

**Supplementary Table. Exemplificative human genes with experimentally confirmed extended cDNA 5' coding region and deduced protein.**

Gene Symbol	Gene Name	RefSeq mRNA GenBank Accession No.	Primer pairs sequence (5' → 3') <sup>a</sup>	RT-PCR product size (Tissue sources)	Error type <sup>b</sup>	GenBank Human EST# <sup>c</sup>	Product length new/reference	No. of new amino acids (% of reference length)
<i>GNB2L1</i>	Guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1	NM_006098	ggaattccatagttggtctc cttgaatgtgcttgttcagag	470 bp (Ovary, Brain)	1	BU172346.1 ES313379.1 BP312588 BP244479	395/317	78 (+25%)
<i>QARS</i>	Glutaminyl-tRNA synthetase	NM_005051	ggatagacgaccttgagcgg gactccgcacatactcaagg	442 bp (Skeletal Muscle, Small Intestine)	1	BI461626.1 BI829834.1 BI463065.1 BM560535	793/775	18 (+2%)
<i>TDP2</i>	Tyrosyl-DNA phosphodiesterase 2	NM_016614	cgcagctgcaccagtttccgag ctcagagatggtttcaggtcg	383 bp (Brain, Bone Marrow)	1	BM554324.1 BG719977.1 BP270589 DA431403	392/362	30 (+8%)

<sup>a</sup> Top: forward primer; bottom: reverse primer (for each gene).

<sup>b</sup> Incomplete exon 1 coding sequence [1]. In these cases, the new coding region is contiguous to the previous known region on the same exon (exon #1). Thus, the actual CDS should be considered as that included between the newly determined "first-AUG" and the known stop.

<sup>c</sup> Four example EST sequences supporting an extended coding sequence at 5' region of the corresponding RefSeq mRNA, resulted from "5'\_ORF\_Extender" software analysis. *GNB2L1*, *QARS* and *TDP2* extensions were supported by a total of 5, 24 and 12 consistent ESTs, respectively. Complete results are available along with the software distribution.