | cacagGTTCC.....AGAAGgtaag | 1.5 kb |
| 12 | 218 | 3088–3305 | tccagGTGAG.....TTAAGgtgag | 2.0 kb |
| 13 | 174 | 3306–3479 | tacagAACCT.....TTAAGgtatg | 358 bp |
| 14 | 2568 | 3480–3850[b] | gatagGGGCC | |

[a] Exon and intron sequences are in upper and lower case letters, respectively.
[b] The cDNA does not contain the complete 3′ UTR.

### 3.4. Sequence of the promoter region of the mouse DLC-1 gene

The sequence of 0.96 kb of the mouse genomic DNA upstream of the 5′ end of exon 1 was determined (Fig. 3). The region from nt −400 to the end of the first exon is characteristic of a CpG island, as it has a 63% G + C content and a CpG/GpC ratio of 0.74. No consensus TATA box sequence or CCAAT element was observed in the putative promoter region, but there were clusters of potential binding sites for transcription factors that recognize GC-rich sequences. The 5′-flanking DNA and the first exon contain 6 GC boxes that can bind transcription factor SP1, five binding sites for transcription factor GCF, and three transcription factor AP-2 binding motifs.

### 3.5. Linkage mapping of the mouse DLC-1 gene (Arhgap7)

A polymorphic microsatellite repeat was identified in intron 8 of the mouse DLC-1 gene, and the repeat was used to map the gene (*Arhgap7*) on the Jackson Laboratory BSS mapping panel (Fig. 4A). *Arhgap7* co-segregated with

```
-960  GAAGAAGCCAACTCCAGGACAGGGCTATGGGTCCCTTTTGCTGGCTGTTGTATGCTCCAG

-900  AGGAGAAACCATTTTGCCCATAGTAAACACTCTGGGTGCTGAGTGAATGCGTTTCCATCA

-840  GTTAAAGTAACACACACAAAAAGCACCAACTTACTCCAAGGTAAATTTTTATGAAAATTG

-780  ATGTGGACTTTTAAACACCTTCCACTGATTATATTCTGTCCCAAGTTTCAAAAATGGAGC

-720  TTCTGATGACATAAGGAGCCGGCAACTAAGCTGCCGGAGAATCATTTGAGCCTGTGGGGG

-660  GAAAAATGATCAGCATCTGAGATCATAAACCTGCAGGCTAGGGCCAGTTCGTGAGCAATT

-600  CTAAGAGCTGTGACCATAGTCCAAATACCAGTTAGAAATATACTGGATAGAGCTGCGCAC

-540  GCGGTATACTGACTCTCCCCACGTGCTGGGGAAAGCTTCCACAGCTGTGTTTAATCCCCA

-480  CAAACTCTTTACTCTTGCAAGCTTTAGCAAGAAATGGGTGCCCCCAAAGCCCGGATACTT

-420  CTTTTCCCTATTTTCTGACCAAGCGATCTATCTTTAACCTCCGCACTTCGCGCAAGTTTT

-360  TTTGTGATGCCAACTGGAATGAGAGGCTGCGCGTACATAAAACCACGCAAGAAAAGGCAG

-300  AGGGGAAATTGAGCACGAGGCCCGGACCTCTGCACTCAGCTAAAGCCTCCAAGCTTTGCC

-240  TTTACCCCAGAGGCCAGTAGTTGGCGACTTTTAGGCGCTTCCCGCCGCCGCAAAAGTGCA

-180  GACCCTGGGAGAACATTCCAGGAAAGGGGGAAGGAGGCGTGGGGCGGGACCTGGGCGAGC

-120  TTTGGGCGGGGCCTGGTGTGCGGTGGGTGGAGTCAGAGCACAGCCCGAGGCCGCAACGTG

 -60  AATCCCGCCCCCGGGGGCGGGCTCAGTAGGAAGGCGCGGCCCCTAGGGCGCGCAGAGAGC

   1  AGAGCACTTCTGCGTTTGGGTGGAGGGCGGTGCGCTCCAGCCTGAGCCATGGCTGTGCGT

  61  CCTCGCTGTTGGAGCCACGGCTCCCCAGCTCCGTGCCCCGCTCCCTGAGAGTGCTCCCTT

 121  CGCGGTGGCAATCTAAAACCCACGATTTTGCCCGAGCTGGGGCGAAGCGTAAGGAAGCTG

 181  CGAACCAGGATGTGCTGACGACCGCGAGGGGCTCGCGTCCCGGCTGCCACCGTGGGTCCC

 241  GACGTGGGATCCCGATTACTTCTGGCAGCCTCGACGTTCCCAGTGCGCTCCCGCCGCCCT

 301  GCGCCGACCTTAATGTGTAGAGACGAGCCGGACACCATGATCCTAACACgtaagatggac
                                   MetIleLeuThr
```

