J Hum Genet (2003) 48:164–169
DOI 10.1007/s10038-002-0001-3

ORIGINAL ARTICLE

Jinzhong Chen · Chaoneng Ji · Shaohua Gu
Enpeng Zhao · Jianliang Dai · Lu Huang
Ji Qian · Kang Ying · Yi Xie · Yumin Mao

Isolation and identification of a novel cDNA that encodes human yrdC protein

Received: 15 November 2002 / Accepted: 17 December 2002 / Published online: 1 March 2003
© The Japan Society of Human Genetics and Springer-Verlag 2003

Abstract In the course of detecting the interaction protein of RBBP10 by yeast two-hybridization, we isolated a novel cDNA that encodes a putative human protein with yrdC domain. It is named human yrdC protein. Because the cDNA contains an open reading fragment (ORF) without a 5′ in- frame stop codon, 5′ RACE and 3′ RACE were proceeded to produce the full-length cDNA. An 1825 bp cDNA was isolated from human placenta, which encodes a putative protein of 279 amino acids. The protein contains a sua5-yciO-yrdC domain. Blast analysis against the human genome database of Genbank revealed that the gene contains five exons, and assigned the gene to human chromosome 1p34.2. A transcript about 2.5 kb is ubiquitously expressed in human tissues. The gene is highly conserved during evolution.

Keywords Human yrdC · RACE · Yeast two-hybrid · Expression

Introduction

The yrdC protein family consists of a series of highly conserved proteins that contain sua5-yicO-yrdC domain. The sua5-yicO-yrdC domain appeared either as one domain of a multiple domain protein or as a single domain protein. As the latter case, the sua5 gene from yeast, was identified as a suppressor of a translation initiation defect in cytochrome c and is required for normal growth in yeast. sua5 could act either at the transcriptional or the posttranscriptional levels to compensate for an aberrant translation start codon in the cyc gene. The sua5–1 allele enhances the iso-1-cytochrome c steady state level in the cyc1–1019 mutant from 2% to approximately 60% of normal (Cyc+) and also confers a marked slow growth phenotype. sua5 null mutants lack cytochrome a.a3 and fail to grow on lactate or glycerol medium. These results define that sua5 is necessary for normal cell growth (Na et al. 1992; Hampsey et al. 1991; Klima et al. 1996). The crystal structure of the YrdC protein, the sau 5 equivalent of *Escherichia coli,* reveals a large concave surface on one side that exhibits a positive electrostatic potential. The conserved basic amino acids located at its floor suggest that YrdC may be a nucleic acid binding protein. An investigation of YrdC's binding affinities for RNA and DNA fragments demonstrates that YrdC binds preferentially to double-stranded RNA (Teplova et al. 2000). However, the actual function of the Sua5 protein remains unknown.

In the previous study, we cloned a RBBP 10, a homolog of RBBP9, which binds to RB protein (Chen et al. 2002; Zhang et al. 1998). In order to reveal the proteins that interact with RBBP 10, the yeast two-hybridization was employed. Using RBBP10 as a bait protein, we screened the fetal brain library. An abundant positive clone that encodes a putative protein with an yrdC domain was isolated. The full-length cDNA was isolated by using both 5′ and 3′ RACE. Northern blot and RT-PCR revealed the transcript was expressed ubiquitously in multiple human tissues with different intensities.

Sequence data from this article has been deposited with the GenBank/EMBL Database with Accession No. AY172561.

