## 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 vrdC domain. It is named human vrdC 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  $\cdot$  RACE  $\cdot$  Yeast two-hybrid  $\cdot$  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

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

J. Chen  $\cdot$  C. Ji  $\cdot$  S. Gu  $\cdot$  E. Zhao  $\cdot$  J. Dai  $\cdot$  J. Qian K. Ying  $\cdot$  Y. Xie  $\cdot$  Y. Mao ( $\boxtimes$ ) 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 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.

#### 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'ccagectettgetgagtagatg3', 5'ggagacttgaccaaacctetggeg3') 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 ( $\alpha$ -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-3phosphate dehydrogenase (G3PDH) primers (appendix of MTC kit: 5'cggattggttgaaggtcgagtcaa3'; 5'catgtgggccatgaggtccacca3') 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<sup>6</sup> 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' genespecific primer: 5'atggaacgctcggaggagctcaac3') and gsp2 (human yrdC 3' gene-specific primer: 5'caggtaggacgcatgtgagggggag3') 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

Fig. 3 The cDNA nucleotide sequence and deduced aminoacid sequence of human yrdC. The *number* at the right of each line indicates the position of the first nucleotide. The openreading 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*  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.

00-0	gg A	<i>TG</i> TC	CT CC	CG GG	CG CC	GT CO	GG TO	GC AG	GG GG	G AT	IG AG	GG GC	CC GC	CG G1	G GG	CT GO	C AC	GC G1	G GG	G T1	rg ac	GC GA	G GGG	74
	М	S	Р	A	R	R	С	R	G	M	R	A	Α	v	Α	Α	S	v	G	L	S	Е	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
Р	Α	G	S	R	S	G	R	L	F	R	Р	Р	s	Р	Α	Р	А	А	Р	G	Α	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	Р	G	S	G	A	v	Q	А	А	S	Р	Е	R	Α	G	w	Т	Е	Α	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	Α	v	Α	Е	L	R	A	G	Α	v	v	Α	v	Р	Т	D	Т	L	Y	G	L	A	С	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	С	S	A	A	L	R	A	V	Y	R	L	K	G	R	S	E	A	K	Р	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	С	L	G	R	v	A	D	v	Y	R	Y	С	R	v	R	v	Р	Е	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	Р	G	Р	v	Т	L	v	М	Е	R	s	Е	Е	L	N	K	D	L	N	Р	F	Т	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
Р	L	v	G	I	R	I	Р	D	Н	A	F	М	Q	D	L	A	Q	M	F	Е	G	Р	L	191
GCT	СТС	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	Ε	E	F	Q	D	L	W	Р	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	Р	E	С	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
<u>v</u>	v	D	L	S	v	Р	G	К	F	c	-								P	C	т		4	263
ATC	CTC								•	G	I	I	R	Р	G	С	A	L	Ľ	5	1	T	A	
	CIU	CAA	CAG	AAG	TAC	GGA	CTG	СТС	ccc	G TCA	I CAT	I GCG	R TCC	P TAC	G CTG	C TGA	A aact	L	gaag	5 gcagg	ı gaagg	T	Α	868
