J Hum Genet (2001) 46:483–486 © Jpn Soc Hum Genet and Springer-Verlag 2001

SHORT COMMUNICATION

**Satoru Matsuda · Chisako Iriyama · Shouichi Yokozaki
Yasukatu Ichigotani · Naoki Shirafuji · Kenichi Yamaki
Tetsuo Hayakawa · Michinari Hamaguchi**

Cloning and sequencing of a novel human gene that encodes a putative target protein of Nesh-SH3

Received: March 12, 2001 / Accepted: April 23, 2001

Abstract By using a conventional two-hybrid technique with an Src homology 3 (SH3) domain of Nesh as the bait protein, a novel full-length cDNA was isolated and sequenced from a human placenta cDNA library. This cDNA consists of 3023 bp and has a predicted open reading frame that encodes 486 amino acids. It possesses an SH3 binding motif, a nuclear targeting sequence, and no catalytic domain. Overall, it has no similarity to known molecules involved in a signaling cascade. Polymerase Chair reaction-based mapping with both a monochromosomal hybrid panel and radiation hybrid cell panels localized the gene on human chromosome 3q12 near the marker D3S1271.

Key words Two-hybrid · Nesh · SH3 · Chromosome 3q12 · RH mapping

Introduction

Src homology 3 (SH3) is a small domain containing about 50 amino acid residues that bind to proline-rich sequences of proteins. Biological processes often require such specific protein–protein interactions in the formation of signaling complexes (Buday 1999). In this article we describe the cloning and characterization of a novel human gene that encodes a predicted protein with a consensus sequence of SH3 binding. We named the molecule Tarsh because it seemed to be a putative target of Nesh-SH3. Nesh was recently identified as a novel protein containing an SH3 domain and proline-rich sequences similar to e3B1 (Miyazaki et al. 2000), whose biochemical and physiological role has not yet been clarified. The identification and characterization of a binding protein to Nesh is necessary to elucidate further precise mechanisms of its cellular function. Accordingly, we tried to identify potential Nesh partners to gain an understanding of Nesh function. The use of the yeast two-hybrid system enabled us to isolate the *Tarsh* gene. The interaction has not yet been characterized in detail by other commonly used protein-binding assays. However, because Tarsh has never been described before, we were prompted to further characterize this gene.

Methods, results, and discussion

We used the SH3 domain (corresponding to amino acids 294–366) of human Nesh for bait plasmids in the yeast two-hybrid screens, and identified a clone from a cDNA library constructed from human placenta poly(A)$^+$ RNA. The clone cDNA was reconfirmed to be interaction-positive by using an independent yeast clone containing the bait. DNA sequencing of the insert region was then performed by using a Sequenase Kit (Amersham, Cleveland, OH, USA) and an automatic sequencer for dideoxy sequencing (Amersham) according to the supplier's instruction. We used the insert nucleotide sequence as a probe for further screening to yield cDNAs encoding full-length and for searching the gene database to confirm the sequence and the gene structure (Matsuda et al. 1996; Matsuda et al. 2000a; Matsuda et al. 2000b; Miyazaki et al. 2000).

The determined nucleotide sequence and predicted amino acid sequence are shown in Fig. 1. The cDNA of 3023 bp contains an uninterrupted open reading frame of 1458 bp and a poly(A) tail. The Tarsh protein was thus

S. Matsuda (✉) · C. Iriyama · Y. Ichigotani · M. Hamaguchi
Department of Molecular Pathogenesis, Nagoya University School of Medicine, 65 Tsurumai-cho, Showa-ku, Nagoya 466-8550, Japan
Tel. +81-52-744-2463; Fax +81-52-735-8167
e-mail: smatsuda@med.nagoya-u.ac.jp

S. Yokozaki · K. Yamaki · T. Hayakawa
Second Department of Internal Medicine, Nagoya University School of Medicine, Nagoya, Japan

N. Shirafuji
Internal Medicine, Teikyo University School of Medicine, Tokyo, Japan

The nucleotide sequence data reported in this article have been deposited with the DDBJ, EMBL, and GenBank data libraries under accession number AB056106.