J. Chen · C. Ji · S. Gu · E. Zhao · J. Dai · J. Qian
K. Ying · Y. Xie · Y. Mao (✉)
State Key Laboratory of Genetic Engineering,
School of Life Science, Fudan University,
200433, Shanghai, P.R. China
E-mail: ymmao@fudan.edu.cn
Tel.: +86-21-65642502
Fax: +86-21-65642502

L. Huang
United Gene Holding, 1111 Zhongshan BeiEr Rd., 200092,
Shanghai, P.R.China

Materials and methods

Yeast two-hybridization

The MATCHMAKER LexA two-hybrid system and human fetal brain MATCHMAKER Lex libraries were purchased from Clon-

tech. The ORF of RBBP 10 was cloned into plasmid pLexA multiple-clone sites between that of EcoR I and Hind III. The yeast two-hybridization was performed according to the recommendations of the manufacturer. The sequential transformation protocol was adopted. All the positive clones were classed by the length of PCR using the AD fusion sites specific primer (5′ccagcctcttgctgagtggagatg3′, 5′ggagacttgaccaaacctctggcg3′) and yeast clone hybridization. The independent clones were verified by the yeast mating test. The PCR products were sequenced on a PE3700 sequencer.

Isolation of the full length cDNA of human yrdC

By yeast two-hybridization, we isolated a 660 bp cDNA fragment that shares great similarity to Flj23476 (BC008984). No in-frame stop code was found in 5′ terminus. Based on the cDNA sequence, two primers (human yrdC 5′ gene-specific primer (gsp1), 5′atggaacgctcggaggagctcaac3′; human yrdC 3′ gene-specific primer (gsp2), 5′caggtaggacgcatgtgaggggag3′) were designed to perform both 3′ and 5′ RACE. The SMARTRACE kit was purchased from Clontech and the experiment was performed according to the manufacturer's recommendations. The human placenta total RNA (appendix of SMARTRACE kit) was used as template of RACE. Both the products of 5′ and 3′ RACE were cloned into T-vector and sequenced on a PE3700 sequencer.

Bioinformation analysis of human yrdC

The sequences from yeast two-hybrid and RACE were analyzed using blast program of NCBI (www.ncbi.bim.nih.gov). The protein conserved domain analysis and alignment was performed on Expasy (www.expasy.org).

Expression pattern analysis of human yrdC

Premade human multiple tissue Northern blots I (MTN I) was purchased from Clontech. The random primer label kit was purchased from Promega. The sequenced 660 bp PCR product of human yrdC was used as a probe. The probe was labeled with (α-32P)dATP and hybridized to MTN I . After autoradiography, the images were analyzed with the OptiQuant Image Analysis software (PACKARD).

Human multiple tissue cDNA panel (MTC) and Advantage 2 DNA polymerase were purchased from Clontech. The cycle-time limited PCR was used for examining the expression pattern with the same primers used in RACE. The PCR with glyceraldehyde-3-phosphate dehydrogenase (G3PDH) primers (appendix of MTC kit: 5′cggattggttgaaggtcggagtcaa3′; 5′catgtgggccatgaggtccaccac3′) served as a control. The PCR was performed according to the manufacturer's recommendations.

Results

With RBBP10 as a bait protein, we screened total 10^6 independent clone of a premade human fetal brain Lex active domain fusion library, and 248 positive clones were isolated. One of the most abundant positive clones contains a 660 bp cDNA that is very similar to a putative human cDNA, FLJ23476 (BC008984, corresponding to nt 349–1009). This cDNA fragment contains a ORF that encodes a polypeptide with 127 amino acids that contain a sua5-yciO-yrdC domain.

Using this 660 bp cDNA as an initial sequence, we screened the human dbEST. There are 53 high quality EST (E value $=0$) corresponding to the cDNA and resembling a putative cDNA of about 2 kb. However, there is not a 5′ in-frame stop codon yet.

Based on the 660 bp cDNA, we designed two primers that were used for RACE. We have isolated a DNA fragment about 600 bp in 5′ RACE and a DNA fragment about 1500 bp in 3′ RACE (Fig. 1). These two DNA fragments overlap each other and resemble an 1825 bp cDNA. Blast analysis against a human genome database revealed that the gene is located to human chromosome 1p34.2 and contains five exons (Fig. 2). All exon and intron boundaries conform to the AG-GT rule. The gene spans about 5.2 kb genome sequence.

