CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Evidence of efficient stop codon readthrough in four mammalian genes
Authors	Loughran, Gary;Chou, Ming-Yuan;Ivanov, Ivaylo P.;Jungreis, Irwin;Kellis, Manolis;Kiran, Anmol M.;Baranov, Pavel V.;Atkins, John F.
Publication date	2014
Original Citation	Loughran, G., Chou, M.-Y., Ivanov, I. P., Jungreis, I., Kellis, M., Kiran, A. M., Baranov, P. V. and Atkins, J. F. (2014) 'Evidence of efficient stop codon readthrough in four mammalian genes', Nucleic Acids Research, 42(14), pp. 8928-8938. doi: 10.1093/nar/gku608
Type of publication	Article (peer-reviewed)
Link to publisher's version	https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gku608 - 10.1093/nar/gku608
Rights	© 2014, the Authors. Published by Oxford University Press on behalf of Nucleic Acids Research. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited. - https://creativecommons.org/licenses/by/4.0/
Download date	2026-10-31 02:29:47
Item downloaded from	https://hdl.handle.net/10468/5016

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Supplementary Figure 1. Comparative evidence for 3 new readthrough candidates: (A-C) Genomic alignments of the readthrough extensions of *ACP2*, *AQP4*, and *MAPK10* with other eutherian mammal genomes, colour coded using CodAlignView ("CodAlignView: a tool for visualizing protein-coding constraint", I Jungreis, M Lin, M Kellis, in preparation). The high concentrations of synonymous substitutions (light green) and conservative amino acid changes (dark green), and relatively low concentrations of radical amino acid changes (red) and frame shifted regions (orange) is characteristic of protein-coding regions. PhyloCSF scores for these extensions are higher than 99% of similar-sized non-coding regions, indicating that they have been functional at the amino acid level in much of the eutherian mammal tree. Horizontal lines in the *ACP2* alignment indicate paired bases in the predicted RNA stem loop. (D) Conserved RNA stem loop 3' of the UGA stop codon in *ACP2*, as predicted by RNAz.

Supplementary Figure 2. Per-codon PhyloCSF scores of the predicted readthrough extensions of the seven readthrough candidates (blue stars) relative to the distributions for coding and non-coding regions. Non-coding distribution was computed from "third ORFs", i.e., the region from the first in-frame stop codon after an annotated stop codon to the next in-frame stop codon, excluding regions that overlap the coding region of another annotated transcript. Coding distribution was computed from same-sized regions directly 5' of the annotated stop codon. Regions less than 10 codons or for which the aligned species encompass less than 10% of the branch length of the 29-mammals tree were excluded because PhyloCSF has insufficient power on such regions. All seven readthrough regions have scores near the middle of the coding distribution and higher than 99.96% of non-coding regions.

Supplementary Figure 3. Nucleotide and codon alignments of the regions corresponding to the end of the main ORF, the readthrough ORF and the flanking sequences of (A) *AQP4*, (B) *MAPK10*, (C) *OPRK1*, (D) *OPRL1* and (E) *SACM1L*. The alignments in A, B, C and D were generated by custom python scripts and clustalW2. First the given region in the main ORF was conceptually translated into a peptide sequence. The conceptual peptide sequences were aligned using ClustalW2. This amino acid alignment was back-translated to a corresponding nucleotide alignment based on codons corresponding to amino acids in the alignment using the initial nucleotide sequences. Finally, the common ancestral (root) sequence was inferred using PyCogenet python module v. 1.5.3 (1) using a ClustalW guide tree. The alignment shown in E used a variant of the procedure for A-D where an alignment was performed directly on the nucleotide sequence with ClustalX2. This nucleotide alignment was then manually adjusted and converted into a corresponding protein sequence alignment which was used to generate a guide tree and evaluate the root sequence as described earlier. Synonymous and non-synonymous substitutions were determined relative to the inferred root sequence shown on top of each alignment. In these alignments, consensus (root) codons (and gaps) are highlighted in white, synonymous codons are highlighted in green, non-synonymous codons are highlighted in brown and stop codons are highlighted in magenta. Abbreviated species identifiers of the sequences used are indicated on the left. Full names of the species whose sequences are used in A-E: *Ailuropoda melanoleuca, Alligator mississippiensis, Anolis carolinensis, Bos Taurus, Callithrix*