Fig. 3. Sequence at the 5′ end of the mouse DLC-1 gene. The sequence is numbered relative to the start of exon 1, as determined by RNA ligase-mediated 5′ RACE. Intron sequences are in lower case letters. Six GC boxes (GGGCGG or the reverse complement) are underlined. Shaded nucleotides denote matches to the consensus sequence for transcription factor GCF (Bina and Crowley, 2001): BVSCGSSSCB, where B = T/C/G, V = A/C/G, and S = G/C. Potential transcription factor AP-2 binding sites (Imagawa et al., 1987) are overlined: YSSCCMNSSS, where Y = C/T and M = A/C. The GCF site at nt −29 to −20 and the AP-2 site at nt −144 to −135 are on the bottom strand. The amino acid sequence, beginning at the putative translation start codon at nt 337–339, is placed below the nucleotide sequence.

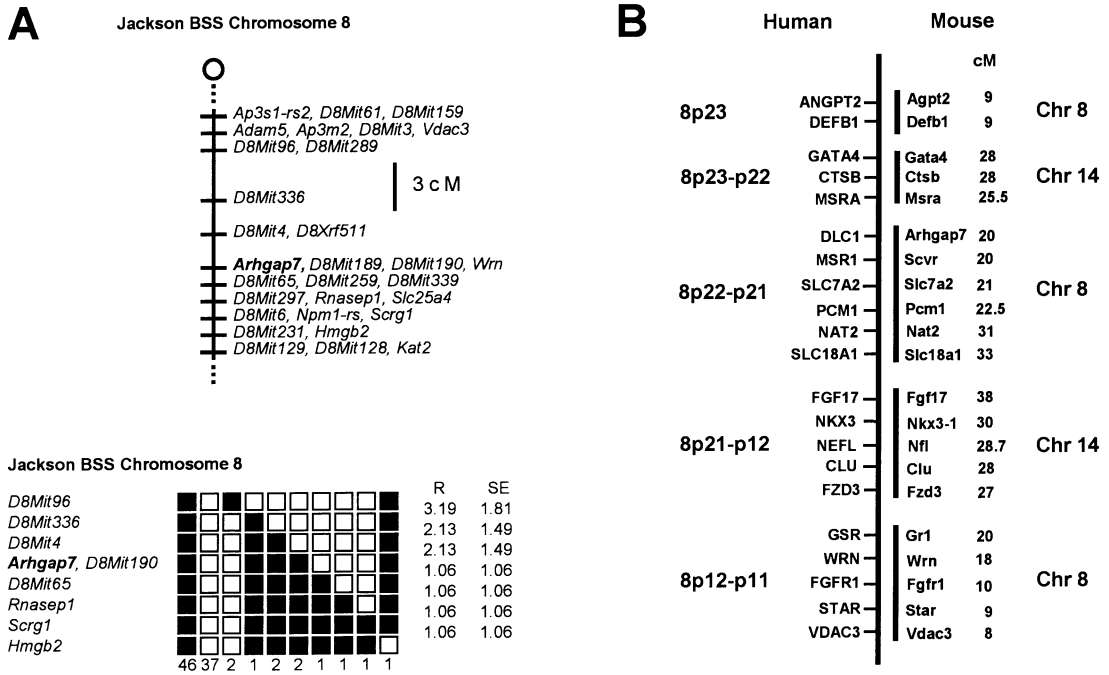


Fig. 4. Linkage mapping of the mouse DLC-1 gene (*Arhgap7*) and comparison of the gene order on the syntenic region on human chromosome 8p. (A) Map and haplotype figures from The Jackson Laboratory BSS backcross showing part of chromosome 8 with loci linked to *Arhgap7*. The map is depicted with the centromere toward the top. A 3 cM scale bar is shown to the right of the figure. Loci mapping to the same position are listed in alphabetical order. In the haplotype figure loci are listed in order with the most proximal at the top. The black boxes represent the C57BL6/JEi allele and the white boxes the SPRET/Ei allele. The number of animals with each haplotype is given at the bottom of each column of boxes. The percent recombination (R) between adjacent loci is given to the right of the figure, with the standard error (SE) for each R. Raw data from The Jackson Laboratory can be obtained from the World Wide Web address http://www.jax.org/resources/documents/cmdata. (B) Comparison of the genes on human chromosome 8p with the orthologous genes on the syntenic regions of mouse chromosomes 8 and 14. The relative order of the human genes was obtained from the NCBI Draft Sequence of the Human Genome and is subject to change. The cytogenetic locations of the genes are indicated. The positions, in cM, of the mouse orthologues on chromosomes 8 and 14 were obtained from the NCBI Human-Mouse Homology Map (http://www.ncbi.nlm.nih.gov/Homology/human8.html) and the Mouse Genome Database (http://www.informatics.jax.org/).