The 1825 bp cDNA contains an 840 bp (nt 6–845) ORF that encodes a putative protein with 279 amino acid residues. Four polyadenylation signals are after the ORF (nt 1391, 1395, 1605, 1659). No 5′ in-frame stop codon is found corresponding to this ORF (Fig. 3).

The putative protein contains a sua5-yciO-yrdC domain (residues 86–253), a GTP-binding elongation factors signature (residues 229–241), and a leucine zipper pattern (residues 259–279) (Fig. 3). Blast analysis suggests that the protein is highly conserved from *E. coli* to humans. Human yrdC protein shares 90% identity to murine hypothetical protein XP_131673, 54% to *Drosophila melanogaster* yrdC protein NP_611502, 47% to an *Arabidopsis thaliana* unknown protein BAB09832, 37% to *Caenorhabditis elegans* hypothetical protein Y48C3A (NP_496826), 22% to *Saccharomyces cerevisiae* sua5 (X64319) and 26% to *E. coli* yrdC protein (NP_417741). These proteins are different in size, while the sua5-yciO-yrdC domain is the only conserved domain (Fig. 4).

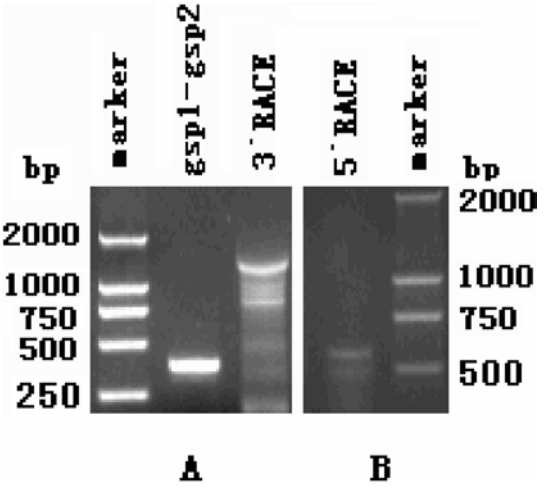

Fig. 1 The RACE results of yrdC. gsp1 (human yrdC 5′ gene-specific primer: 5′atggaacgctcggaggagctcaac3′) and gsp2 (human yrdC 3′ gene-specific primer: 5′caggtaggacgcatgtgaggggag3′) were used to amplify the fragment between nt462 and nt842. 3′ RACE used primer gsp1 and UPM (Clontech, Universal Primers Mixture). 5′ RACE used primer gsp2 and UPM

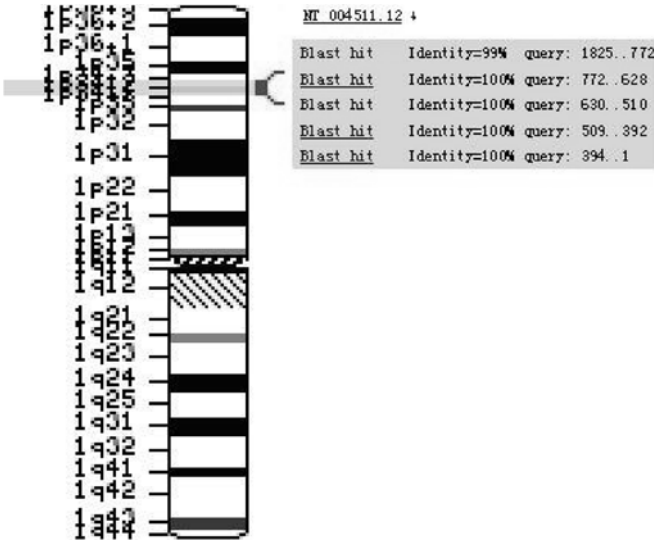

Fig. 2 The chromosome location of the human yrdC gene. Human yrdC was located to human chromosome 1p34.2. The gene contains five exons

Northern blot analysis detected a single band about 2.5 kb hybridized to all blots with different intensities. The transcript is high in liver and pancreas (Fig. 5). MTC based PCR verified the expression pattern revealed by Northern blot analysis. The PCR bands presented here appeared in 36 cycles of PCR, while the G3PD control group reached saturation at 24 cycles. The transcript is expressed ubiquitously in human tissues. The transcript is high in tumor tissues (Fig. 6).