jacchus, Callorhinchus milii, Camelus ferus, Canis lupus familiaris, Ceratotherium simum, Chrysemys picta bellii, Condylura cristata, Cricetulus griseus, Danio rerio, Dasypus novemcinctus, Dicentrarchus labrax, Elephantulus edwardii, Equus caballus, Erinaceus europaeus, Felis catus, Ficedula albicollis, Gadus morhua, Gallus gallus, Gasterosteus aculeatus, Geospiza fortis, Haplochromis burtoni, Homo sapiens, Latimeria chalumnae, Lepisosteus oculatus, Leucoraja erinacea, Lipochromis sp., Loxodonta africana, Macaca mulatta, Macropus eugenii, Maylandia zebra, Mchenga conophoros, Meleagris gallopavo, Melopsittacus undulates, Monodelphis domestica, Mus musculus, Mustela putorius furo, Myotis davidii, Neolamprologus brichardi, Nomascus leucogenys, Ochotona princeps, Odobenus rosmarus, Oncorhynchus mykiss, Orcinus orca, Oreochromis niloticus, Ornithorhynchus anatinus, Orycteropus afer, Oryctolagus cuniculus, Oryzias latipes, Otolemur garnettii, Ovis aries, Pelodiscus sinensis, Perca flavescens, Pimephales promelas, Procavia capensis, Pteropus alecto, Pteropus vampyrus, Pundamilia nyererei, Python molurus, Rattus norvegicus, Saimiri boliviensis boliviensi, Salmo salar, Sarcophilus harrisii, Sus scrofa, Taeniopygia guttata, Takifugu rubripes, Tarsius syrichta, Tetraodon nigroviridis, Torpedo californica, Trichechus manatus, Tursiops truncates, Vicugna pacos, Xenopus laevis, Xenopus tropicalis, Xiphophorus maculates. The sequences used in the alignments here were obtained from GenBank using the BLAST (tblastn) algorithm where the translated human sequence was used as bait. The target databases during this search were Expressed Sequence Tags (EST), the Whole-Genome Shotgun contigs, and Transcriptome Shotgun Assembly. All non-mammalian orthologous sequences that could be identified, as well as representative number of mammalian sequences were collected and used in the alignments.

Supplementary Figure 4. Amino acid conservation of OPRK1 and OPRL1 readthrough extensions. The conservation is shown as amino acid logogram (A) of the OPRK1, and (B) the ORPL1 readthrough extensions, respectively. The five highly conserved amino acids near their C-terminal ends, that also show sequence similarity between the two paralogs, are boxed in read. To build the conservation plots the nucleotide sequence alignment for the *OPRK1* and *OPRL1* readthrough extensions (Supplementary Figures 3C and D) were converted into amino acid sequence alignments and used to generate logogram utilizing WebLogo (2).