several markers that have been placed at about 20 cM from the centromere on the consensus map of mouse chromosome 8 (available on line at http://www.informatics.jax.org/ccr/searches/index.cgi). This is consistent with the previously reported cytogenetic localization of the gene to bands A4-B2 of chromosome 8 (Yuan et al., 1999). No known mouse mutations or tumor susceptibility loci have been mapped in the vicinity of *Arhgap7*, but the gene was tightly linked to *Wrn*, the mouse homologue of the Werner syndrome protein gene, which encodes a RecQ-like DNA helicase (Shen and Loeb, 2000). The human WRN and DLC-1 genes have been mapped to the short arm of chromosome 8 but localize to bands p12 and p21.3–22, respectively (Shen and Loeb, 2000; Yuan et al., 1998). Comparison of the gene order on human chromosome 8p11-p23 to that on proximal mouse chromosome 8 shows that the synteny has been interrupted by translocations and inversions (Fig. 4B).

### *3.6. Screening for loss of the DLC-1 gene in mouse hepatocellular carcinomas*

The polymorphic marker in the DLC-1 gene has the potential for use in loss of heterozygosity studies to determine whether deletion of the DLC-1 gene occurs during tumorigenesis in mice. Hepatocellular carcinomas develop in double transgenic mice that overexpress c-*myc* and TGF-α in the liver (Murakami et al., 1993). The PCR assay was able to discriminate between the DLC-1 alleles in the c-*myc* (B6 × CBA) and TGF-α (CD-1) transgenic lines (Fig. 5). Analysis of five HCC DNA samples from the double transgenic mice showed that allelic loss of the DLC-1 gene had occurred in one tumor.

## 4. Discussion

The DLC-1 gene encodes a putative tumor suppressor that may regulate cell growth, morphology, and adhesion. Knowledge of the DLC-1 genomic organization will provide a foundation for understanding how DLC-1 expression is controlled in normal tissues and the means by which it is down-regulated in many cancers. In this communication we report the structure of the mouse DLC-1 gene, which consists of 14 exons that extend over 47 kb of genomic DNA. The exon/intron organization of the mouse gene is very similar to that of the human, as determined from the draft sequence of the human genome. Recently two human cDNA sequences have been deposited in GenBank that are ~99.5% identical to that of

