Genome-scale analysis of human mRNA 5′ coding sequences based on expressed sequence tags (EST) database

RT-PCR details

Standard reverse transcription conditions were: 2 μg of total RNA, Moloney murine leukemia virus reverse-transcriptase (Promega, Madison, WI; used with the companion buffer) 400 U, oligo dT-15 2.5 μM, random nonamers 2 μM, dNTPs 500 μM each. An RNA denaturation step was performed at 95°C for 5 min before the addition of primers and enzyme. RT reaction was performed in a final volume of 50 μL for 60 min at 42°C.

PCR experiments were performed in a 25 μL final volume, containing 2 μL of cDNA, 1 U Taq polymerase (TaKaRa, Shiga, Japan) with companion reagents (0.2 mM each dNTPs, 2 mM $MgCl_2$, 10 PCR buffer) and 0.2-0.3 μM of each primer. An initial denaturation step of 2 min at 94°C, followed by 40-48 cycles of 30 s at 94°C, 30 s at the indicated annealing temperature (T_a, 61-64°C), 30 s at 72°C, and a final extension of 7 min at 72°C. In one case (*TDP2* cDNA), an additional step of reamplification (20 cycles) was conducted as above, starting from 1 μL of sample obtained after the excision of the expected band from agarose gel and its subsequent syringe-squeezing [1].

Primers pairs were designed with "Amplify3" software [2] following standard criteria and are listed in Supplementary Table available at: http://apollo11.isto.unibo.it/suppl/.

All RT-PCR products obtained were gel analyzed following a standard method [3], purified using a GenElute kit (Sigma-Aldrich, St. Louis, MO), and then subjected to automated sequence analysis of both DNA strands for each fragment, using the same primers utilized in the respective PCR reactions. BigDye chain-terminator method (Applied Biosystems, Carlsbad, CA) was used with an automated Applied Biosystems ABI 3730 DNA automated sequencer.

Sequence extension details for GNB2L1, QARS and TDP2 transcripts and proteins

GNB2L1 (located on 5q35.3), also known as *RACK1* (Receptor for activated C-kinase 1), is an ubiquitously expressed gene encoding a protein, homologous to the G protein β subunit, which can coordinate the interaction of a variety of key signaling molecules. It is believed to play a central role in many biological processes (cell growth, translation, apoptosis, migration, cell cycle, cell division) [4]. Interestingly, while in interaction studies this protein was retrieved as a prey starting from many other proteins as a bait (interactions listed in the Entrez Gene entry at: http://www.ncbi.nlm.nih.gov/gene/10399), to date no study appears to have been designed using GNB2L1 as a bait. Such a study could have the advantage of expressing a more complete product for GNB2L1, including 78 amino acids at its amino terminus (Fig. 2A in the paper), which could reveal additional interactions compared with the product encoded by the currently known cDNA. For example, when Fomenkov et al. [5] reported the extension of the interacting *GNB2L1* cDNA CDS, it appeared not to include the region described here, and the interaction was localized to the C-terminal region of GNB2L1 protein. Notably, our analysis also identifies an in-frame stop codon upstream of the newly determined *GNB2L1* start codon. This would suggest that the extended coding region at 5′ for this mRNA is now complete, since the use of possibly existent further start codons would hesitate in translation stopping upstream of the translation start in the correct frame.

QARS (glutaminyl-tRNA synthetase) gene maps on 3p21.31 and encodes an aminoacyl-tRNA synthetase. Functional data are summarized at the Entrez Gene entry: http://www.ncbi.nlm.nih.gov/gene/5859. The *QARS* cDNA was used in at least one study of human protein-protein pair wise interaction [6] and it was derived from human ORFeome v1.1 database [7], which, in its current 5.1 version (http://horfdb.dfci.harvard.edu/), still lacks codons we have determined here by cDNA sequencing. In fact, in the ORFeome, the cDNA clones from which the *QARS* ORF was deduced reported only few bases upstream of the putative start codon, thus hampering the individuation of further in-frame upstream codons, which were instead identified by

our EST-based analysis. Although, in this case, the extra amino acids are few (Fig. 2B in the paper), making unlikely significant changes in the interaction study, this finding stresses the need for gene annotation refinements. As in the case of *GNB2L1*, the presence of an in-frame stop codon upstream of the newly determined *QARS* start codon suggests that the extended coding region at 5′ for this mRNA is now complete.

The *TDP2* gene, located on 6p22.3-p22.1, encodes a member of a superfamily of divalent cation-dependent phosphodiesterases and also has several interactions described (summary at Entrez Gene http://www.ncbi.nlm.nih.gov/gene/51567), where its cDNA was not used as a bait. Its known CDS appears to lack 30 conserved amino acids corresponding to the protein amino terminus (Fig. 2C in the paper).

References

[1] Q. Li, C.L. Ownby, A rapid method for extraction of DNA from agarose gels using a syringe, BioTechniques 15 (1993) 976-978.

[2] W.R. Engels, Contributing software to the internet: the Amplify program, Trends Biochem. Sci. 18 (1993) 448-450.

[3] L.G. Davis, W.M. Kuehl, J.F. Battey, Basic methods in molecular Biology, Appleton & Lange, Norwalk, 1994.

[4] A. McCahill, J. Warwicker, G.B. Bolger, M.D. Houslay, S.J. Yarwood, The RACK1 scaffold protein: a dynamic cog in cell response mechanisms, Mol. Pharmacol. 62 (2002) 1261-1273.

[5] A. Fomenkov, R. Zangen, Y.P. Huang, M. Osada, Z. Guo, T. Fomenkov, B. Trink, D.

Sidransky, E.A. Ratovitski, RACK1 and stratifin target DeltaNp63alpha for a proteasome degradation in head and neck squamous cell carcinoma cells upon DNA damage, Cell Cycle 3 (2004) 1285-1295.

[6] J.F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G.F. Berriz, F.D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, N. Klitgord, C. Simon, M. Boxem, S. Milstein, J. Rosenberg, D.S. Goldberg, L.V. Zhang, S.L. Wong, G. Franklin, S. Li, J.S. Albala, J. Lim, C. Fraughton, E. Llamosas, S. Cevik, C. Bex, P. Lamesch, R.S. Sikorski, J. Vandenhaute, H.Y. Zoghbi, A. Smolyar, S. Bosak, R. Sequerra, L. Doucette-Stamm, M.E. Cusick, D.E. Hill, F.P. Roth, M. Vidal, Towards a proteome-scale map of the human protein-protein interaction network, Nature 437 (2005) 1173-1178.

[7] J.F. Rual, T. Hirozane-Kishikawa, T. Hao, N. Bertin, S. Li, A. Dricot, N. Li, J. Rosenberg, P. Lamesch, P.O. Vidalain, T.R. Clingingsmith, J.L. Hartley, D. Esposito, D. Cheo, T. Moore, B. Simmons, R. Sequerra, S. Bosak, L. Doucette-Stamm, C. Le Peuch, J. Vandenhaute, M.E. Cusick, J.S. Albala, D.E. Hill, M. Vidal, Human ORFeome version 1.1: a platform for reverse proteomics, Genome Res. 14 (2004) 2128-2135.