

SUPPLEMENTARY TABLES

Supplementary Table 1

Distribution of sequence reads. The triplicate batches from the cycloheximide footprints are labeled “THP-1 -puro” and sequence reads from the puromycin experiments are labeled “THP-1 +puro”. Only unique hits to annotated RefSeq sequences (right column) were analyzed further after the removal of ribosomal contamination (column: “Ribosomal matches”).

Replicate	Total no. reads	No. ribosomal matches	No. unique hits to annotated RefSeq sequences
THP-1 -puro #1	96,712,346	69562236	14762803
THP-1 -puro #2	101,960,552	71142855	18087352
THP-1 -puro #3	97,677,373	68161986	18556057
THP-1 +puro #1	88,213,584	73869737	5585473
THP-1 +puro #2	95,946,401	81697924	5930977
THP-1 +puro #3	99,128,594	80998596	7189773

Supplementary Table 2

Positive TIS signals emitted by the neural network.

Analyzed transcripts	5062
Positions tested	1,304,713
Positive signals	14,464
Merged positive signals (over 2 bp window)	10,386
Merged positive signals in target regions	8710
Classified TISs	7251
Transcripts with at least one predicted TIS in target region	4364
Transcripts without any positive TIS signal in target region	698

Supplementary Table 3

Comparative analysis of TIS prediction in the pooled data and the three individual replicates.

Given are the predicted TIS in 1720 transcripts matching the minimum expression criteria of 20x coverage at least one nucleotide position.

Replicate	Number of detected TIS
Pool, Replicates 1, 2, 3	1626
Pool, Replicates 1, 2	194
Pool, Replicates 1, 3	188
Pool, Replicates 2, 3	149
Pool, Replicates 1	82
Pool, Replicates 2	76
Pool, Replicates 3	50
Replicate 1 only	325
Replicate 2 only	275
Replicate 3 only	215
Pool only	20
Replicates 1, 2	59
Replicates 1, 3	46
Replicates 2, 3	38
Replicates 1, 2, 3	13

Supplementary Table 4

Estimation of false positive rate among the newly identified TIS: A total of 5166 TISs (i.e. the number of newly annotated TIS in the 5' UTR identified in this study) were randomly drawn from the analyzed 5'UTRs and the 3 basepair annotation window was analyzed for the presence of either an AUG or near cognate TIC. Ten replicates of these random draws are shown in the Table below. A mean of 1970.2 of these randomly drawn TISs did not have an AUG or near cognate TIC in its 3 basepair neighborhood as used in the TIC annotation pipeline. This corresponds to 38.1 % of the randomly drawn TISs. This implies that not more than 577 (i.e. $220 / 0.381$) of the 5166 TISs are due to random noise, implying that 88.8% might be correctly annotated by our pipeline. These calculations hinge on several assumptions and thus have to be viewed with some caution: First, we cannot prove, that the random proportion of unclassifiable TIS (i.e. $\sim 1/3$) is a lower limit to the expected proportion of unclassifiable TIS among the false positive predictions made by the neural network. The random proportion may be larger or smaller depending upon the (unknown) way in which TIS classifiability influences the prediction specificity of the network. Second, the above statement also hinges on the assumption that all true positive TIS predictions would be classifiable, for which cannot be stated with certainty either.

No.	AUG or near-cognate	other
1	3261	1905
2	3211	1955
3	3217	1949
4	3222	1944
5	3164	2002
6	3190	1976
7	3174	1992

8	3183	1983
9	3196	1970
10	3140	2026
mean	3195.8	1970.2
Rate false positive in a random drawing		38.1 %
Estimation false positive		576.9
Rate false positive		11.2 %

Supplementary Table 5

Genes with non-AUG-initiated N-terminal protein extension and experimentally validated uORFs which did not meet the minimal read coverage criteria or are not covered by the correct isoform. Genes with non-AUG-initiated N-terminal protein extension were selected from the “database of mRNA sequences with non-AUG start codons” (<http://bioinfo.iitk.ac.in/>) and with known functional uORFs were identified from the report by Calvo et al. 2009.

RefSeq ID	Gene	Classification	low translation	wrong Isoform	Reference
NM_031895	CACNG8	non-AUG start	yes		Ivanov et al. 2011
NM_002006	FGF2	non-AUG start	yes		E Arnaud et al. 1999, H Prats et al. 1989
NM_021182	HMHB1	non-AUG start	yes		Dolstra et al. 1999)
NM_001128619	LUZP6	non-AUG start	yes		Xiong et al. 2006)
NM_199072	MDFIC	non-AUG start	yes		Thébault et al. 2000)
NM_001098579	MRVI1	non-AUG start	yes		Shaughnessy et al. 1999)
NM_014293	NPTXR	non-AUG start	yes		Dodds et al. 1997)
NM_022002	NR1I2	non-AUG start	yes		Lehmann et al. 1998)
NM_001134939	OAZ3	non-AUG start	yes		Ivanov et al. 2011
NM_001172438	PEG10	non-AUG start	yes		Heike Lux et al. 2010
NM_001166304	PIGX	non-AUG start	yes		Ashida et al. 2005
NM_002701	POU5F1	non-AUG start	yes		Xia Wang et al. 2009
NM_175886	PRPS1L1	non-AUG start	yes		Taira et al. 1990
NM_001169117	STIM2	non-AUG start	yes		R T Williams et al. 2001
NM_021961	TEAD1	non-AUG start	yes		J H Xiao et al. 1991
NM_003214	TEAD3	non-AUG start	yes		Jacquemin et al. 1997
NM_052883	TXNRD3	non-AUG start	yes		Gerashchenko et al. 2010
NM_001025366	VEGFA	non-AUG start	yes		Meiron et al. 2001, Tee and Jaffe 2001, Huez et al. 2001
NM_003111	SP3	non-AUG Start		yes	Hernandez et al. 2002
NM_000392	ABCC2	uORF	yes		Yuanyuan Zhang et al. 2007
NM_031850	AGTR1	uORF	yes		Mickey M Martin et al. 2006
NM_000707	AVPR1B	uORF	yes		Nomura et al. 2001, Rabadan-Diehl et al. 2007
NM_012104	BACE1	uORF	yes		Rogers et al. 2004

NM_001166	<i>BIRC2</i>	uORF	yes		Warnakulasuriyarachchi et al. 2003
NM_004345	<i>CAMP</i>	uORF	yes		Wu et al. 2002
NM_001001548	<i>CD36</i>	uORF	yes		Griffin et al. 2001
NM_000784	<i>CYP27A1</i>	uORF	yes		Lodhi et al. 2003
NM_004448	<i>ERBB2</i>	uORF	yes		Child et al. 1999, Mehta et al. 2006, Spevak et al. 2006
NM_001122742	<i>ESR1</i>	uORF	yes		Kos et al. 2002, Pentecost et al. 2005
NM_005257	<i>GATA6</i>	uORF	yes		Takeda et al. 2004
NM_001190468	<i>GDNF</i>	uORF	yes		Tanaka et al. 2001
NM_004810	<i>GRAP2</i>	uORF	yes		Guyot et al. 2002
NM_005577	<i>LPA</i>	uORF	yes		Zysow et al. 1995
NM_005372	<i>MOS</i>	uORF	yes		Steel et al. 1996
NM_138768	<i>MYEOV</i>	uORF	yes		de Almeida et al. 2006
NM_022162	<i>NOD2</i>	uORF	yes		Rosenstiel et al. 2007
NM_001145281	<i>OPRM1</i>	uORF	yes		Kyu Young Song et al. 2007
NM_000965	<i>RARB</i>	uORF	yes		Reynolds et al. 1996
NM_003745	<i>SOCS1</i>	uORF	yes		Schlüter et al. 2000
NM_003189	<i>TAL1</i>	uORF	yes		Calkhoven et al. 2003
NM_000459	<i>TEK</i>	uORF	yes		Park et al. 2006
NM_000460	<i>THPO</i>	uORF	yes		Stockklausner et al. 2006
NM_001025366	<i>VEGFA</i>	uORF	yes		Bastide et al. 2008
NM_004185	<i>WNT2B</i>	uORF	yes		Tang et al. 2008
NM_000050	<i>ASS1</i>	uORF		yes	Pendleton et al. 2005
NM_004064	<i>CDKN1B</i>	uORF		yes	Göpfert et al. 2003
NM_005194	<i>CEBPB</i>	uORF		yes	Lincoln et al. 1998
NM_001018077	<i>NR3C1</i>	uORF		yes	Diba et al. 2001
NM_001113491	<i>SEPT9</i>	uORF		yes	McDade et al. 2007

Supplementary Table 6

Conservation among primates of the TICs at classified network-identified TISs (N-TERM: in frame N-terminal protein extension, uORF: upstream open reading frame; ovORF: CDS-overlapping uORF). P values refer to a t-test comparing the mean conservation score among case and control TICs. Significant p values after Bonferroni correction for multiple testing (i.e. $p < 0.0016$) are marked in bold italics. DIFF: difference in mean Conservation Score between case and control TICs, CI: 95% confidence interval.