Discussion

In the course of screening the fetal brain LexA active domain library with RBBP 10 as bait protein for RBBP interaction protein, we isolated a 660 bp cDNA. The cDNA encodes a 220-residue polypeptide that contains the sequence of a putative protein predicted by NCBI, FLJ23476. This protein contains a sua5-yciO-yrdC domain; it is named human yrdC protein.

Fig. 3 The cDNA nucleotide sequence and deduced amino-acid sequence of human yrdC. The *number* at the right of each line indicates the position of the first nucleotide. The open-reading frame (ORF) is *uppercased*. The first atg is in *italic type*. The polyadenylation signals are *shaded*. The yrdC domain is in *bold type*. The GTP-binding elongation factors signature is *underlined*. The leucine zipper pattern is *boxed*

```
ggcgg ATG TCT CCG GCG CGT CGG TGC AGG GGG ATG AGG GCC GCG GTG GCT GCC AGC GTG GGG TTG AGC GAG GGG    74
      M  S  P  A  R  R  C  R  G  M  R  A  A  V  A  A  S  V  G  L  S  E  G                          23
CCT GCT GGC TCC CGG AGC GGT CGC CTC TTC CGC CCG CCG AGT CCC GCT CCG GCG GCC CCC GGC GCC CGG CTG    146
  P  A  G  S  R  S  G  R  L  F  R  P  P  S  P  A  P  A  A  P  G  A  R  L                          47
TTG CGG CTC CCG GGG AGC GGG GCC GTG CAG GCC GCG AGC CCG GAG CGC GCC GGC TGG ACC GAG GCG CTG CGG    218
  L  R  L  P  G  S  G  A  V  Q  A  A  S  P  E  R  A  G  W  T  E  A  L  R                          71
GCC GCC GTG GCC GAG CTG CGC GCC GGC GCC GTG GTG GCC GTC CCC ACC GAT ACG CTG TAC GGC CTG GCC TGC    290
  A  A  V  A  E  L  R  A  G  A  V  V  A  V  P  T  D  T  L  Y  G  L  A  C                          95
GCG GCG AGC TGC TCG GCG GCT CTG CGC GCT GTG TAC CGC CTC AAG GGT CGC AGC GAG GCC AAG CCT CTG GCC    362
  A  A  S  C  S  A  A  L  R  A  V  Y  R  L  K  G  R  S  E  A  K  P  L  A                          119
GTA TGC CTC GGC CGC GTG GCC GAC GTC TAC AGA TAC TGC CGT GTG AGA GTA CCT GAG GGG CTC CTG AAA GAC    431
  V  C  L  G  R  V  A  D  V  Y  R  Y  C  R  V  R  V  P  E  G  L  L  K  D                          143
CTA CTG CCA GGA CCA GTG ACC CTG GTG ATG GAA CGC TCG GAG GAG CTC AAC AAG GAC CTA AAC CCT TTT ACG    503
  L  L  P  G  P  V  T  L  V  M  E  R  S  E  E  L  N  K  D  L  N  P  F  T                          167