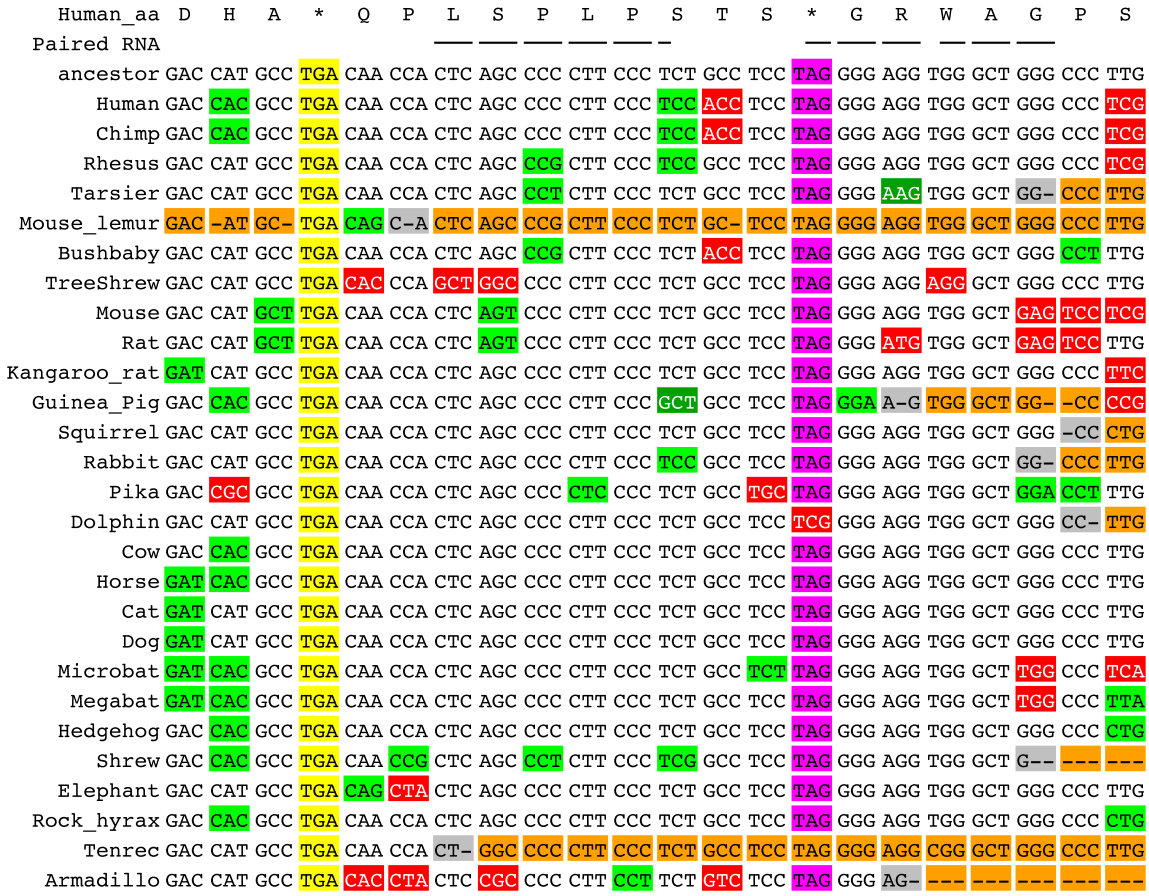
Supplementary Figure 5. (a) Dual luciferase assay indicating readthrough efficiencies of *OPRL1* constructs with only one 5' nt and with 3' nested deletions showing the importance of the nt immediately 3'. (B) The effect of nt identity immediately 5' of the *OPRL1* UGA on readthrough efficiency. (C) The effect of replacing the UGA stop codon with either UAA or UAG on the readthrough efficiencies of wild type *OPRL1* and *OPRL1*-CARYYA.

REFERENCES

1. Knight, R., Maxwell, P., Birmingham, A., Carnes, J., Caporaso, J.G., Easton, B.C., Eaton, M., Hamady, M., Lindsay, H., Liu, Z. *et al*. (2007) PyCogent: a toolkit for making sense from sequence.Genome Biol., 8, R171.

2. Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. Genome Res., 14, 1188-1190.