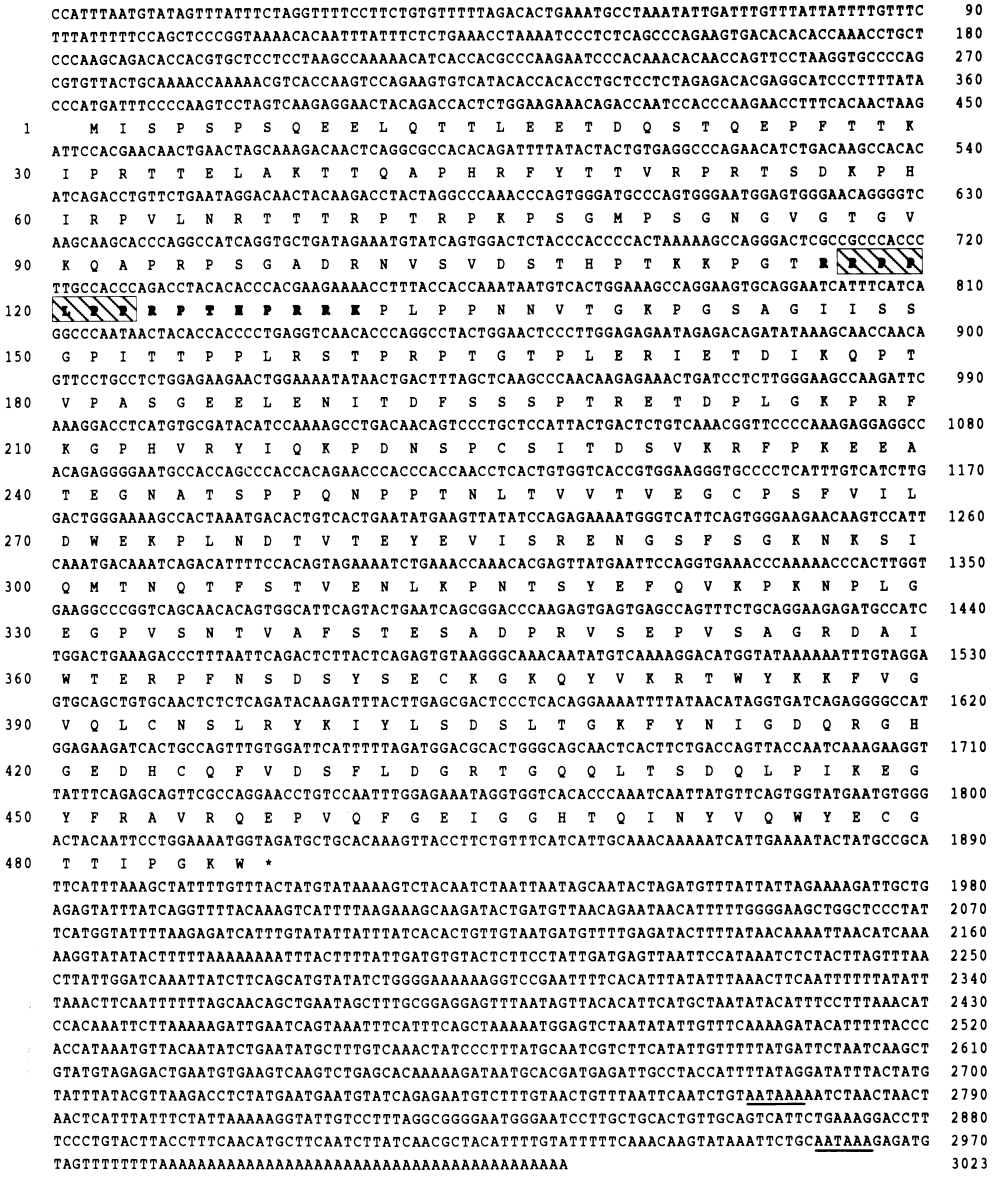

Fig. 1. Representation of the *Tarsh* cDNA and sequences. Nucleotide and predicted amino acid sequences of the *Tarsh* gene. The complete sequence of the cDNA was determined and confirmed by the primer walking strategy with dideoxy sequencing. The nucleotide sequence data have been submitted to the GenBank and EMBL data libraries under accession number AB056106. Numbering of the nucleotide and amino acid sequences is shown on the *right* and *left*, respectively. Polyadenylation signals (AATAAA) exist in the 3′-noncoding DNA sequence (nucleotides 2772–2777 and 2959–2964). The original sequence of the two-hybrid positive clone started from nucleotide 141 and ended at nucleotide 3023. The Src homology 3 (SH3) binding consensus sequence (amino acids 117–122) is indicated by a *shaded box*, and the predicted bipartite nuclear targeting sequence (amino acids 116–130) is *boldfaced*