CCT CTT GTA GGC ATT CGG ATT CCT GAT CAT GCT TTT ATG CAA GAC TTG GCT CAG ATG TTT GAG GGT CCG CTT    575
  P  L  V  G  I  R  I  P  D  H  A  F  M  Q  D  L  A  Q  M  F  E  G  P  L                          191
GCT CTC ACT AGT GCC AAC CTC AGC TCC CAG GCC AGT TCT CTG AAT GTC GAG GAG TTC CAG GAT CTC TGG  CCT    647
  A  L  T  S  A  N  L  S  S  Q  A  S  S  L  N  V  E  E  F  Q  D  L  W  P                          215
CAG TTG TCC TTG GTT ATT GAT GGG GGA CAA ATT GGG GAT GGC CAG AGC CCC GAG TGT CGC CTT GGC TCA ACT    719
  Q  L  S  L  V  I  D  G  G  Q  I  G  D  G  Q  S  P  E  C  R  L  G  S  T                          239
GTG GTT GAT TTG TCT GTG CCC GGA AAG TTT GGC ATC ATT CGT CCA GGC TGT GCC CTG GAA AGT ACT ACA GCC    791
  V  V  D  L  S  V  P  G  K  F  G  I  I  R  P  G  C  A  L  E  S  T  T  A                          263
ATC CTC CAA CAG AAG TAC GGA CTG CTC CCC TCA CAT GCG TCC TAC CTG TGA aactctgggaagcaggaaggccc          868
  I  L  Q  Q  K  Y  G  L  L  P  S  H  A  S  Y  L  &                                               279
aagacctggtgctggatactatgtgtctgtccactgacgactgtcaaggcctcatttgcagaggccaccggagctagggcactagcctgact      960
tttaaggcagtgtgtctttctgagcactgtagaccaagcccttggagctgctggtttagccttgcacctggggaaaggatgtatttatttgt     1052
attttcatatatcagccaaaagctgaatggaaaagttaagaacattcctaggtggccttattctaataagtttcttctgtctgttttgtttt     1144
tcaattgaaaagtaatttaaataacagatttagaatctagtgagagcctcctctctggtgggtggtggcatttaaggttcaaaccagccagaa     1236
gtgctggtgctgtttaaaaagtctcaggtggctgcgtgtggtggctcatgcctgtaatcccaacattctgggaggcccaggcgggagaactg     1328
cttgagcccaggagttcagaatcagcctgggcaacatagcaatactccgtctcataaaaattaataaataaaaagtctcaggtgaccaaagg     1420
ctcctgaagctagaaccaggtttggataaagattgaagagccacaggccactcttccctctgagccattgggcctagtggtgtcatgtattg     1512
taattgctcgcagggagagcagtctttttggtgtaatagtgggatgtctgcttagttggcaggggttcagtccaaatggaagaatattggga     1604
aataaacctccactatccttttatagccagggacttttttcttatttattcataaaataaattatagttaattatacccataacacctttatt     1696
taaatccagtgttctccgcagccttttgtctatttatatgtgtaccaagtgttaaacataattattattgggcatttgaactttgtttttct     1789
ttaaagaaatgctgctattaaacatatttgtaaatgg                                                             1825
```