TYPE	CODON	MEAN.CASE	MEAN.CONT	DIFF	CI	P	N.CASE	N.CONT
CDS	ATG	0.580	0.202	0.378	0.358 - 0.397	$<10^{-10}$	2299	403
NTERM	AAG	0.661	0.254	0.407	0.183 - 0.631	1.32×10^{-3}	6	793
NTERM	ACG	0.253	0.259	-0.006	-0.063 - 0.050	0.754	51	894
NTERM	AGG	0.300	0.292	0.008	-0.101 - 0.117	0.882	16	2727
NTERM	ATA	0.179	0.194	-0.015	-0.162 - 0.132	0.832	5	152
NTERM	ATC	0.239	0.236	0.003	-0.064 - 0.070	0.901	33	494
NTERM	ATG	0.265	0.205	0.060	-0.111 - 0.230	0.473	6	184
NTERM	ATT	0.272	0.222	0.051	-0.049 - 0.151	0.316	19	328
NTERM	CTG	0.267	0.240	0.027	0.001 - 0.054	0.045	270	1887
NTERM	GTG	0.218	0.268	-0.050	-0.095 - -0.005	0.030	80	1613
NTERM	TTG	0.225	0.256	-0.031	-0.083 - 0.021	0.240	60	976
uORF	AAG	0.314	0.278	0.036	-0.047 - 0.120	0.389	29	1558
uORF	ACG	0.351	0.263	0.087	0.049 - 0.126	1.03×10^{-5}	143	898
uORF	AGG	0.353	0.291	0.061	0.007 - 0.116	0.028	66	2800
uORF	ATA	0.393	0.213	0.180	0.098 - 0.262	3.08×10^{-5}	34	308
uORF	ATC	0.323	0.235	0.088	0.041 - 0.136	2.79×10^{-4}	99	531
uORF	ATG	0.469	0.206	0.263	0.239 - 0.286	$<10^{-10}$	924	397
uORF	ATT	0.387	0.248	0.140	0.093 - 0.186	1.17×10^{-8}	103	634
uORF	CTG	0.254	0.244	0.011	-0.004 - 0.026	0.150	1045	2039
uORF	GTG	0.273	0.270	0.003	-0.022 - 0.027	0.753	342	1645
uORF	TTG	0.335	0.257	0.078	0.044 - 0.111	7.95×10^{-6}	209	889
ovORF	AAG	0.287	0.268	0.019	-0.070 - 0.109	0.667	20	1472
ovORF	ACG	0.252	0.260	-0.008	-0.047 - 0.031	0.695	103	912

ovORF	AGG	0.328	0.287	0.041	-0.041 - 0.123	0.326	28	2304
ovORF	ATA	0.449	0.181	0.267	0.131 - 0.404	2.88×10^{-4}	14	158
ovORF	ATC	0.266	0.236	0.030	-0.022 - 0.082	0.258	71	546
ovORF	ATG	0.573	0.208	0.365	0.317 - 0.413	$<10^{-10}$	116	338
ovORF	ATT	0.253	0.247	0.006	-0.058 - 0.071	0.846	47	540
ovORF	CTG	0.224	0.243	-0.019	-0.036 - -0.003	0.023	664	1988
ovORF	GTG	0.271	0.267	0.005	-0.024 - 0.033	0.675	215	1557
ovORF	TTG	0.309	0.258	0.051	0.011 - 0.091	0.012	128	992

Supplementary Table 7

Local rate of variation, among primates, of the TICs at classified network-identified TISs (N-TERM: in frame N-terminal protein extension, uORF: upstream open reading frame; ovORF: CDS-overlapping uORF). The local rate of variation was estimated using SiPhy (Garber et al. 2009). P values refer to a t-test comparing the mean local rate of variation among case and control TICs. Multiple alignments and the phylogenetic tree model were taken from the UCSC genome browser (<http://hgdownload.cse.ucsc.edu/goldenPath/hg19/multiz46way/maf/>, <http://hgdownload.cse.ucsc.edu/goldenPath/hg19/phastCons46way/primates.mod>). Significant p values after Bonferroni correction (i.e. $p < 0.0016$) are marked in bold italics. DIFF: difference in mean local rate of variation between case and control TICs, CI: 95% confidence interval.

TYPE	CODON	MEAN.CASE	MEAN.CONT	DIFF	CI	P	N.CASE	N.CONT
CDS	ATG	0.154	1.792	-1.638	-1.828 - -1.448	<i><10⁻¹⁰</i>	2299	403
NTERM	AAG	0.759	1.337	-0.579	-1.282 - 0.125	0.101	6	793
NTERM	ACG	1.561	1.736	-0.174	-0.615 - 0.266	0.436	51	894
NTERM	AGG	1.153	1.305	-0.152	-0.734 - 0.429	0.601	16	2727
NTERM	ATA	1.826	1.696	0.130	-1.071 - 1.331	0.821	5	152
NTERM	ATC	1.616	1.604	0.012	-0.530 - 0.554	0.890	33	494
NTERM	ATG	0.891	1.693	-0.802	-2.073 - 0.470	0.205	6	184
NTERM	ATT	1.570	1.486	0.084	-0.758 - 0.927	0.843	19	328
NTERM	CTG	1.531	1.581	-0.050	-0.289 - 0.188	0.680	270	1887
NTERM	GTG	1.806	1.364	0.442	-0.048 - 0.931	0.077	80	1613
NTERM	TTG	1.966	1.492	0.474	-0.052 - 1.000	0.077	60	976
uORF	AAG	1.519	1.249	0.270	-0.409 - 0.949	0.431	29	1558
uORF	ACG	1.444	1.729	-0.285	-0.589 - 0.019	0.067	143	898
uORF	AGG	1.113	1.315	-0.202	-0.503 - 0.099	0.188	66	2800
uORF	ATA	1.587	1.686	-0.100	-0.723 - 0.524	0.752	34	308
uORF	ATC	1.646	1.605	0.040	-0.379 - 0.459	0.813	99	531
uORF	ATG	0.771	1.754	-0.983	-1.189 - -0.779	<i><10⁻¹⁰</i>	924	397
uORF	ATT	0.955	1.444	-0.489	-0.779 - -0.200	<i>0.001</i>	103	634

uORF	CTG	1.513	1.547	-0.033	-0.163 - 0.097	0.615	1045	2039
uORF	GTG	1.318	1.401	-0.083	-0.285 - 0.119	0.421	342	1645
uORF	TTG	1.400	1.457	-0.057	-0.310 - 0.196	0.659	209	889
ovORF	AAG	1.817	1.283	0.534	-0.318 - 1.386	0.215	20	1472
ovORF	ACG	1.501	1.741	-0.240	-0.556 - 0.076	0.136	103	912
ovORF	AGG	1.109	1.326	-0.217	-0.770 - 0.336	0.437	28	2304
ovORF	ATA	1.685	1.867	-0.182	-1.611 - 1.249	0.799	14	158
ovORF	ATC	1.497	1.594	-0.097	-0.500 - 0.307	0.642	71	546
ovORF	ATG	0.850	1.714	-0.864	-1.204 - -0.523	7.97×10^{-7}	116	338
ovORF	ATT	1.819	1.402	0.417	-0.122 - 0.956	0.129	47	540
ovORF	CTG	1.619	1.614	0.005	-0.153 - 0.163	0.794	664	1988
ovORF	GTG	1.294	1.383	-0.088	-0.317 - 0.140	0.451	215	1557
ovORF	TTG	1.697	1.441	0.257	-0.106 - 0.619	0.165	128	992