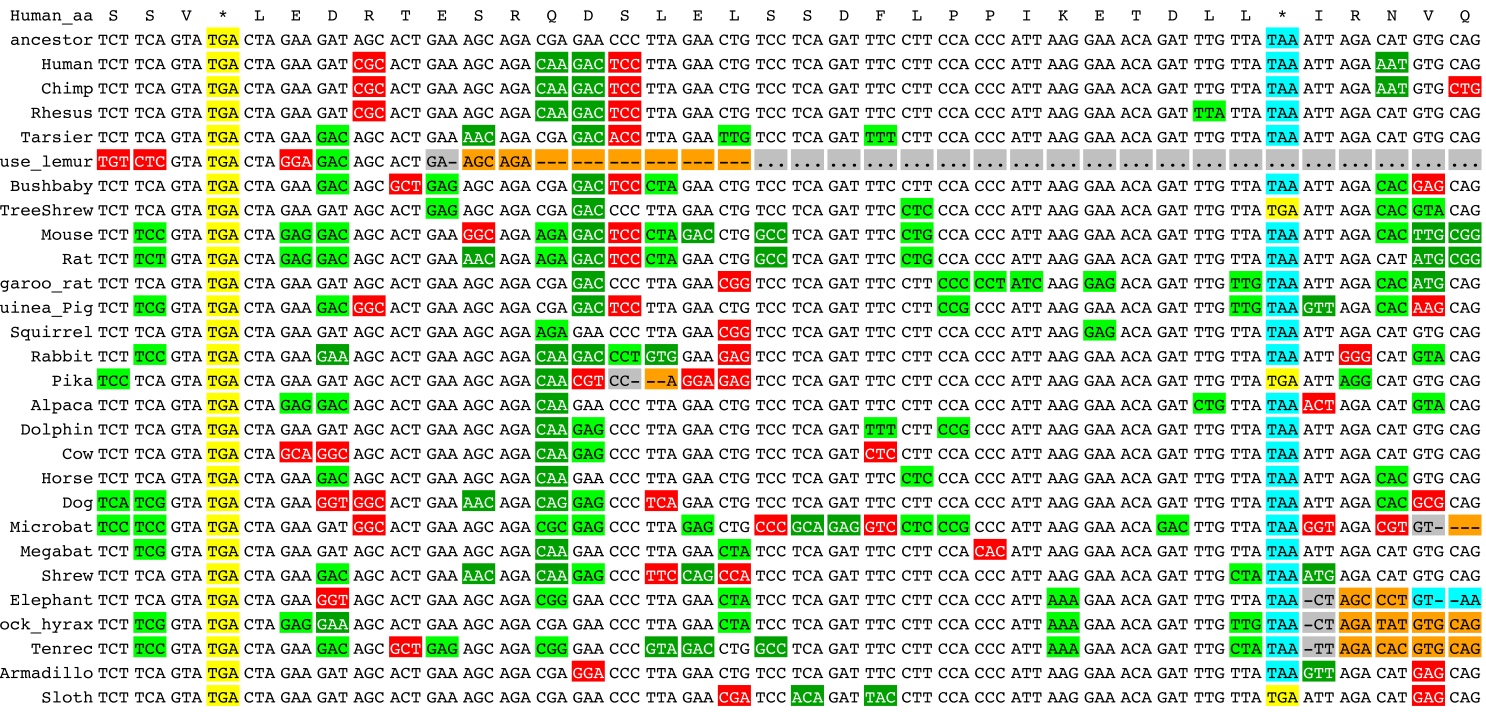
Supplementary Figure 1A



ACP2

No Change	GAC
Synonymous	GAT
Conservative	GAA
Radical	GGG
TAA Stop	TAA
TAG Stop	TAG
TGA Stop	TGA
Indel	GA-
Frameshifted	GAC

Supplementary Figure 1B



AQP4

No Change	GAC
Synonymous	GAT
Conservative	GAA
Radical	GGG
TAA Stop	TAA
TAG Stop	TAG
TGA Stop	TGA
Indel	GA-
Frameshifted	GAC