I	L	CAA Q	CAG Q	AAG K	TAC Y	GGA G	CTG L	CTC L	CCC P	TCA S	I CAT H	I GCG A	R TCC S	P TAC Y	G CTG L	C TGA &	A aact	L	gaag	5 gcagg	ı gaagş	T gccc	A	868 279
I aaga	L	CAA Q ggtgd	CAG Q tgga	AAG K atact	TAC Y tatg	GGA G	CTG L tgtco	CTC L cacts	CCC P gacga	TCA S	I CAT H caag	I GCG A ggcct	R TCC S tcatt	P TAC Y	G CTG L	C TGA & gccad	A aact	L	gaag gaag	s gcagg cacta	I gaagg	T gccc	A	868 279 960
I aaga ttta	L L accta	CAA Q ggtgc cagtg	Q Q tgga	AAG K ataci	TAC Y tatgi	GGA G tgtct gcact	CTG L tgtco tgtag	CTC L cactg	CCC P gacga	G TCA S actgt	I CAT H caag ggago	I GCG A ggcct	R TCC S tcatt	P TAC Y ttgca	G CTG L agagg cttg	C TGA & gccac gcaco	A aact ccgga ctggg	L ctgg ngcta	ggaag aggagg aggat	s gcagg cacta	agaagg agcct	T gccc tgact	A	868 279 960 1052
I aaga ttta atti	L acctg aggo	Q ggtgd cagtg tatat	Q Q ctgga gtgto ccago	AAG K atact cttto ccaaa	TAC Y tatgi ctgag	GGA G tgtct gcact	CTG L tgtco tgtag	CTC L cacts gacca	CCC P gacga tagco	TCA S actgt ccttg	I CAT H tcaag ggago	I GCG A ggcct ctgct	R TCC S catt tggtt	P TAC Y ttgca ttagc	G CTG L agagg ccttg attct	C TGA & gccac gcaco taata	A aact ccgga ctgga aagtt	L ctgg agcta ggaaa	ggaag agggg aggat totgt	s gcagg cacta tgtat tctgt	agcct tttat	T gccc tgact tttgt gtttt	A	868 279 960 1052 1144
I aaga ttta atti tcaa	L accts aggo ttcat	CAA Q ggtgc cagtg tatat	Q Q ctgga gtgto ccago gtaat	AAG K atact cttto ccaaa	TAC Y tatg ctgag aagct ataac	GGA G tgtct gcact tgaat	CTG L tgtco tgtag tggaa tttag	CTC L cacts gacca aaagt gaato	CCC P gacga nagco taag	TCA S actgt ccttg gaaca	I CAT H ccaag ggago attco agcct	I GCG A ggcct ctgct ctagg	R TCC S tcatt tggtt gtggc	P TAC Y ttgca ttagc ctta ggtgg	G CTG L agagg ccttg attct ggtgg	C TGA & gccac gcacc taata gtggo	A aact ccgga ctgga agtt cattt	L ageta ggaaa ttett	ggaag agggo aggat tetgt	s gcagg tgtat tctgt	agcot tttat ttttg	T gccc tgact tttgt gtttt cagaa	A 2	868 279 960 1052 1144 1236
I aaga ttta atti tcaa gtgo	L accts aaggo ttcat attga	CAA Q ggtgc cagtg tatat aaaaag tgctg	Q ctgga stgtc ccago staat sttta	AAG K atact cttto ccaaa ttaaa	TAC Y tatgi taga aagci ataac	GGA G tgtct gcact tgaat cagat	CTG L tgtco tgtag tggaa tttag	CTC L cacts gacca aaagt gaato	CCC P gacga iagco itaag itaag	TCA S ictgt ictgt icttg gaaca igaga ggtgg	I CAT H ccaag ggagc attco agcct gctca	I GCG A ggcct ctgct ctagg tcctc	R TCC S tcatt tggtt tggt tggt tctg	P TAC Y ttgca ttago cctta ggtgg aatco	G CTG L agagg ccttg attct ggtgg ccaao	C TGA & gccac gcacc taata gtggc catto	A aact ccgga ctggg aagtt cattt	L ageta ggaaa ttett aagg	ggaag agggo aggat tetgt gttca cccag	s gcagg cacta tgtat tctgt naaco ggcgg	agcct tttat tttt cagco ggaga	T gccc tgact tttgt gtttt cagaa hactg	A ; ; ;	868 279 960 1052 1144 1236 1328