Supplementary Table 8

Conservation of TIC usage in human and mouse. The multiple alignment files of the mouse assembly to the human genome (hg19/GrCh37) was taken from the UCSC Genome Browser <http://hgdownload.cse.ucsc.edu/goldenPath/hg19/multiz46way/maf/> The human 5'UTR regions were aligned to the appropriate mouse genome (mm9) position. For mapping of the mouse transcripts only those transcripts with at least one TIS and with a distinct assignment were used for further analysis. TICs for the analyzed categories canonical, N-terminal extension, uORF, and overlapping uORF were taken from Ingolia et al. 2011

	Human	Conserved in mouse	Mouse	Conserved in human
TIC total	3294	2141 (65%)	7391	4458 (60.3%)
canonical	1321	1293 (97.8%)	2350	2312 (98.4%)
N- terminal extension	243	98 (40.3%)	404	157 (38.9%)
uORF	970	455 (46.9%)	1748	833 (47.6)
Overlapping uORF	681	263 (38,6%)	1041	427 (41%)

Supplementary Table 9

TIC usage in human and mouse. Mouse TISs were aligned to the human genome and TIC usage of TISs used in both species are compared.

	mouse											
		ATG	CTG	GTG	TTG	AAG	AGG	ATA	ATC	ATT	ACG	other
human	ATG	1279	1	3	0	0	0	1	0	0	2	0
	CTG	1	94	8	3	0	0	0	0	0	1	0
	GTG	0	4	33	2	0	0	0	0	0	2	0
	TTG	0	2	0	11	0	0	1	1	0	0	0
	AAG	0	0	0	0	1	0	0	0	0	0	0
	AGG	0	0	0	0	0	0	0	0	0	0	0
	ATA	0	0	0	0	0	0	3	0	0	0	0
	ATC	0	0	0	0	0	0	0	5	1	0	0
	ATT	0	0	0	0	0	0	0	1	6	0	0
	ACG	0	2	1	0	0	0	0	0	0	19	0
	other	0	0	0	0	0	0	0	1	0	1	0

TIC usage in human and mouse uORFs. Mouse TISs were aligned to the human genome and TIC usage of TISs used in both species are compared.

[illegible]

Supplemnetary Table 11

TIC usage in human and mouse overlapping uORF. Mouse TISs were aligned to the human genome and TIC usage of TISs used in both species are compared.

[illegible]

Supplementary Table 12

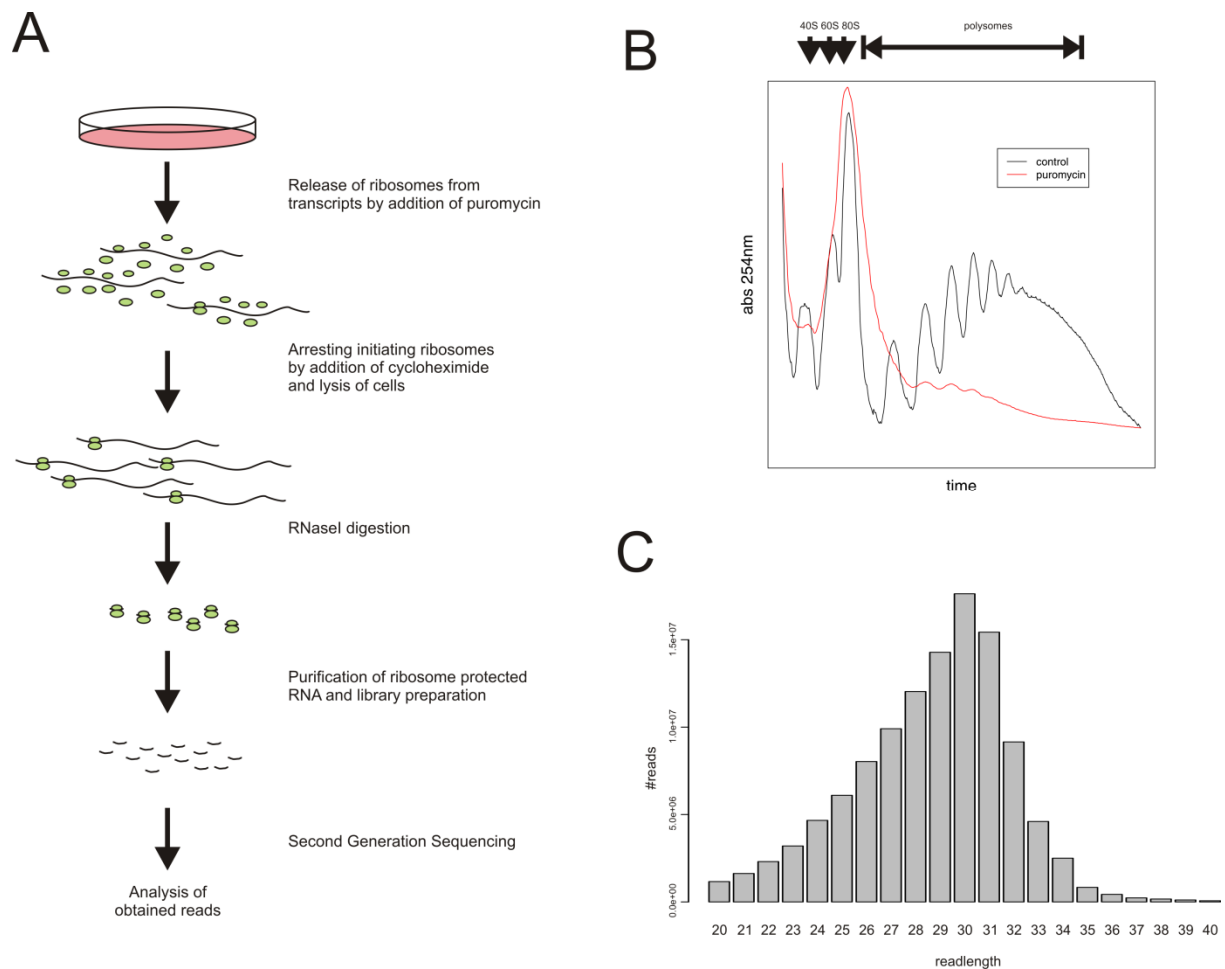
Transcript based analysis of TIS usage in human and mouse. Mouse transcripts harboring a experimentally determined TIS were compared to human transcripts with a TIS.