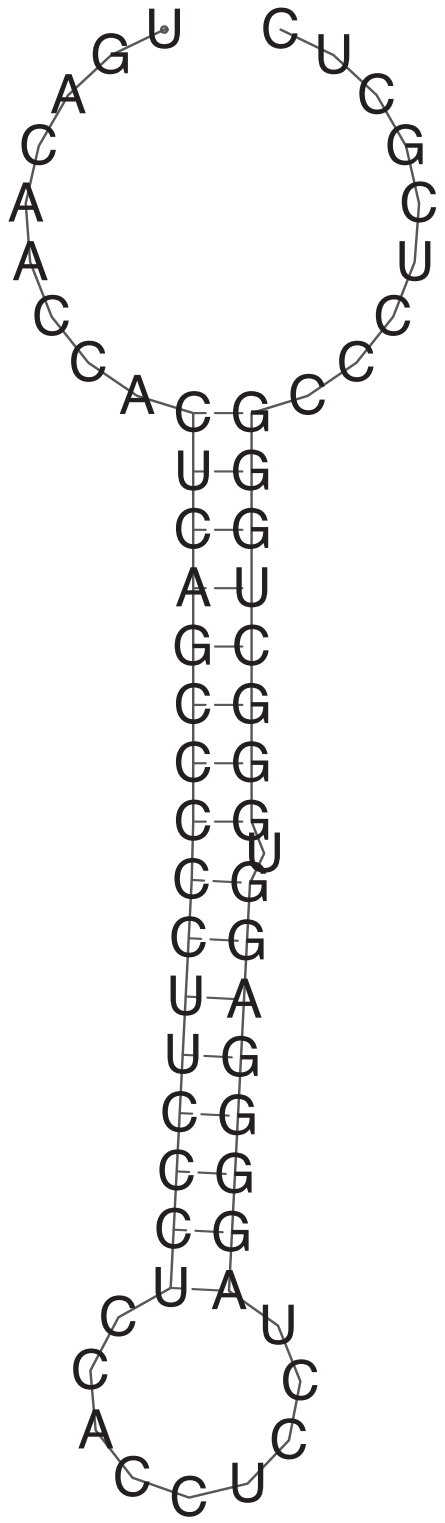
Supplementary Figure 1C

Human_aa	C	C	R	*	L	A	A	C	L	R	N	P	A	F	F	R	R	*		C		D		G		T
ancestor	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	TG	---	T	---	GAT	GGA	---	ACA
Human	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	TG	---	T	---	GAT	GGA	---	ACA
Chimp	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	TG	---	T	---	GAT	GGA	---	ACA
Rhesus	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	TG	---	T	---	GAT	GGA	---	ACA
Tarsier	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAT	CCA	GCG	TTC	TTC	AGG	AGA	TGA	TG	---	T	---	GAT	GGA	---	ACA
Mouse_lemur	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	CG	---	T	---	GAT	GGA	---	GCA
Bushbaby	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	TG	---	T	---	GAT	GGA	---	ACA
TreeShrew	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	TG	---	A	---	GAT	GGA	---	ACA
Mouse	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	CG	---	C	---	GAT	AGA	---	ACA
Rat	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	CG	---	CCAT		GAT	AGA	---	ACA
Kangaroo_rat	TGC	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCA	TTC	TTC	AGG	AGA	TGA	TA	---	C	---	GAT	AGA	---	ACA
Guinea_Pig	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	TTG	CGG	AAC	CCA	GCG	TTC	TTC	AGG	AGC	TGA	GG	---	G	---	GAT	GGA	---	ACA
Squirrel	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	TG	---	T	---	GAT	GGA	---	ACA
Rabbit	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	TG	---	T	---	GAT	GGA	---	ACA
Pika	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	TGATGT		---		GAT	GGA	---	ACA
Alpaca	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCG	GCG	TTC	TTC	AGG	AGA	TGA	TG	---	T	---	GAT	GGA	---	ACA
Dolphin	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	ACG	TTC	TTC	AGG	AGA	TGA	TG	---	T	---	GAT	GGA	---	ACA
Cow	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	TG	---	T	---	GAT	GGA	---	ACA
Horse	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	TG	---	T	---	GAT	GGAGCAC		ACA
Cat	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCA	TTC	TTC	AGG	AGA	TGA	TG	---	C	---	GAT	GGA	---	ACA
Dog	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	TG	---	T	---	GAT	GGA	---	ACA
Microbat	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGC	AGA	TGA	TG	---	G	---	GAT	GGA	---	ACA
Megabat	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	TG	---	C	---	GAT	GGG	---	ACA
Hedgehog	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	AGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	CA	---	T	---	GAC	GGA	---	ACA
Shrew	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	CA	---	T	---	GAT	GGA	---	ACA
Elephant	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	TG	---	C	---	GAT	GGA	---	ACA
Rock_hyrax	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	TG	---	C	---	GAT	GGA	---	ACA
Tenrec	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	TG	---	C	---	GAT	GGG	---	ATA
Armadillo	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	TG	---	G	---	GAT	GGA	---	ACA
Sloth	TGT	TGC	AGG	TGA	CTA	GCC	GCC	TGC	CTG	CGA	AAC	CCA	GCG	TTC	TTC	AGG	AGA	TGA	TG	---	T	---	GAT	GGA	---	ACA

MAPK10

No Change	GAC
Synonymous	GAT
Conservative	GAA
Radical	GGG
TAA Stop	TAA
TAG Stop	TAG
TGA Stop	TGA
Indel	GA-
Frameshifted	GAC

Supplementary Figure 1D



Supplementary Figure 2