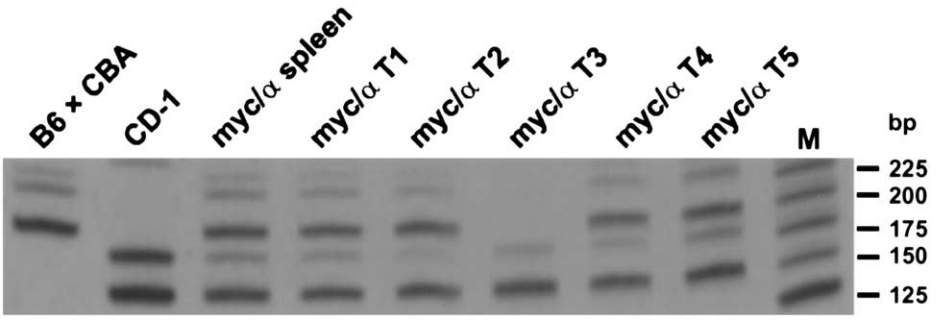


Fig. 5. Genotyping of mouse liver tumors with the DLC-1 microsatellite marker. PCR was performed with the intragenic repeat primers and genomic DNA from the control B6 × CBA and CD-1 mouse strains, the spleen of a c-*myc*/TGF-α (myc/α) double transgenic mouse, and from five liver tumors (T1 to T5) that arose in the c-*myc*/TGF-α mice. The sizes of the molecular weight markers (M), in bp, are indicated. The products amplified from the B6/CBA allele were resolved as a major band of ~200 bp and a minor band of ~230 bp, while the CD-1 allele yielded a major product of ~150 bp and a minor one of ~180 bp. The T3 tumor DNA shows loss of the B6/CBA allele.

exons 2–14 of the human DLC-1 gene but differ at the 5′ ends. The divergent 5′ ends are located upstream of exon 1 of the DLC-1 gene in the NCBI sequence map of human chromosome 8, suggesting that the DLC-1-related transcripts arise from the use of alternative promoters. A 3.7 kb cDNA (FLJ21891 fis, GenBank AK025544) appears to contain a different first exon, and a 7.4 kb cDNA (KIAA1723, GenBank AB051510; Nagase et al., 2000) has an additional 1678 bp upstream of the start of exon 2. The human DLC-1 gene may thus be part of a complex transcription unit with the potential to encode polypeptides with different amino termini. The 7.4 kb KIAA1723 cDNA would account for the size of the major human DLC-1 mRNA detected on Northern blots; however, the first ATG codon in-frame with the longest ORF is unlikely to be the start codon, according to the scanning model for initiation of translation (Kozak, 2000). The most abundant mouse DLC-1 mRNA of 6.5 kb is consistent with the size predicted from the genomic DNA and cDNA sequences, and so far only one transcription start site for the murine gene has been detected by 5′ RACE. Characterization of the endogenous human and mouse DLC-1 polypeptides will be necessary to determine which of the isoforms predicted from the cDNA sequences are present in cells.

By Northern blotting the mouse DLC-1 gene was found to be expressed in all tissues examined. The GC-rich promoter region of the mouse gene, which lacks a TATA box and has clustered binding sites for transcription factors SP1, GCF, and AP-2, is typical of the promoters of widely expressed housekeeping genes. Further studies will be required to demonstrate that this region can direct transcription and to determine whether the genomic DNA contains additional regulatory elements that are responsible for the high levels of DLC-1 mRNA in some tissues, such as heart and liver. The presence of two short upstream ORFs in the 5′ UTR of the mouse DLC-1 mRNA, which are conserved in the rat and human p122/DLC-1 sequences, indicates that expression of the gene may also be regulated at the level of translation initiation, a common occurrence in genes involved in signal transduction and growth control (Kozak, 2000).

