

Supplementary Table 2 Profile of the 5'-end transcripts in HEK293 cells.

Tag sequence	Tag count	Related UniGene cluster	Related refseq	Gene
GCTGTGACAGCCACACAGC	286	Hs.71346	NM_005382	Homo sapiens neurofilament 3 (150kDa medium) (NEF3), mRNA
CTTTCCCTGTGGCAGCAG	171			Multiple hit to genome
CTCTTTCCTTGGCCTAAAG	127			Multiple hit to genome
CTCTTCCGC CGTGTCCG	120	Hs.75309	NM_001961	Eukaryotic translation elongation factor 2
TACCTGGTTGATCCCTGCC	117			Multiple hit to genome
CTGGTGTATCCCTGCCAGT	89			Multiple hit to genome
AACGGGTAGCCTGAGGAG	83	Hs.274402,Hs.75452	NM_005345,NM_005346	Heat shock 70kDa protein 1A
AGTAGCAGCAGCGCCGGG	75	Hs.252400	NM_031243,NM_002137	Heterogeneous nuclear ribonucleoprotein A2/B1
CTCATTGAACCTCGCCTGC	68			Multiple hit to genome
GTGACAGCCACACGCCCC	66	Hs.71346	NM_005382	Homo sapiens neurofilament 3 (150kDa medium) (NEF3), mRNA
AGTTCCGATCGGTAGCCGG	57	Hs.71346	NM_005382	Multiple hit to genome
ACGCTGTGACAGCCACAC	56			Homo sapiens neurofilament 3 (150kDa medium) (NEF3), mRNA
CTTTTCGCAACGGGTTT	55			Multiple hit to genome
AATTTCTACGGCCACCGG	54	Hs.446628	NM_001007	Ribosomal protein S4, X-linked
ACGGCCGAGACCCGCTCC	53	Hs.426930,Hs.510444	NM_001101	Actin, beta
CTTTCACCTGCAAGCGGGC	52	Hs.5662,Hs.509234	NM_006098	Guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1
ATATTTCTTACTCTCTCG	48			Homo sapiens X (inactive)-specific transcript (XIST) on chromosome X
CTCAGTCCGCGCTGCCAG	43	Hs.14376,Hs.500737	NM_006708	Actin, gamma 1
ATTCCTAGTTAAGGGCGC	42	Hs.268849		Glyoxalase I
AAATTGTTCGCCAGCCGC	37	Hs.15589	NM_004774	Multiple hit to genome
CCTCCCTATCACACGCCG	37			PPAR binding protein
CTTTTCGCCCCGTGGACGC	37			Multiple hit to genome
AGTACAGCTCCGGCCGCC	35	Hs.402752	NM_003487,NM_139215	TAF15 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 68kDa
CACCTGTTTGCAGGCTGC	34	Hs.146550		Multiple hit to genome
CTCTCTTTCACGTCAAGG	33	Hs.5662,Hs.509234	NM_006098	Myosin, heavy polypeptide 9, non-muscle
AGTTGCTTCA GCGTCCCG	30	Hs.446628,Hs.449634	NM_005348	Guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1
AGTTCCTCCGGCGTACGGC	29	Hs.211602	NM_006306	Heat shock 90kDa protein 1, alpha
GTCCGTA CTG CAGAGCCG	29	Hs.353170	NM_004343	SMC1 structural maintenance of chromosomes 1-like 1 (yeast)
AAAAACGGCCAGCCTGAGG	27	Hs.75452		Calreticulin
AFTTCGTCTTAGCCACGC	26	Hs.202166		FLJ38698
AGGCATTTGAGGACGCCAG	25			Heterogeneous nuclear ribonucleoprotein HI (H)
AGTGGGCCGGACCCGCCGG	25	Hs.192374	NM_003299	Hit to genome
CCCAATTTCTACGGCCAC	25	Hs.446628	NM_001007	Tumor rejection antigen (gp96) 1
CTCGTTGGCAGTAGTGC	25	Hs.380118,Hs.460941		Ribosomal protein S4, X-linked
GTGCTGCAGCCGCTGCCG	25	Hs.2795		RNA binding motif protein, X-linked
CATTTTCGCTTAGCCACG	24	Hs.202166	NM_005566	Lactate dehydrogenase A
CTCTTTCCTTAAGCAGCC	24			Heterogeneous nuclear ribonucleoprotein HI (H)
GACTAATTTGTTGGCGGC	24	Hs.280311		Multiple hit to genome
ACCCTATTCA TTTCTACC	23	Hs.279806	NM_004396	Myosin, heavy polypeptide 10, non-muscle
CCTTTTCGCCCCGTGACG	23			DEAD (Asp-Glu-Ala-Asp) box polypeptide 5
AGTATCTGTGGGTACCCG	22	Hs.433455,Hs.331035	NM_001428	Multiple hit to genome
CAATTTTACGGCACC	22	Hs.446628	NM_001007	Enolase 1, (alpha)
GCACACAGGCATCCATCC	22	Hs.107600	NM_006158	Ribosomal protein S4, X-linked
AGTGACGGGTATTTGCTG	19	Hs.75337,Hs.467172	NM_004741	Neurofilament, light polypeptide 68kDa
CTCTTTCAGCCAGCGCC	19			Nucleolar and colicel-body phosphoprotein 1
CTTTTCCGCCCGTCC	19	Hs.374596	NM_003295	Multiple hit to genome
GGCTCTTGTCTTGGCTG	19	Hs.180909	NM_181696,NM_181697	Tumor protein, translationally-controlled 1
ATATAGAGGCTGGGGGGG	18	Hs.427152		Peroxisomal protein 1
ATTAACCGGTTGCAGGGG	18	Hs.78996,Hs.449476	NM_002592	High density lipoprotein binding protein (vigitin)
CCTTTTGGCTCTCTGACC	18			Proliferating cell nuclear antigen
				Multiple hit to genome

Top 50 5'-end transcripts expressed in HEK293 cells are listed. The tag sequence represents the 18-bp SAGE tag. Tags associated with their corresponding UniGene/EST are listed. All of 5' SAGE tag data are available at <http://sage.gi.k.u-tokyo.ac.jp/>.