I aaga ttta atti tcaa gtgo ctta	L accts aaggo ttcat attga ctggt	CAA Q ggtgc cagtg tatat aaaag tgctg ccagg	Q ctgga stgto ccago staat staat staat	AAG K atact cttto ccaaa ttaaa aaaaa tcaga	TAC Y tatg aagct ataac agtc1 aatca	GGA G tgtct tgaat tgaat tagat	CTG L tgtco tgtag tggaa tttag gtggo	CTC L cacts gacca aaagt gaato ctgcs caaca	CCC P gacga tagco taag tagt gtgtg	TCA S ictgt ccttg gaaca gaga gaga gtgg caata	I CAT H ccaag ggago attco agcct gctca actco	I GCG A ggcct ctgct ctagg tcctc atgcc atgcc	R TCC S catt tggtt tggt tggt tgtg ctctg ctgta	P TAC Y ttgca ttago cctta ggtgg aatco aaaaa	G CTG L agaga cctta attci ggtga ccaac	C TGA & gccac gcacc taata gtggo catto ataaa	A aact ccgga ctgga agtt cattt ctgga ataaa	L ctgg agcta ggaaa ttctt aagg aggo aagt	ggaag aggga aggat totgi totgi gttca cocag	s gcagg tgtat tctgt aaaco ggcgg aggtg	agcc1 tttat ttttat cagco ggaga gacca	T gccc tgact tttgt gtttt cagaa hactg		868 279 960 1052 1144 1236 1328 1420
I aaga ttta atti tcaa gtgo ctta	L accts aaggo ttcat attga attga gagco	CAA Q ggtgc cagtg tatat aaaaag tgctg ccagg agcta	Q etgga gtgtc ccago gtaat gttta gagtt agaao	AAG K atact ccaaa tcaga ccagg	TAC Y tatge aaget ataac agtet aatca	GGA G tgtct tgaat tcaga tcaga agcct ggata	CTG L tgtco tgtag tggaa tttag gtggo tgggo aaaga	CTC L cacts gacca aaagu gaato ctgcs caaca attga	CCC P gaoga taagoo taag tagt gtgtg tagg	TCA S actgt gaaca gaga gaga gaga gatgg caata	I CAT H dutted ageet agetea acted caggo	I GCG A ggcctt ctgct ctagg cctc itgcc cgtct	R TCC S ccatt sggtt sggtt stggc ctcts ccata	P TAC Y ttgca ttgca cctta ggtgg aatco aaaaaa cccto	G CTG L agagaga ccttg ggtgg ccaac attca attca	C TGA & gccac gcacc gcacc gcact catto htaaa gccat	A aact ccgga aagtt cattt cttggg itaaa	L cctgg ggaaa ttctt aagg gaggo aagt ggccta	E ggaag aggat aggat aggat aggat agttca agtgg	s cacta tgtat tctgt aaaacc ggcgg aggtg gtgtc	gaagg agcct tttat ttttg ggaga gacca catg1	T gccc tgact tttgt gtttt cagaa aactg aaagg		868 279 960 1052 1144 1236 1328 1420 1512
I aaga ttta atti tcaa gtgo ctta taa	L accts aaggo ttcat attga ctggt gagco ctgaa	CAA Q ggtgc cagtg tatatat tatat tgctg agcta tcgca	Q Q ttgga gtgtc ccago gtaat gttta gagti agaao	AAG K attact ctttto ccaaa aaaaaa tcaaa ccaag agago	TAC Y tatg aaagc ataac agtc1 gtttg cagto	GGA G tgtct gcact tgaat cagat tcagg agcct ggata	CTG L tgtcd tgtaa tggaa tttag gtggo tgggo aaaaga ttgg	CTC L cacts gacca aaaagt gaato ctgca cacaca attga tgtaa	CCC P gacga aagco ttaag ttagt ttagt aagag aagag	TCA S actgt actgt gaaca gaga gaga gaga gaga g	I CAT H ccaag ggagco attco caggo atgcto actoo caggo	I GCG A ggcct ctgct ctagg cctc atgcc cgtct ccact	R TCC S ccatt tggtt ttggc ttcts ccata ccata	P TAC Y ttgca ttagc cctta ggtgg aaacc ccctc ttggc	G CTG L agagg cettg atter ggtgg ccaac attaa ccaac ccaac	C TGA & gccace gcace gcace catte gcatte gccat gccat	A aact ccgga ctggg aagtt cattt ctggg ataaa ctggg	L cctgg ggaaa ttctt aagg ga ga	ggaag aggaag aggat cctgi cccag ccccag ccccag agtgg	s gcagg tgtat tctgt aaacc ggcgg ggcgg ggggg ggggg ggggg ggggg ggggg ggggg gggg	gaagg agcct tttat ttttg cagco ggaga gacca catg1 atat1	T gccc tgact tttgt gtttt cagaa aactg aaagg cattg	A	868 279 960 1052 1144 1236 1328 1420 1512 1604