The mouse DLC-1 gene had previously been mapped by FISH to proximal chromosome 8, and identification of a polymorphic intragenic microsatellite marker allowed us to localize the gene more precisely by linkage analysis. *Arhgap7* is tightly linked to *Wrn*, the orthologue of the human WRN gene mutated in Werner syndrome, a disorder associated with premature ageing, genomic instability, and an increased incidence of neoplasia (Shen and Loeb, 2000). Deletion of *Wrn* along with *Arhgap7* in cancers may contribute to the accumulation of genetic alterations during tumor progression. Comparative mapping of the syntenic region on human chromosome 8p shows that it has undergone breakage, inversion, and dispersal to chromosomes 8 and 14 in the mouse. Loss of heterozygosity analysis using a high density of polymorphic markers indicated that there may be several distinct minimally deleted regions on 8p in human HCC and ovarian tumors (Pineau et al., 1999; Lassus et al., 2001). It will be of interest to determine whether there is a relationship between the evolutionary breakpoints on 8p and the sites of genomic instability in tumors, and whether the DLC-1 locus resides in a region susceptible to chromosomal rearrangements.

The information on the mouse DLC-1 gene structure reported here will make it possible to investigate whether the gene is inactivated by deletions and mutations in various mouse models of tumor development. Our initial results indicate that allelic deletion of DLC-1 may occur in a subset of mouse hepatocellular carcinomas. Analysis of additional tumors will help to elucidate the role of alterations in the DLC-1 gene in the pathways that lead to HCC and other malignancies. Characterization of the murine locus will also provide a basis for analyzing the physiological function of DLC-1 through the use of gene targeting to produce mice with a null mutation in the gene.

## References

Arbieva, Z.H., Banerjee, K., Kim, S.Y., Edassery, S.L., Maniatis, V.S., Horrigan, S.K., Westbrook, C.A., 2000. High-resolution physical map and transcript identification of a prostate cancer deletion interval on 8p22. Genome Res. 10, 244–257.

Bina, M., Crowley, E., 2001. Sequence patterns defining the 5′ boundary of human genes. Biopolymers 59, 347–355.

Durkin, M.E., Keck-Waggoner, C.L., Popescu, N.C., Thorgeirsson, S.S., 2001. Integration of a c-*myc* transgene results in disruption of the mouse *Gtf2ird1* gene, the homologue of the human GTF2IRD1 gene hemizygously deleted in Williams–Beuren syndrome. Genomics 73, 20–27.

Emi, M., Fujiwara, Y., Nakajima, T., Tsuchiya, E., Tsuda, H., Hirohashi, S., Maeda, Y., Tsuruta, K., Miyaki, M., Nakamura, Y., 1992. Frequent loss of heterozygosity for loci on chromosome 8p in hepatocellular carcinoma, colorectal cancer, and lung cancer. Cancer Res. 52, 5368–5372.

Homma, Y., Emori, Y., 1995. A dual functional signal mediator showing RhoGAP and phospholipase C-δ stimulating activities. EMBO J. 14, 286–291.

Imagawa, M., Chiu, R., Karin, M., 1987. Transcription factor AP-2 mediates induction by two different signal transduction pathways: protein kinase C and cAMP. Cell 51, 251–260.

Kim, C.A., Phillips, M.L., Kim, W., Gingery, M., Tran, H.H., Robinson, M.A., Faham, S., Bowie, J.U., 2001. Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression. EMBO J. 20, 4173–4182.

Kozak, M., 2000. Do the 5′ untranslated domains of human cDNAs challenge the rules for initiation of translation (or is it vice versa)? Genomics 70, 396–406.

Lamarche, N., Hall, A., 1994. GAPs for rho-related GTPases. Trends Genet. 10, 436–440.

Lassus, H., Laitinen, M.P.E., Anttonen, M., Heikinheimo, M., Aaltonen, L.A., Ritvos, O., Butzow, R., 2001. Comparison of serous and mucinous ovarian carcinomas: distinct pattern of allelic loss at distal 8p and expression of transcription factor GATA-4. Lab. Invest. 81, 517–526.

Murakami, H., Sanderson, N.D., Nagy, P., Marino, P.A., Merlino, G., Thorgeirsson, S.S., 1993. Transgenic mouse model for synergistic effects of nuclear oncogenes and growth factors in tumorigenesis: interaction of c-*myc* and transforming growth factor-α in hepatic oncogenesis. Cancer Res. 53, 1719–1723.

Nagase, T., Kikuno, R., Hattori, A., Kondo, Y., Okumura, K., Ohara, O., 2000. Prediction of the coding sequences of unidentified human genes. XIX. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. DNA Res. 7, 347–355.