	Human	Mouse	Both
Transcripts	2216	2216	2216
canonical	1192	2042	1135
N-terminal extension	292	296	73
uORF	932	774	530
Overlapping uORF	734	680	323

SUPPLEMENTARY FIGURES

Supplementary Figure 1

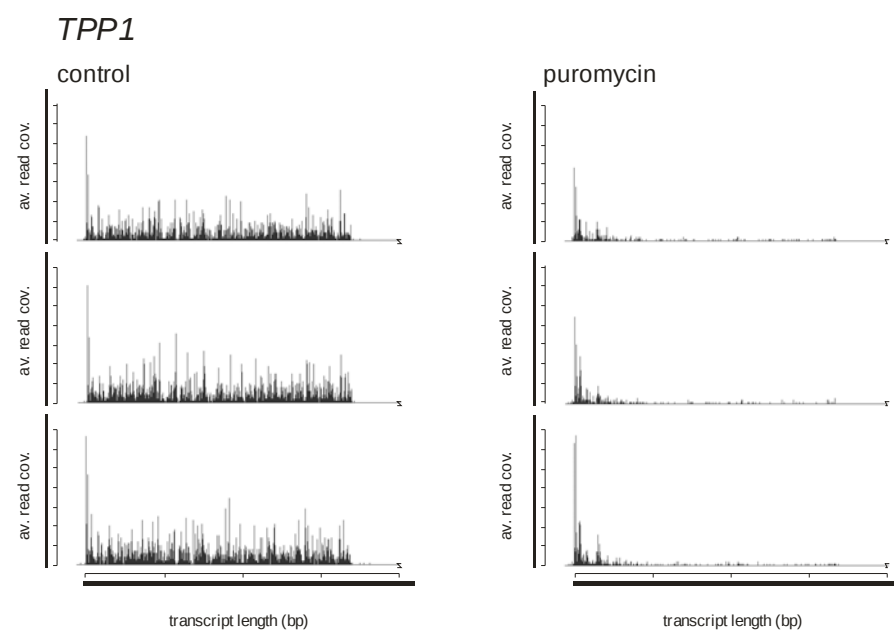
Translational initiation site enrichment in ribosome footprints by puromycin treatment. Panel A: Schematic illustration of experimental procedure. Panel B: Sucrose density analysis of extracts after cycloheximide and puromycin treatment, with loss of polysomal fractions indicating enrichment by translation-initiating ribosomes. Panel C: Length distribution of ribosome footprint sequences after excluding nuclease-generated rRNA contaminations.



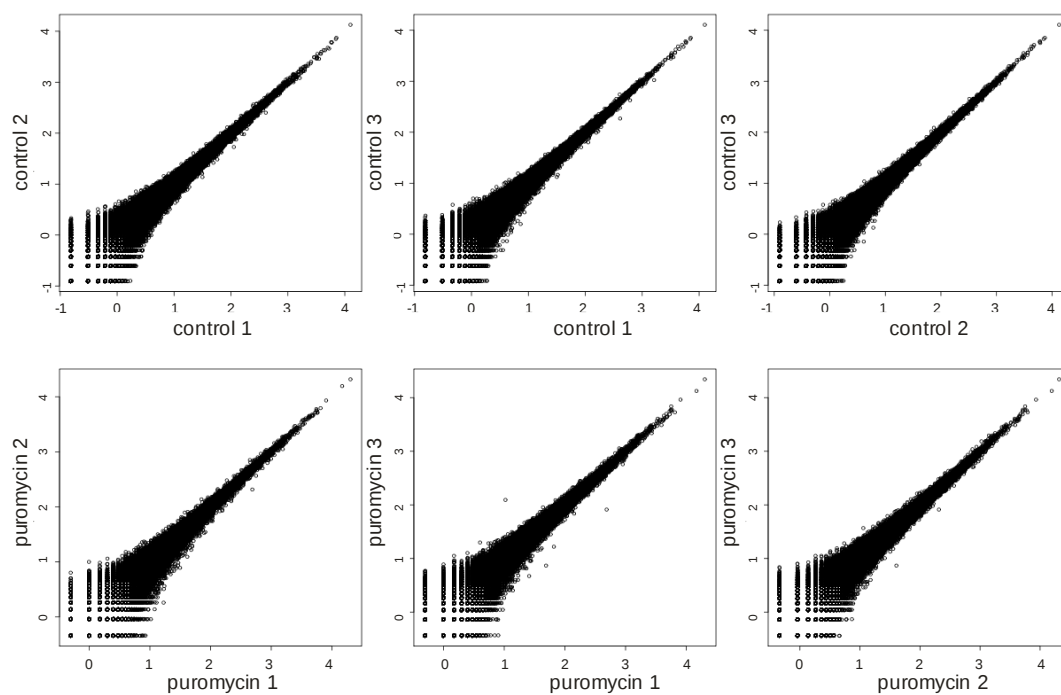
Supplementary Figure 2

Stability of ribosomal footprints with and without puromycin treatment (“control”). Panel A: Results of triplicate experiment on the *TPP1* gene. Panel B: Pair-wise correlation of the read coverage per nucleotide between triplicates.