I aaga ttta atti tcaa gtgo ctta taa aata	L accts aaggo ttcat attga gagco ctgaa ttgct aaacc	CAA Q ggtgc cagtg tatat tatat aaaaag tgctg ccagg agcta agcta cccagg ccagg	Q Q ttgga gtgtgtc ccago gtaat gttta gagtt agaad aggga actat	AAG K atact ccttta aaaaaa tcaga ccaga agago tcctt	TAC Y tatgi tatga aagct aataad agtci aatca gttig gttig cagto ttata	GGA G tgtct tgaat tgaat tcagg	CTG L tgtcc tgtag tggac ttgg aaagg ttggg agggg	CTC L cactg gacca gacca gacca caaca ctgca caaca attga actti	CCC P gacga gac gac	TCA S actgt actgt gaaca gaga gaga gaga gaga g	I CAT H ggagc attco agoot agoot caggo attco caggo attgo	I GCG A ggcct ctgct ctgct atgcc ccctc atgcc ccctc atgcc tccact	R TCC S ccatt cggtt dtggo ctctg ccata ccata ccata	P TAC Y ttgca ttagc cctta ggtgg aatco aaaaaa cccto ttggc	G CTG L agagg ccttg ccttg ggtgg ccaac attaa ctgag caggg	C TGA & gccac gcacc taata gccac gcatto ataaa gccat ggtto tagtt	A aact ccgga aagti catti ctggg ataaa ctggg cagto cagto	L cctgg agcta ggaaa ttct1 caagg gaggo tccta gccta ccaaa	ggaag aggga aggga aggga aggga aggga agtto agtto agtto agtgg agtgg agtgga agtgga aggga aggaag	s gcagg tgtat tctgt aaacc ggcgg gggtg gtgtc aaggaa	agcc1 ttta1 ttttg cagcc ggaga catg1 tta1	T gccc tgact tttgt gtttt cagaa aactg aaagg caggga ttatt	A	868 279 960 1052 1144 1236 1328 1420 1512 1604 1696
I aaga ttta atti tcaa gtgo ctta ctta taa aata taaa	L accta aaggo ttcat attga ctggt ctggt ttgct aaacco atcca	CAA Q ggtgc cagtg tatat tatat tatat tatat tgctg agcta tcgca	CAG Q ctgga gtgto ccago gtaat gttta ggga aggga actat ctcto	AAG K atact atact ccaaa ttaaa aaaaaa tcaag agago tcctt	TAC Y tatgi tatgi aaagci aaagci aataac gtttg cagto ttata agcci	GGA G tgtct tgaat caga tcaga ggata ctttt agcca tttt	CTG L tgtcc tgtag tgga tttag gtgga ttgg aaagg gtcta	CTC L cactg gacca aaagt gaato ctgcg caaca attga acttt attta	CCC P gacga hagco ttaag ttagg ttagg htago htagg	TCA S actgt ccttg gaaca gaga gaga gaga gaga g	I CAT H ccaag ggagc aggagc aggagc aggagc aggagc aggagc aggagc aggagc aggagc aggagc aggagc aggagc t ttat tta	I GCG A ggcct etgct etgct atgcc cctc atgcc cctc catgct tctagt	R TCC S ccatt sggtt sggtt stggc stcts sccata ccata ccata ccata scctto ttagt	P TAC Y ttgca ttgca cctta ggtgg aatco aaaaa cccto ttggo ttaaa	G CTG L aggagg ccttg gtgg gtgg ccaac attaa ccgggg ttaa ttaa	C TGA & gccac gcac taata gtggc catto ataaa gcat catto ataaa gccat catto	A aact ccgga ctggg aagtt cattt ctggg cagtc cagtc cagtc cagtc	L cctgg agcta ggaaa ttct1 caagg gaggo ccta aagt gccta ccaaa ccaaa catao	E ggaag aggat aggat aggat aggat agta agta agta agta agta agta agta agta agta agta agta agta agta aggaag	s cacta tgtat tctgt aaaacc ggcgg ggtgt aaggtg taaca	agcc1 tttat ttttg cagcc ggaga gacca catg1 atat1 acct1	T gccc gccc tttgt gtttt cagaa aactg aaagg cattg cggga ttatt		868 279 960 1052 1144 1236 1328 1420 1512 1604 1696 1789