Ng, I.O., Liang, Z-D., Cao, L., Lee, T.K., 2000. DLC-1 is deleted in primary hepatocellular carcinoma and exerts inhibitory effects on the proliferation of hepatoma cell lines with deleted DLC-1. Cancer Res. 60, 6581–6581.

Pineau, P., Nagai, H., Prigent, S., Wei, Y., Gyapay, G., Weissenbach, J., Tiollais, P., Buendia, M.-A., Dejean, A., 1999. Identification of three distinct regions of allelic deletions on the short arm of chromosome 8 in hepatocellular carcinoma. Oncogene 18, 3127–3134.

Ponting, C.P., Aravind, L., 1999. START: a lipid-binding domain in StAR, HD-ZIP and signalling proteins. Trends Biochem. Sci. 24, 130–132.

Rowe, L.B., Nadeau, J.H., Turner, R., Frankel, W.N., Letts, V.A., Eppig, J.T., Ko, M.S.H., Thurston, S.J., Birkenmeier, E.H., 1994. Maps from two interspecific backcross DNA panels available as a community genetic mapping resource. Mamm. Genome 5, 253–274.

Sahai, E., Olson, M.F., Marshall, C.J., 2001. Cross-talk between Ras and Rho signalling pathways in transformation favours proliferation and increased mobility. EMBO J. 20, 755–766.

Schultz, J., Ponting, C.P., Hofmann, K., Bork, P., 1997. SAM as a protein interaction domain involved in developmental regulation. Protein Sci. 6, 249–253.

Sekimata, M., Kabuyama, Y., Emori, Y., Homma, Y., 1999. Morphological changes and detachment of adherent cells induced by p122, a GTPase-activating protein for Rho. J. Biol. Chem. 274, 17757–17762.

Shen, J.-C., Loeb, L.A., 2000. The Werner syndrome gene. The molecular basis of RecQ helicase-deficiency diseases. Trends Genet. 16, 213–220.

Van Aelst, L., D'Souza-Schorey, C., 1997. Rho GTPases and signaling networks. Genes Dev. 11, 2295–2322.

Wilson, P.J., McGlinn, E., Marsh, A., Evans, T., Arnold, J., Wright, K., Biden, K., Young, J., Wainwright, B., Wicking, C., Chenevix-Trench, G., 2000. Sequence variants of *DLC1* in colorectal and ovarian tumours. Hum. Mutat. 15, 156–165.

Yokota, T., Yoshimoto, M., Akiyama, F., Sakamoto, G., Kasumi, F., Nakamura, Y., Emi, M., 1999. Localization of a tumor suppressor gene associated with the progression of human breast carcinoma within a 1-cM interval of 8p22–p23.1. Cancer 85, 447–452.

Yuan, B-Z., Miller, M.J., Keck, C.L., Zimonjic, D.B., Thorgeirsson, S.S., Popescu, N.C., 1998. Cloning, characterization and chromosomal localization of a gene frequently deleted in human liver cancer (DLC-1) homologous to rat *Rho*GAP. Cancer Res. 58, 2196–2199.

Yuan, B.-Z., Yang, Y., Keck-Waggoner, C.L., Zimonjic, D.B., Thorgeirsson, S.S., Popescu, N.C., 1999. Assignment and cloning of mouse *Arhgap7* to chromosome 8A4-B2, a conserved syntenic region of human chromosome 8p22 → p21. Cytogenet. Cell Genet. 87, 189–190.

Yuan, B.-Z., Thorgeirsson, S.S., Eyfjord, J., Gudmundsdottir, K., Popescu, N.C., 2000. Genomic deletion, and DNA hypermethylation of DLC-1, a candidate tumor suppressor gene in several human solid tumors. Proc. Am. Assoc. Cancer Res. 41, 105.

Zohn, I.M., Campbell, S.L., Khosravi-Far, R., Rossman, K.L., Der, C.J., 1998. Rho family proteins and Ras transformation: the RHOad less traveled gets congested. Oncogene 17, 1415–1438.