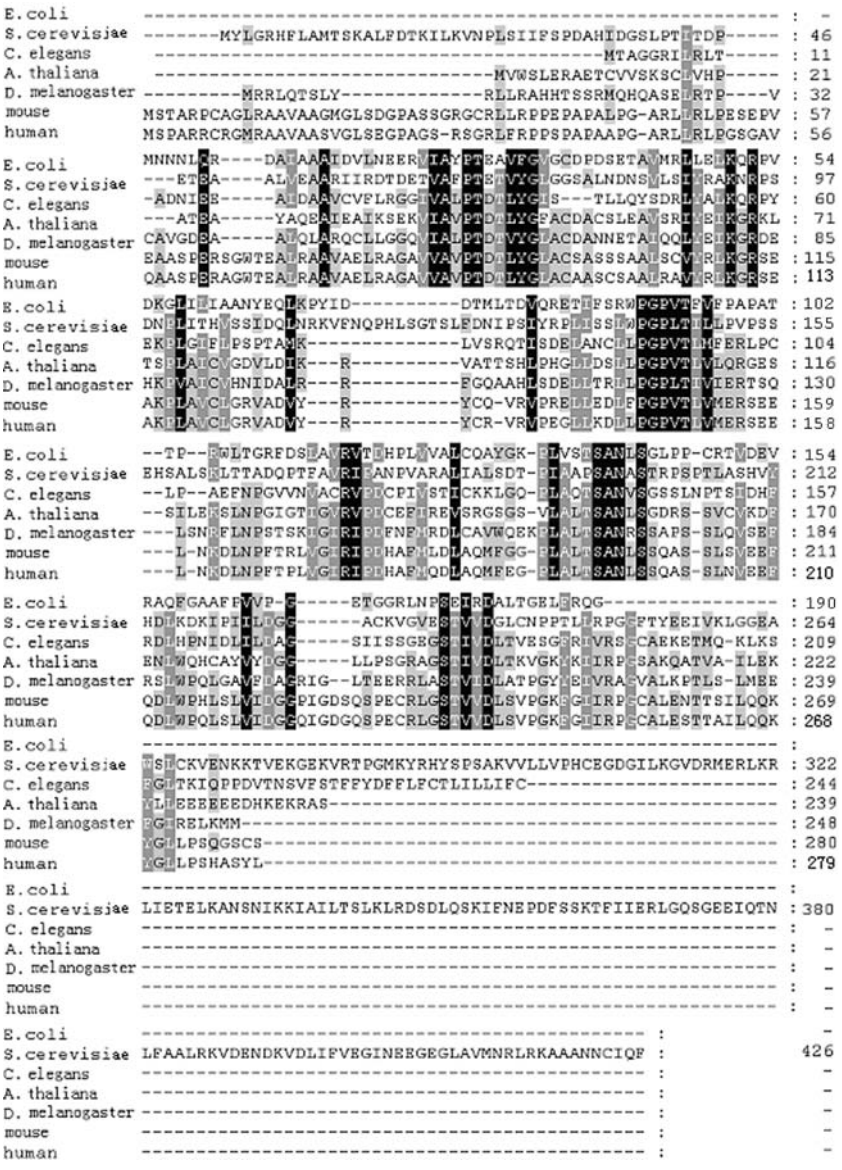

Fig. 4 Protein sequence alignment of human yrdC and its homolog of *E. coli*, *S. cerevisiae*, *C. elegans*, *A. thaliana*, *D. melanogaster* and mouse. The *numbers* indicate the positions of amino acids. The conserved residues are marked with different *shadows*

Using the 660 bp cDNA as an initial sequence, blast analysis against human dbEST found a cluster of 53 high quality ESTs that resembled an 1825 bp putative cDNA. Both 5′ and 3′ RACE verified this cDNA sequence. Human yrdC gene is located to chromosome 1p34.2. It spans 5.2 kb genome that contains five exons and four introns. All exon–intron boundaries conform to the AG_GT rule. Northern blot analysis shows the transcript is about 2.5 kb. Based on the above data, the full-length cDNA of human yrdC has been cloned.

The 1825 bp cDNA contains an 837 bp ORF that encodes a 279-residue protein. Four polyadenylation signals followed the stop codon. Although there is not a 5′ in-frame stop codon found in the cDNA, we are sure the ORF is completed.

Blast analysis revealed the yrdC protein is highly conserved from *E. coli* to human. Human yrdC protein shares 90% identity to murine hypothetical protein XP_131673, 54% to *D. melanogaster* yrdC protein NP_611502, 47% to an *A. thaliana* unknown protein BAB09832, 37% to *C. elegans* hypothetical protein Y48C3A (NP_496826), 22% to *S. cerevisiae* sua5 (X64319) and 26% to *E. coli* yrdC protein (NP_417741). However, these proteins are different in size, the only conserved part is yrdC domain. There is a sua5 domain found in yeast protein sua5 protein, it is different from other homologs.

Human yrdC protein contains a sua5-yciO-yrdC domain, a GTP-binding elongation factors signature, and a leucine zipper pattern. YrdC protein family is highly conserved during the course of evolution, but its actual function is not clear yet.