A

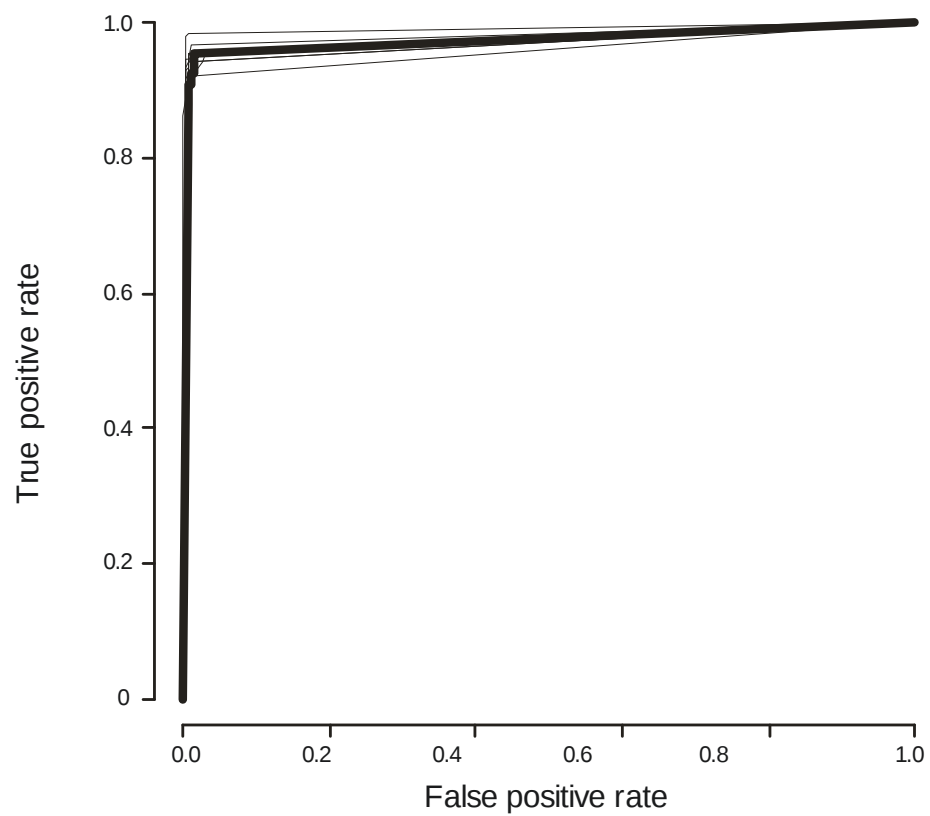


B



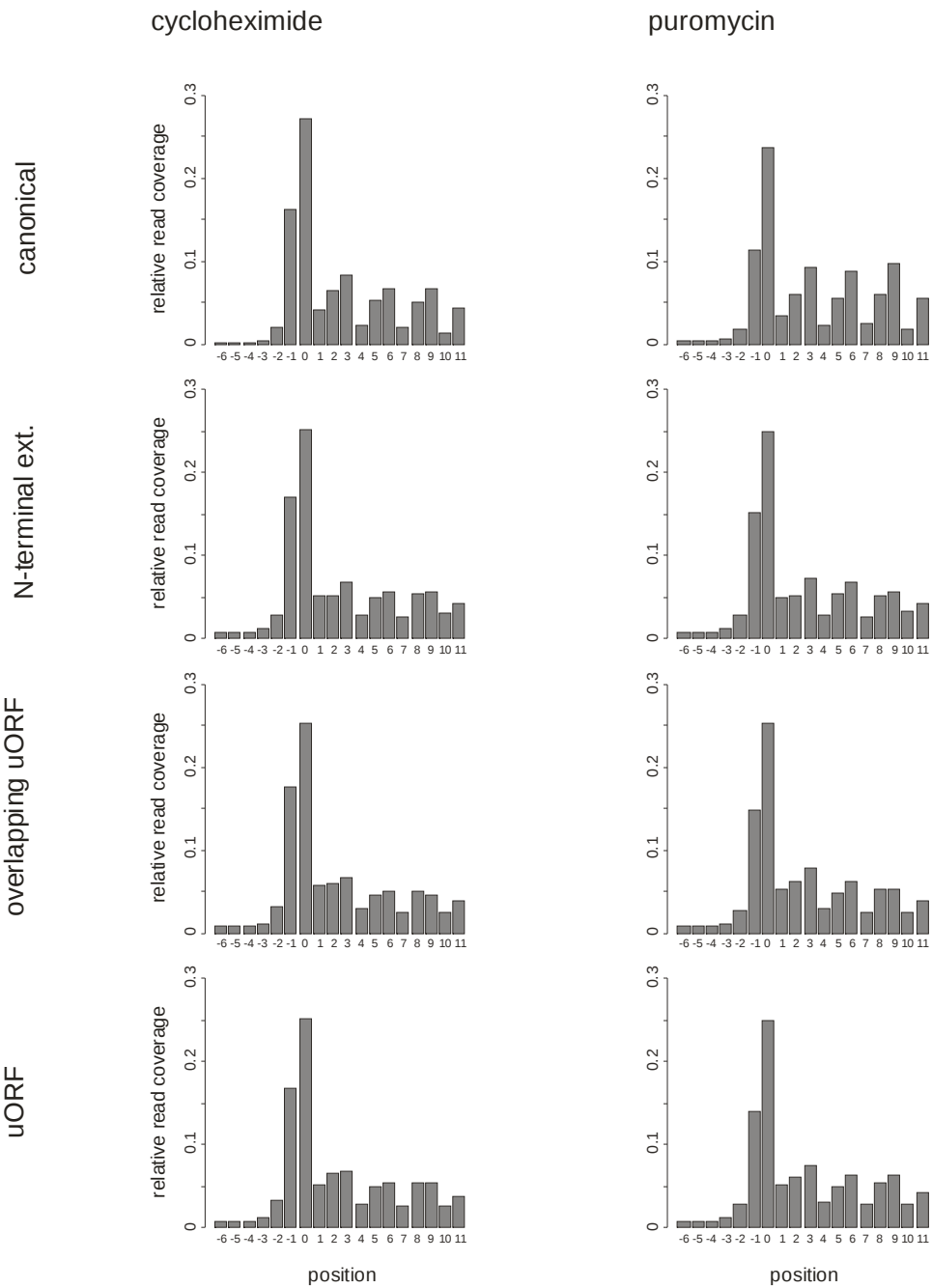
Supplementary Figure 3

Characterization of the used Neural Network. Panel A: ROC curves of ten neural networks trained on a manually curated dataset. The neural network with the median performance (indicated by a bold line) was used for all further analyses.



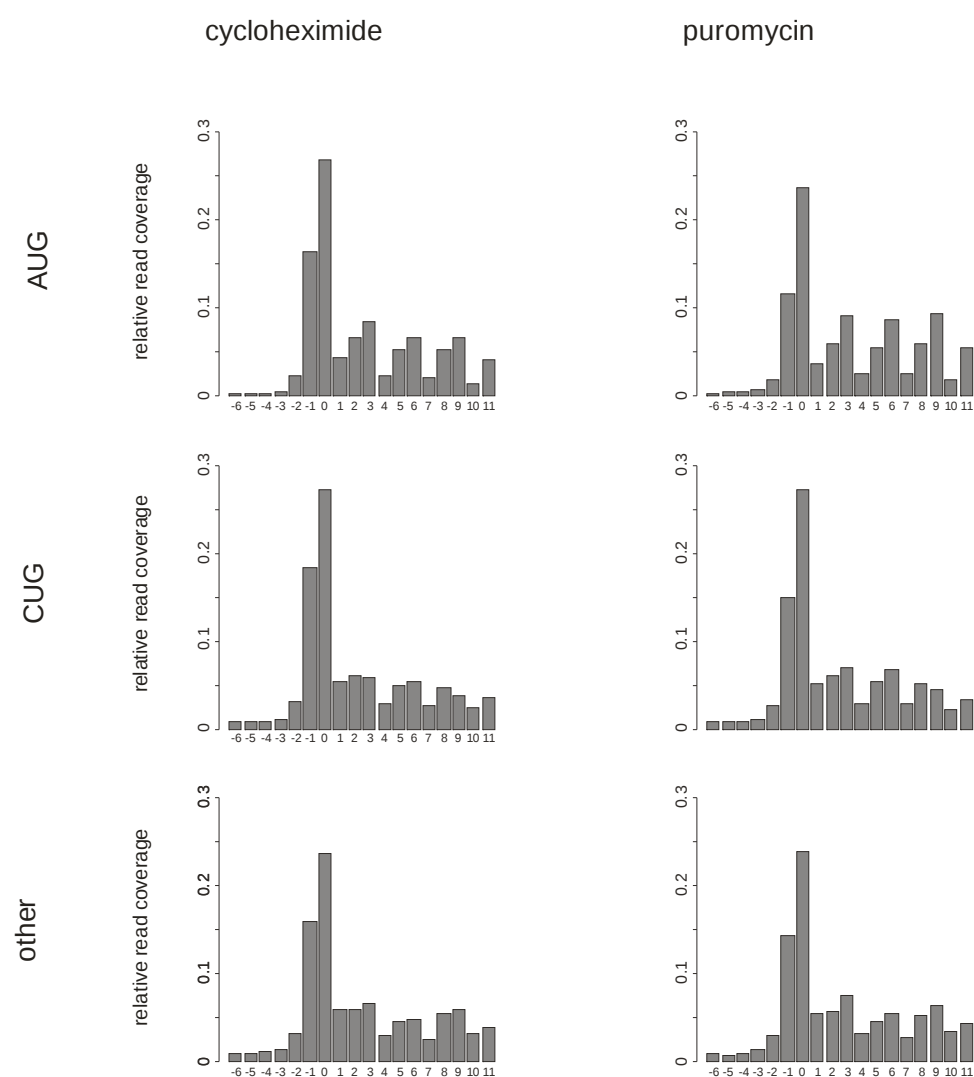
Supplementary Figure 4

Relative read coverage around the newly identified TISs.