predicted to consist of 486 amino acids. There is a potential ATG start codon (Kozak 1989) from position 364. An expected binding site to Nesh-SH3 is located in the amino-terminal region of Tarsh. The amino acid sequence RPPLPP, indicated by the shaded box in Fig. 1, is completely matched to the consensus sequence of the class 1 SH3-binding motif. Interestingly, this site overlaps the consensus sequence of the bipartite nuclear targeting signal motif (Robbins J et al. 1991). There might be an underlying mechanism by which the cytoplasmic localization of Tarsh is regulated.

Searching the nucleotide and protein databases GenBank, EMBL, EST, SWALL, and PIR revealed that Tarsh did not have high homology with other proteins. However, a fragment of human expressed sequence tag (EST) with complete nucleotide identity to Tarsh was identified in the EST database (accession number T17344). The predicted amino acid sequence of human *Tarsh* was 69% identical and 74% homological to that of a mouse embryonal partial cDNA sequence (accession number AK003938), suggesting that these are cognate genes. Analysis of the genomic sequences of *Tarsh* revealed that *Tarsh* has 18 exons. Confirming the binding activity between Nesh and Tarsh and identifying other Tarsh binding proteins might be important to further understand the biological function of Tarsh.

The tissue distribution of *Tarsh* transcript in various tissues was examined by using cycle-limited reverse transcription coupled polymerase chain reaction (RT-PCR) as described previously (Seki et al. 1999; Matsuda et al. 2000c). Primers used for RT-PCR correspond to the coding rather than noncoding region of the gene (5′-CAGCAACTC ACTTCTG-3′ and 5′-TCTGTTAACATCAGTA-3′); the expected product corresponds to nucleotides 1672–2040). Template cDNAs from the human tissues were purchased from Clontech (Palo Alto, CA, USA). As shown in Fig. 2,

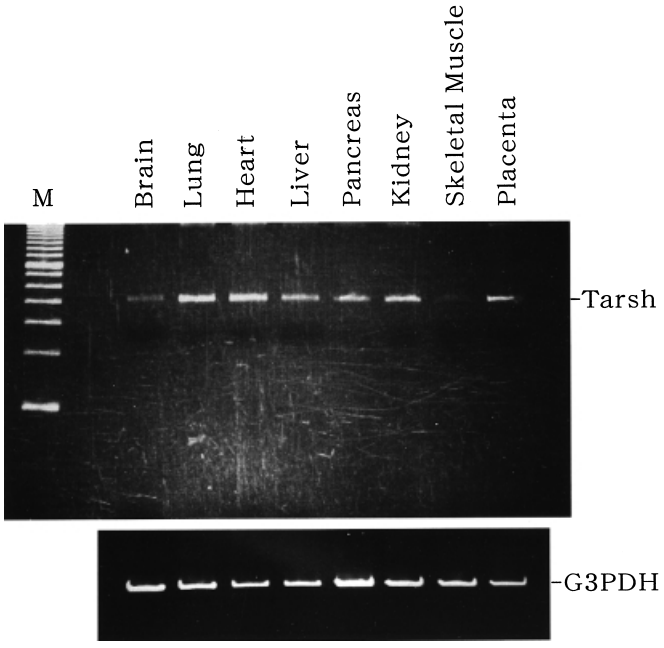

Fig. 2. Expression of the *Tarsh* gene in multiple normal human tissues. Reverse transcription and amplification by polymerase chain reaction with the specific primers for the *Tarsh* gene were performed to analyze *Tarsh* expression (*top panel*). The eight tissues examined are indicated above each lane. The template cDNA for brain (*lane 1*), lung (*lane 2*), heart (*lane 3*), liver (*lane 4*), pancreas (*lane 5*), kidney (*lane 6*), skeletal muscle (*lane 7*), and placenta (*lane 8*) of the human normal tissues were purchased from Clontech (Palo Alto, CA, USA). *Lane M*, marker. The *bottom panel* shows the expression of glycerol-3-phosphate dehydrogenase (*G3PDH*) that was analyzed as a control