GTP-binding elongation factors signature is found in elongation factors that catalyze the elongation of peptide chains in protein biosynthesis. This region is conserved in both EF-1alpha/EF-Tu as well as EF-2/EF-G

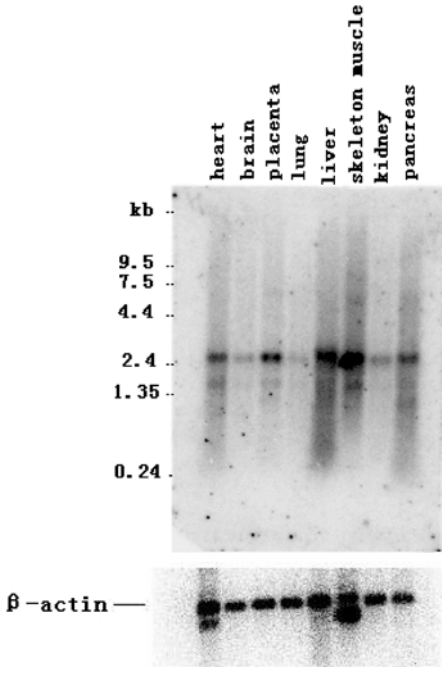

Fig. 5 Northern blot analysis of human yrdC. The RNA size marker positions are indicated on the *right*. A transcript with length of 2.5 kb was expressed in all detected tissues with different intensities. The hybridization signal of beta actin served as a control

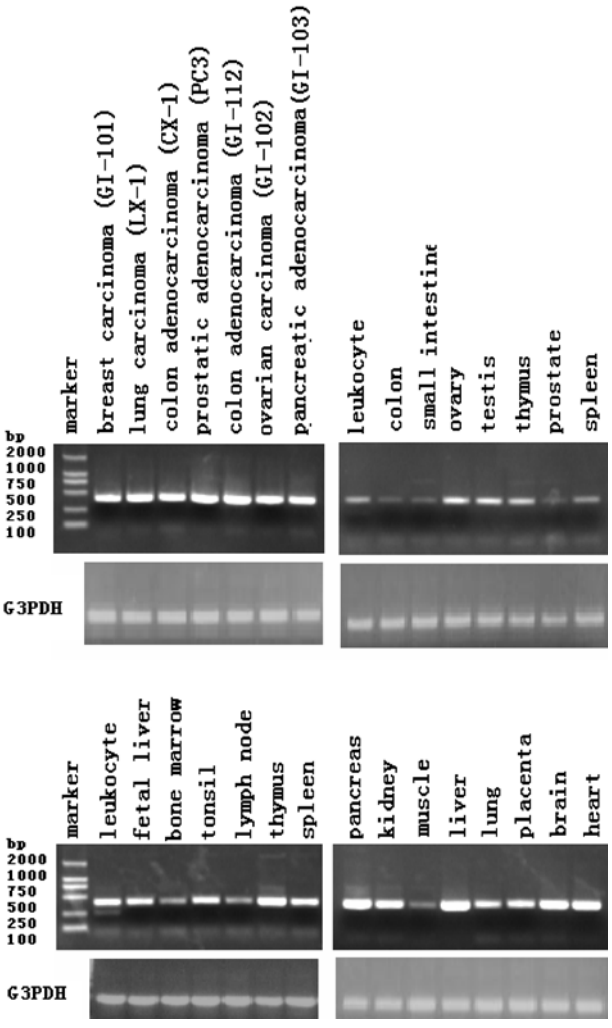

Fig. 6 Expression pattern of human yrdC revealed by RT-PCR. The adult MTC I, MTC II, MTC panels of the immune system, and human tumor MTC panels were detected by RT-PCR with the primers of human yrdC (gsp1 and gsp2). The PCR was performed according to the manufacturer's instruction for 36 cycles (94°C 30 s, 68°C 2 min, 36 cycles). The G3PDH primers mixture (Clontech) was used for a control PCR (24 cycles, 94°C 30 s, 68°C 2 min)

and thus seems typical for GTP-dependent proteins which bind non-initiator tRNAs to the ribosome (Moller et al. 1987; Moldave 1985). The GTP-binding elongation factor family also includes several proteins related to cell growth, such as the yeast omnipotent nonsense codons suppressor protein SUP2 and rat statin S1 (Lee et al. 1992; Hoshino et al. 1989).