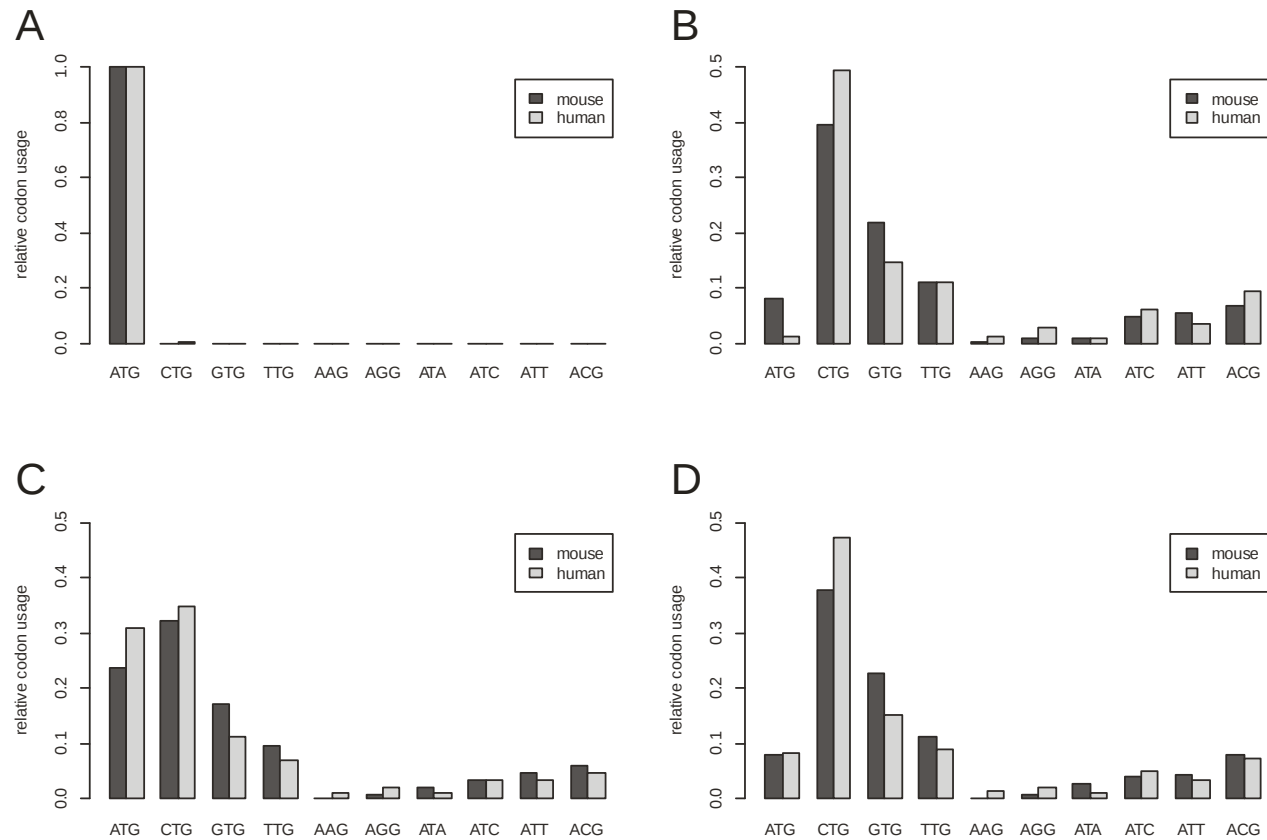
Supplementary Figure 5

Relative read coverage around newly identified TICs, according to type of embedded TIC. A higher read coverage of the second codon at TISs with an AUG TIC rather than a near-cognate TIC (ratio of peak at start codon versus second codon AUG: 3.2; CUG: 4.5; other: 3.6) may be indicative of a delayed initiation kinetic pertaining to the former type of TIS.



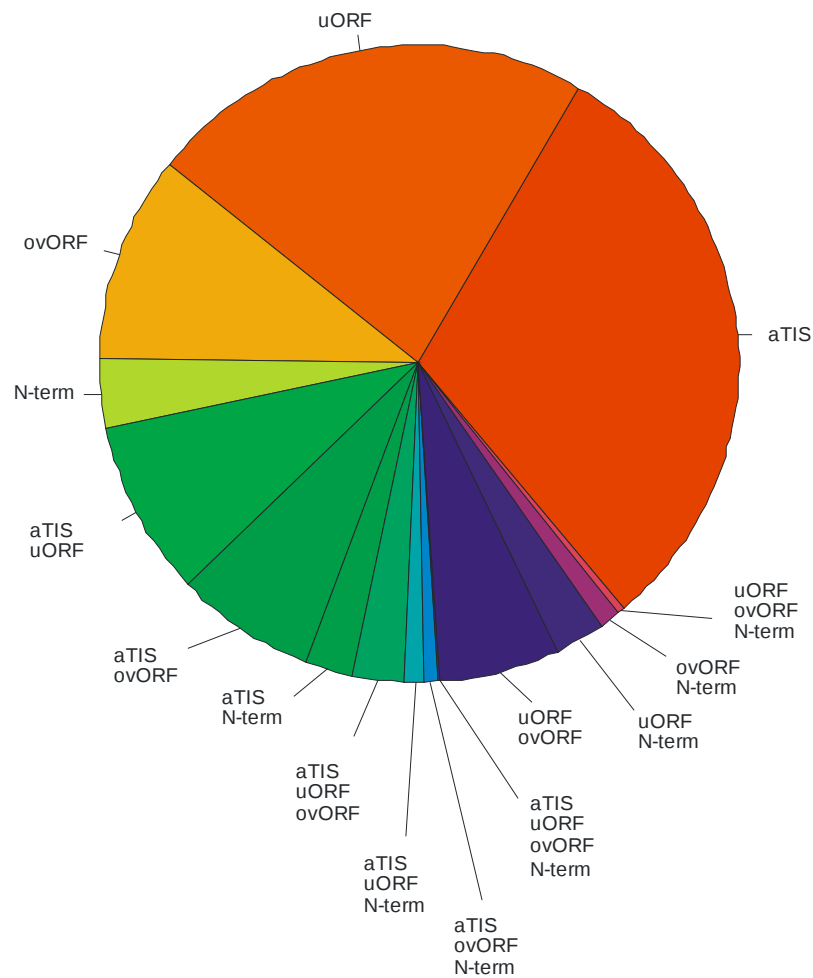
Supplementary Figure 6

Codon usage in human and mouse at canonical TISs (Panel A), N-terminal extensions (Panel B), uORFs (Panel C) and overlapping uORFs (Panel D). Mouse data were taken from Ingolia et al. 2011.



Supplementary Figure 7

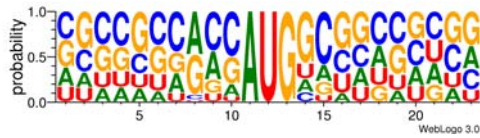
Distribution of transcripts according to the functional classification of neural network-predicted TIS (graphical representation of Table 3 of the main text). aTIS = annotated TIS; uORF = upstream open reading frame; ovORF = CDS-overlapping uORF; N-term = N-terminal protein extension.



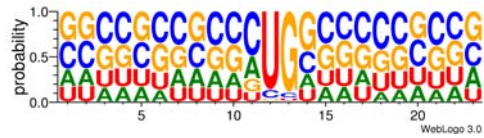
Supplementary Figure 8

Presence of a Kozak consensus sequence in newly identified TISs in different functional categories. The analysis was performed using the weblogo 3.0 software (Crooks et al. 2004).

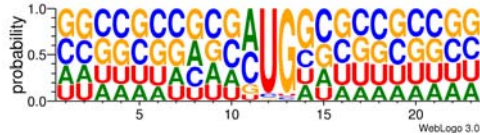
canonical



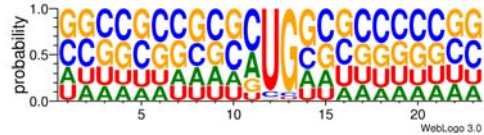
N-terminal extension



uORF



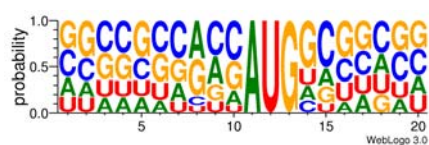
ovORF



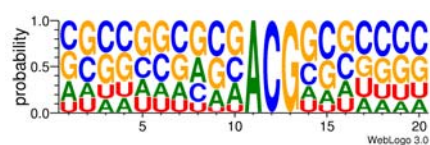
Supplementary Figure 9

Analysis of potential consensus sequences in the vicinity of the newly identified TISs in the by codon category. Sequences were plotted using weblogo 3.0 (Crooks et al. 2004). For the newly identified near-cognate codons, no consensus sequence is detectable.