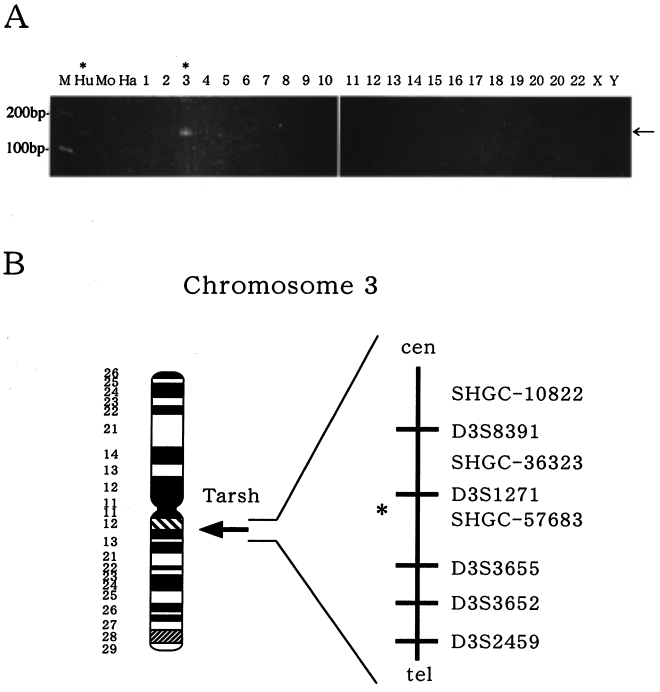

Fig. 3. Chromosomal assignment of the *Tarsh* gene. **A** Representative polymerase chain reaction (PCR)-based monochromosomal somatic cell hybrid mapping of *Tarsh*. Primers from the 3′ untranslated region of the *Tarsh* cDNA were used. *Lanes 1–22* and *lanes X* and *Y* represent human chromosomes. *Lanes Hu*, *Mo*, and *Ha* represent cell hybrids containing genomic DNA from human, mouse, and hamster, respectively. A single product of the expected size was generated from chromosome 3 and human genomic DNA (indicated by *asterisk*). *Lane M*, size marker. The *arrow* indicates the expected size of PCR product. **B** Chromosomal localization of the *Tarsh* gene in a schematic ideogram of human chromosome 3. The approximate corresponding cytogenetic location of the gene on chromosome 3q12 by the results of radiation hybrid panel mapping is shown. The *asterisk* shows the putative localization of the *Tarsh* gene. *Cen*, Centromere; *tel*, telomere

by using cycle-limited RT-PCR (28 cycle amplified), expression of *Tarsh* mRNA was detected at significant levels in several adult tissues (brain, heart, lung, liver, pancreas, kidney, and placenta) but not in skeletal muscle. This result suggests that *Tarsh* might be ubiquitously expressed except for skeletal muscle, although the level of expression varies. In this experiment genomic DNA might not be amplified, because primers designed from the noncoding region of *Tarsh* could not be amplified (data not shown). Considering its expression profile, Tarsh might have a fundamental function in a broad spectrum of adult tissue cells.

To determine the chromosomal localization of the *Tarsh* gene, we performed PCR-based monochromosomal somatic cell hybrid mapping (Quantum, Montreal, QC, Canada) with a set of 3′ untranslated region-primers (5′-TGCTGCACAAAGTTAC-3′ and 5′-CATCTAGTATTG CTAT-3′). This study indicated that the *Tarsh* gene is located on human chromosome 3 (Fig. 3A). To further refine the subchromosomal location of the gene, we used radiation hybrid (RH) panels (Stanford G3 and Genebridge 4). Both of the linkage analyses of the PCR results (the data vector for *Tarsh* of the Stanford G3 and the Genebridge 4RH panels: 10000 00000 00000 00000 00000 01010 00020 11000 10010 00001 00001 00000 00010 00010 00100 01010 000 and 00000 10110 01100 01111 01000 00110 01000 10100 11001 00010 11101 00011 01100 10010 10101 00010 00001 01010 100, respectively) showed consistently that the *Tarsh* gene was linked to several markers adjacent to marker D3S1271 in chromosome 3q12 with logarithm of differences score values higher than 16.