Leucine zipper pattern is a conserved motif related to protein dimerization and nucleic acid binding. It is present in some transcript factors (Zahnow 2002; Busch and Sassone-Corsi 1990). However, the consensus sequence is short and it is too unspecific. Because the homolog of human yrdC, *E. coli* yrdC, was a verified nucleic acid binding protein, this motif might be critical for the function of human yrdC.

Both Northern blot analysis and MTC based RT-PCR revealed that human yrdC was expressed ubiquitously in human tissues. The transcript is high in liver, pancreas, brain and some tumor tissue, where protein synthesis is on a relatively high level. The result suggests that human yrdC might be a house-keeping gene and might relate to protein synthesis.

Based on the structure character and the homolog's function of human yrdC protein, we deduce that human yrdC might be a translation related protein. Its interaction with RBBP10 might be unspecific. The actual function needs further study.

Acknowledgements We thank Professor Nolan for his critical reading of the manuscript. We gratefully acknowledge the financial support from the National Nature Science Foundation of the People's Republic of China (grant number: 30270519).

References

Busch SJ, Sassone-Corsi P (1990) Dimers, leucine zippers and DNA-binding domains. Trends Genet 6:36–40

Chen JZ, Yang QS, Wang S, Meng XF, Ying K, Xie Y, Mao YM (2002) Cloning and expression of a novel retinoblastoma binding protein cDNA, RBBP10. Biochem Genet 40:273–282

Hampsey M, Na JG, Pinto I, Ware DE, Berroteran RW (1991) Extragenic suppressors of a translation initiation defect in the cyc1 gene of *Saccharomyces cerevisiae*. Biochimie 73:1445–1455

Hoshino S, Miyazawa H, Enomoto T, Hanaoka F, Kikuchi Y, Kikuchi A, Ui M (1989) A human homologue of the yeast GST1 gene codes for a GTP-binding protein and is expressed in a proliferation-dependent manner in mammalian cells EMBO J 8:3807–3814

Klima R, Coglievina M, Zaccaria P, Bertani I, Bruschi CV (1996) A putative helicase, the SUA5, PMR1, tRNALys1 genes and four open reading frames have been detected in the DNA sequence of an 8.8 kb fragment of the left arm of chromosome VII of *Saccharomyces cerevisiae* yeast 12:1033–1040

Lee S, Francoeur AM, Liu S, Wang E (1992) Tissue-specific expression in mammalian brain, heart, and muscle of S1, a member of the elongation factor-1 alpha gene family. J Biol Chem 267:24064–24068

Moldave K, (1985) Eukaryotic protein synthesis. Annu Rev Biochem 54:1109–1149

Moller W, Schipper A, Amons R (1987) A conserved amino acid sequence around Arg-68 of Artemia elongation factor 1 alpha is involved in the binding of guanine nucleotides and aminoacyl transfer RNAs. Biochimie 69:983–989

Na JG, Pinto I, Hampsey M (1992) Isolation and characterization of SUA5, a novel gene required for normal growth in *Saccharomyces cerevisiae*. Genetics 131:791–801

Teplova M, Tereshko V, Sanishvili R, Joachimiak A, Bushueva T, Anderson WF, Egli M (2000) The structure of the yrdC gene product from *Escherichia coli* reveals a new fold and suggests a role in RNA binding. Protein Sci 12:2557–2566

Zahnow CA (2002) CCAAT/enhancer binding proteins in normal mammary development and breast cancer. Breast Cancer Res 4:113–121

Zhang M, Niu CH, Thorgeirsson SS (1998) A retinoblastoma-binding protein that affects cell-cycle control and confers transforming ability. Nat Genet 19:371–374