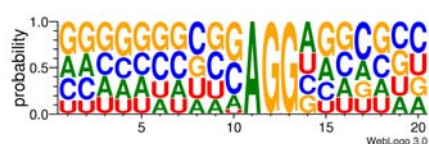
AUG



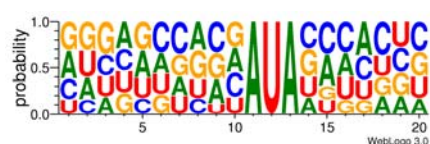
ACG



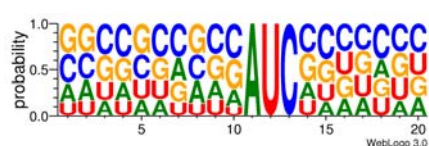
AGG



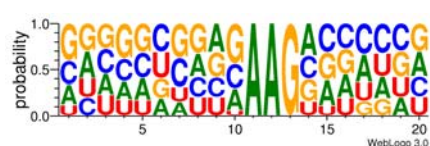
AUA



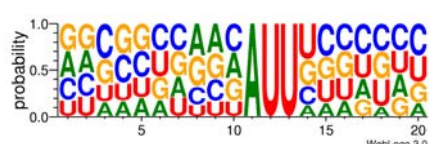
AUC



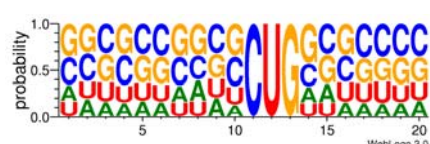
AAG



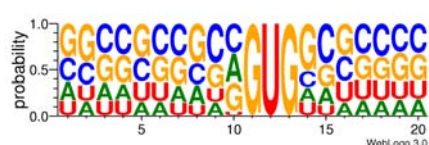
AUU



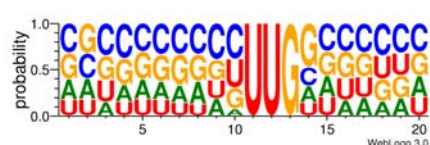
CUG



GUG



UUG



REFERENCES

- de Almeida RA, Heuser T, Blaschke R, Bartram CR, and Janssen JWG. 2006. Control of MYEOV protein synthesis by upstream open reading frames. *J Biol Chem* **281**: 695-704.
- Arnaud E, Touriol C, Boutonnet C, Gensac MC, Vagner S, Prats H, and Prats AC. 1999. A new 34-kilodalton isoform of human fibroblast growth factor 2 is cap dependently synthesized by using a non-AUG start codon and behaves as a survival factor. *Mol Cell Biol* **19**: 505-514.
- Ashida H, Hong Y, Murakami Y, Shishioh N, Sugimoto N, Kim YU, Maeda Y, and Kinoshita T. 2005. Mammalian PIG-X and yeast Pbn1p are the essential components of glycosylphosphatidylinositol-mannosyltransferase I. *Mol Biol Cell* **16**: 1439-1448.
- Bastide A, Karaa Z, Bornes Stéphanie, Hieblot C, Lacazette E, Prats Hervé, and Touriol Christian. 2008. An upstream open reading frame within an IRES controls expression of a specific VEGF-A isoform. *Nucleic Acids Res* **36**: 2434-2445.
- Calkhoven CF, Muller C, Martin R, Kros G, Pietsch H, Hoang T, and Leutz A. 2003. Translational control of SCL-isoform expression in hematopoietic lineage choice. *Genes Dev* **17**: 959-964.
- Calvo SE, Pagliarini DJ, and Mootha VK. 2009. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci* **106**: 7507-7512.
- Child SJ, Miller MK, and Geballe A P. 1999. Translational control by an upstream open reading frame in the HER-2/neu transcript. *J Biol Chem* **274**: 24335-24341.
- Crooks GE, Hon G, Chandonia J-M, and Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome research* **14**: 1188-90.
- Diba F, Watson CS, and Gametchu B. 2001. 5'UTR sequences of the glucocorticoid receptor 1A transcript encode a peptide associated with translational regulation of the glucocorticoid receptor. *J Cell Biochem* **81**: 149-161.
- Dodds DC, Omeis IA, Cushman SJ, Helms JA, and Perin MS. 1997. Neuronal pentraxin receptor, a novel putative integral membrane pentraxin that interacts with neuronal pentraxin 1 and 2 and taipoxin-associated calcium-binding protein 49. *J Biol Chem* **272**: 21488-21494.
- Dolstra H, Fredrix H, Maas F, Coulie PG, Brasseur F, Mensink E, Adema GJ, de Witte TM, Figdor CG, and van de Wiel-van Kemenade E. 1999. A human minor histocompatibility antigen specific for B cell acute lymphoblastic leukemia. *J Exp Med* **189**: 301-308.
- Garber M, Guttman M, Clamp M, Zody MC, Friedman N, and Xie X. 2009. Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics* **25**: 54-62.

- Gerashchenko MV, Su D, and Gladyshev VN. 2010. CUG start codon generates thioredoxin/glutathione reductase isoforms in mouse testes. *J Biol Chem* **285**: 4595-4602.
- Griffin E, Re A, Hamel N, Fu C, Bush H, McCaffrey T, and Asch AS. 2001. A link between diabetes and atherosclerosis: Glucose regulates expression of CD36 at the level of translation. *Nat Med* **7**: 840-846.
- Guyot B, Arnaud S, Phothisrath P, Rigal D, and Mouchiroud G. 2002. Upstream open reading frames regulate translation of Mona/Gads adapter mRNA in the megakaryocytic lineage. *Platelets* **13**: 459-464.
- Göpfert U, Kullmann M, and Hengst L. 2003. Cell cycle-dependent translation of p27 involves a responsive element in its 5'-UTR that overlaps with a uORF. *Hum Mol Genet* **12**: 1767-1779.
- Hernandez EM, Johnson A, Notario V, Chen A, and Richert JR. 2002. AUA as a translation initiation site in vitro for the human transcription factor Sp3. *J Biochem Mol Biol* **35**: 273-282.
- Huez I, Bornes S, Bresson D, Créancier L, and Prats H. 2001. New vascular endothelial growth factor isoform generated by internal ribosome entry site-driven CUG translation initiation. *Mol Endocrinol* **15**: 2197-2210.
- Ingolia NT, Lareau LF, and Weissman JS. 2011. Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell* **147**: 789-802.
- Ivanov IP, Firth AE, Michel AM, Atkins JF, and Baranov PV. 2011. Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic Acids Res* **39**: 4220-4234.
- Jacquemin P, Martial JA, and Davidson I. 1997. Human TEF-5 is preferentially expressed in placenta and binds to multiple functional elements of the human chorionic somatomammotropin-B gene enhancer. *J Biol Chem* **272**: 12928-12937.
- Kos M, Denger S, Reid G, and Gannon F. 2002. Upstream open reading frames regulate the translation of the multiple mRNA variants of the estrogen receptor alpha. *J Biol Chem* **277**: 37131-37138.
- Lehmann JM, McKee DD, Watson MA, Willson TM, Moore JT, and Kliewer SA. 1998. The human orphan nuclear receptor PXR is activated by compounds that regulate CYP3A4 gene expression and cause drug interactions. *J Clin Invest* **102**: 1016-1023.
- Lincoln AJ, Monczak Y, Williams SC, and Johnson PF. 1998. Inhibition of CCAAT/enhancer-binding protein alpha and beta translation by upstream open reading frames. *J Biol Chem* **273**: 9552-9560.
- Lodhi KM, Ozdener MH, and Shayiq RM. 2003. The upstream open reading frame mediates constitutive effects on translation of cytochrome p-450c27 from the seventh in-frame AUG codon in rat liver. *J Biol Chem* **278**: 40647-40657.