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* 

n			
E.Coli S.Cerevisjae C.elegans A.thaliana D.melanogaster mouse human	MYLGRHFLAMTSKALFDTKILKVNPLSIIFSPDAHTDGSLPT TDP		46 11 21 32 57 56
E.coli S.cerevisjae C.elegans A.thaliana D.melanogaster mouse human	MNNNLRDA'AARIDVLNEER IAYETEAVEC GCADDESTA MELEINOSPU ETSAAL'EARRIRDTDETVAPETE VYGIGGSALNDSILSE RARNSPS -ADNIESALDAVCVFLRGGVALETDILYGISTLLQYSDRJAINOSPY ATSAYAQETEAINSEK IAYETDILYGFACDACSLEA SRL EINOSPY EANSPERSGVEAL RARVAELRAGA VAVETDILYGFACDANNETA IQUE EINOSE EAASFERSGVEAL RARVAELRAGA VAVETDILYGFACSASSSAL SOY RINGSE CAASFERAGVEALRARVAELRAGA VAVETDILYGFACSASSSAL SOY RINGSE		54 97 60 71 85 115
B.coli S.cerevisiae C.elegans A.thaliana D.melanogaster mouse human	DKGLI IAANYEQLKPYIDDTMLTDVQRETIFSR#PGPYTeVFPAPAT DN:LTTH/SSIDQLNRKVFNQPHLSGTSLEDNIPSTYRELISSI#PGPTTLLEVPSS EKELGIFIPSPTAMK	: 1 : 1 : 1 : 1 : 1 : 1	102 155 104 116 130 159 158
E.coli S.cerevisjae C.elegans A.thaliana D.melanogaster mouse human	TPFWLTGRFDSLATENT HPL/VALCOAYGK-FDVSDSANLSGLPP-CRT DEV EHSALSELTTADOPTFATRIFANPVARATIALSDT-FLAPSANASTRPSPTLASHV LPAEPNPGVNIACRVPTCPISTICKKLGO-FLAOTSANVSGSSLNPTS DH SILEISLNPGIGTIG RVTCEPTREVSRGSGS-VWALDSANLSGDRS-SVC KD LSNFLNPSTSKIG RIPTFNPIRDLCAVWQEKPLALDSANRSSAPS-SLOVSE L-NDLNPFTRL GERIPTHAFILDLAQMFGG-FLALDSANLSGAS-SLSVEE L-NDLNPFTPL GERIPHAFILDLAQMFGG-FLALDSANLSGAS-SLSVEE		154 212 157 170 184 211 210
E.coli S.cerevisiae C.elegans A.thaliana D.melanogaster mouse human	RAQFGAAF EVVP-GETGGRLN PEEIRCALTGEL ROG- HDIKDKI FIDIGGACKYGVES VIDGLCN PTLIREG PTYESIVKLGGEA RDIHENIDLILAGSIISGEGS IV DLVESG RIVRS CAEKETWQ-KLKS ENIMORCAYWIGGLIPSGRAGS IV DLVEKVGK, KIIRP SAKQATVA-ILEK RSIMPQLGAVFIAFRIG-LTEERRLAS VIDLATPGY EIVRA VALKPILS-LMEE QDIMPHLSLVIIGEPIGDSQSPECRLGS VIDLSVPGK GIIRP CALENTSILQQK QDIMPLSLVIIGEQIGDGQSPECRLGS VIDLSVPGK GIIRP CALESTAILQQK		190 264 209 222 239 269 268
E.Coll S.Cerevisiae C.elegans A.thaliana D.melanogaster mouse human	SICKVENKKTVEKGEKVRTPGMKYRHYSPSAKVVLLVPHCEGDGILKGVDRMERLKR GITKIQPPDVTNSVFSTFFYDFFLFCTLILLIFC- LIEBEEEDHKEKRAS- GRELKMM GILPSGGSCS- IGILPSHASYL-		322 244 239 248 280 279
E.coli S.cerevisjae C.elegans A.thaliana D.melanogaster	LIETELKANSNIKKIAILTSLKLRDSDLQSKIFNEPDFSSKTFIIERLGQSGEEIQTN	:3:	380
human		:	-
E.coli S.cerevisiae C.elegans A.thaliana	LFAALRKVDENDKVDLIFVEGINEEGEGLAVMNRLRKAAANNCIQF	4	- 126 - -
D. melanogaster	:		-
mouse human			-

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

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



**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)

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 cycl 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 retinoblastomabinding protein that affects cell-cycle control and confers transforming ability. Nat Genet 19:371–374