The most likely order of the refined loci is shown in Fig. 3B. Other genes that have been mapped to chromosome 3q12 near the *Tarsh* gene include the sporadic parathyroid adenoma gene (Thompson et al. 1995) and the tumor initiation and progression gene of multiple endocrine neoplasia type 2 (Mulligan et al. 1993). Although there is no evidence of Tarsh involvement in the genesis of these diseases, our precise chromosomal positioning data contributes a positional candidate approach for the disease genes linked to this locus. Future studies will address the biological role of Tarsh with reference to Nesh binding to understand the function of Tarsh in pathogenesis.

Acknowledgments This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas and for COE Research from the Ministry of Education, Science and Culture of Japan, a Grant under the Monbukagakusho International Scientific Research Program. We thank Yu-ki Iwata for her excellent technical assistance.

References

Buday L (1999) Membrane-targeting of signalling molecules by SH2/SH3 domain-containing adaptor proteins. Biochim Biophys Acta 1422:187–204

Kozak M (1989) The scanning model for translation: an update. J Cell Biol 108:229–241

Matsuda S, Kawamura-Tsuzuku J, Ohsugi M, Yoshida M, Emi M, Nakamura Y, Onda M, Yoshida Y, Nishiyama A, Yamamoto T (1996) Tob, a novel protein that interacts with p185erbB2, is associated with anti-proliferative activity. Oncogene 12:705–713

Matsuda S, Katsumata R, Okuda T, Yamamoto T, Miyazaki K, Senga T, Machida K, Thant AA, Nakatsugawa S, Hamaguchi M (2000a) Molecular cloning and characterization of human MAWD, a novel protein containing WD-40 repeats frequently overexpressed in breast cancer. Cancer Res 60:13–27

Matsuda S, Ichigotani Y, Okuda T, Irimura T, Nakatsugawa S, Hamaguchi M (2000b) Molecular cloning and characterization of a novel human gene (HERNA) which encodes a putative RNA-helicase. Biochim Biophys Acta 1490:163–169

Matsuda S, Miyazaki K, Ichigotani Y, Kurata H, Takenouchi Y, Yamamoto T, Nimura Y, Irimura T, Nakatsugawa S, Hamaguchi M (2000c) Molecular cloning and characterization of a novel human gene (NESCA) which encodes a putative adapter protein containing SH3. Biochim Biophys Acta 1491:321–326

Miyazaki K, Matsuda S, Ichigotani Y, Takenouchi Y, Hayashi K, Fukuda Y, Nimura Y, Hamaguchi M (2000) Isolation and characterization of a novel human gene (NESH) which encodes a putative signaling molecule similar to e3B1 protein. Biochim Biophys Acta 1493:237–241

Mulligan LM, Gardner E, Smith BA, Mathew CG, Ponder BA (1993) Genetic events in tumour initiation and progression in multiple endocrine neoplasia type 2. Genes Chromosomes Cancer 6:166–177

Robbins J, Dilworth SM, Laskey RA, Dingwall C (1991) Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. Cell 64:615–623

Seki N, Hattori A, Hayashi A, Kozuma S, Ohira M, Hori T, Saito T (1999) Structure, expression profile and chromosomal location of an isolog of DNA-PKcs interacting protein (KIP) gene. Biochim Biophys Acta 1444:143–147

Thompson DB, Samowitz WF, Odelberg S, Davis RK, Szabo J, Heath H (1995) Genetic abnormalities in sporadic parathyroid adenomas: loss of heterozygosity for chromosome 3q markers flanking the calcium receptor locus. J Clin Endocrinol Metab 80:3377–3380