- Lux H, Flammann H, Hafner M, and Lux A. 2010. Genetic and molecular analyses of PEG10 reveal new aspects of genomic organization, transcription and translation. *PLoS one* **5**: e8686.
- Martin MM, Buckenberger JA, Knoell DL, Strauch AR, and Elton TS. 2006. TGF-beta(1) regulation of human AT(1) receptor mRNA splice variants harboring exon 2. *Mol Cell Endocrinol* **249**: 21-31.
- McDade SS, Hall PA, and Russell SEH. 2007. Translational control of SEPT9 isoforms is perturbed in disease. *Hum Mol Genet* **16**: 742-752.
- Mehta A, Trotta CR, and Peltz SW. 2006. Derepression of the Her-2 uORF is mediated by a novel post-transcriptional control mechanism in cancer cells. *Genes Dev* **20**: 939-953.
- Meiron M, Anunu R, Scheinman EJ, Hashmueli S, and Levi BZ. 2001. New isoforms of VEGF are translated from alternative initiation CUG codons located in its 5'UTR. *Biochem Biophys Res Commun* **282**: 1053-1060.
- Nomura A, Iwasaki Y, Saito M, Aoki Y, Yamamori E, Ozaki N, Tachikawa K, Mutsuga N, Morishita M, Yoshida M, et al. 2001. Involvement of upstream open reading frames in regulation of rat V(1b) vasopressin receptor expression. *Am J Physiol Endocrinol Metab* **280**: E780-787.
- Park E-H, Lee JM, and Pelletier J. 2006. The Tie2 5' untranslated region is inhibitory to 5' end-mediated translation initiation. *FEBS letters* **580**: 1309-1319.
- Pendleton LC, Goodwin BL, Solomonson LP, and Eichler DC. 2005. Regulation of endothelial argininosuccinate synthase expression and NO production by an upstream open reading frame. *J Biol Chem* **280**: 24252-24260.
- Pentecost BT, Song R, Luo M, DePasquale JA, and Fasco MJ. 2005. Upstream regions of the estrogen receptor alpha proximal promoter transcript regulate ER protein expression through a translational mechanism. *Mol Cell Endocrinol* **229**: 83-94.
- Prats H, Kaghad M, Prats AC, Klagsbrun M, Lélías JM, Liauzun P, Chalon P, Tauber JP, Amalric F, and Smith JA. 1989. High molecular mass forms of basic fibroblast growth factor are initiated by alternative CUG codons. *Proc Natl Acad Sci* **86**: 1836-1840.
- Rabadan-Diehl C, Martínez A, Volpi S, Subburaju S, and Aguilera G. 2007. Inhibition of vasopressin V1b receptor translation by upstream open reading frames in the 5'-untranslated region. *J Neuroendocrinol* **19**: 309-319.
- Reynolds K, Zimmer AM, and Zimmer A. 1996. Regulation of RAR beta 2 mRNA expression: evidence for an inhibitory peptide encoded in the 5'-untranslated region. *J Cell Biol* **134**: 827-835.
- Rogers GW, Edelman GM, and Mauro VP. 2004. Differential utilization of upstream AUGs in the beta-secretase mRNA suggests that a shunting mechanism regulates translation. *Proc Natl Acad Sci* **101**: 2794-2799.

- Rosenstiel P, Huse K, Franke A, Hampe J, Reichwald K, Platzer C, Roberts RG, Mathew CG, Platzer M, and Schreiber S. 2007. Functional characterization of two novel 5' untranslated exons reveals a complex regulation of NOD2 protein expression. *BMC genomics* **8**.
- Schlüter G, Boinska D, and Nieman-Seyde SC. 2000. Evidence for translational repression of the SOCS-1 major open reading frame by an upstream open reading frame. *Biochem Bioph Res Co* **268**: 255-261.
- Shaughnessy JD, Largaespada DA, Tian E, Fletcher CF, Cho BC, Vyas P, Jenkins NA, and Copeland NG. 1999. Mrvi1, a common MRV integration site in BXH2 myeloid leukemias, encodes a protein with homology to a lymphoid-restricted membrane protein Jaw1. *Oncogene* **18**: 2069-2084.
- Song KY, Hwang CK, Kim CS, Choi HS, Law P-Y, Wei L-N, and Loh HH. 2007. Translational repression of mouse mu opioid receptor expression via leaky scanning. *Nucleic Acids Res* **35**: 1501-1513.
- Spevak CC, Park E-H, Geballe Adam P, Pelletier J, and Sachs MS. 2006. her-2 upstream open reading frame effects on the use of downstream initiation codons. *Biochem Bioph Res Co* **350**: 834-841.
- Steel LF, Telly DL, Leonard J, Rice BA, Monks B, and Sawicki JA. 1996. Elements in the murine c-mos messenger RNA 5'-untranslated region repress translation of downstream coding sequences. *Cell Death Differ* **7**: 1415-1424.
- Stockklausner C, Breit S, Neu-Yilik G, Echner N, Hentze MW, Kulozik AE, and Gehring NH. 2006. The uORF-containing thrombopoietin mRNA escapes nonsense-mediated decay (NMD). *Nucleic Acids Res* **34**: 2355-2363.
- Taira M, Iizasa T, Shimada H, Kudoh J, Shimizu N, and Tatibana M. 1990. A human testis-specific mRNA for phosphoribosylpyrophosphate synthetase that initiates from a non-AUG codon. *J Biol Chem* **265**: 16491-16497.
- Takeda M, Obayashi K, Kobayashi A, and Maeda M. 2004. A unique role of an amino terminal 16-residue region of long-type GATA-6. *J Biochem* **135**: 639-650.
- Tanaka M, Ito S, and Kiuchi K. 2001. The 5'-untranslated region of the mouse glial cell line-derived neurotrophic factor gene regulates expression at both the transcriptional and translational levels. *Brain res* **91**: 81-95.
- Tang T, Rector K, Barnett CD, and Mao CD. 2008. Upstream open reading frames regulate the expression of the nuclear Wnt13 isoforms. *Biochem Bioph Res Co* **366**: 1081-1088.
- Tee MK, and Jaffe RB. 2001. A precursor form of vascular endothelial growth factor arises by initiation from an upstream in-frame CUG codon. *Biochem J* **359**: 219-226.
- Thébault S, Gachon F, Lemasson I, Devaux C, and Mesnard JM. 2000. Molecular cloning of a novel human I-mfa domain-containing protein that differently regulates human T-cell leukemia virus type I and HIV-1 expression. *J Biol Chem* **275**: 4848-4857.

- Wang X, Zhao Y, Xiao Z, Chen B, Wei Z, Wang B, Zhang J, Han J, Gao Y, Li L, et al. 2009. Alternative translation of OCT4 by an internal ribosome entry site and its novel function in stress response. *Stem Cells* **27**: 1265-1275.
- Warnakulasuriyarachchi D, Ungureanu NH, and Holcík M. 2003. The translation of an antiapoptotic protein HIAP2 is regulated by an upstream open reading frame. *Cell Death Differ* **10**: 899-904.
- Williams RT, Manji SS, Parker NJ, Hancock MS, Van Stekelenburg L, Eid JP, Senior PV, Kazenwadel JS, Shandala T, Saint R, et al. 2001. Identification and characterization of the STIM (stromal interaction molecule) gene family: coding for a novel class of transmembrane proteins. *Biochem J* **357**: 673-685.
- Wu H, Ross CR, and Blecha F. 2002. Characterization of an upstream open reading frame in the 5' untranslated region of PR-39, a cathelicidin antimicrobial peptide. *Mol Immunol* **39**: 9-18.
- Xiao JH, Davidson I, Matthes H, Garnier JM, and Chambon P. 1991. Cloning, expression, and transcriptional properties of the human enhancer factor TEF-1. *Cell* **65**: 551-568.
- Xiong Z, Liu E, Yan Y, Silver RT, Yang F, Chen IH, Chen Y, Verstovsek S, Wang H, Prchal J, et al. 2006. An unconventional antigen translated by a novel internal ribosome entry site elicits antitumor humoral immune reactions. *J Immunol* **177**: 4907-4916.
- Zhang Y, Li W, and Vore M. 2007. Translational regulation of rat multidrug resistance-associated protein 2 expression is mediated by upstream open reading frames in the 5' untranslated region. *Mol Pharmacol* **71**: 377-383.
- Zysow BR, Lindahl GE, Wade DP, Knight BL, and Lawn RM. 1995. C/T polymorphism in the 5' untranslated region of the apolipoprotein(a) gene introduces an upstream ATG and reduces in vitro translation. *Arterioscl Throm Vas* **15**: 58-64.