Supplementary materials for

“Mammalian alternative translation initiation is mostly nonadaptive”

Chuan Xu & Jianzhi Zhang (jianzhi@umich.edu)

Supplementary materials include:
Legends of supplementary figures
Figures S1-S11
Table S1

Legends of supplementary figures

Fig. S1. Correlations between TIS diversity and translational amount among simulated genes for HEK293 cells. ATLI and translational amount are uncorrelated in the generation of the simulated genes. (**A**) Positive correlation between the Simpson index of TIS diversity and translational amount among the simulated genes. (**B**) Positive correlation between the Shannon index of TIS diversity and translational amount among the simulated genes. In (A) and (B), each dot represents a simulated gene. (**C**) Simpson/Shannon index of TIS diversity of a supergene made up of simulated genes rises with the median translational amount of the simulated genes belonging to the supergene. Each triangle/circle represents a supergene. Spearman's rank correlation coefficient (ρ) and associated *P* value are presented.

Fig. S2. Results in Fig. 1 are robust to variation of the analysis. The supergene approach is used here. (**A**) Spearman's correlation between translational amount and TIS diversity among supergenes when only TISs with *Q* value < 0.05 of each gene are considered (see Materials and methods). $P < 0.05$ in all cases except for samples 6 and 7. (**B**) Spearman's correlation between translational amount and TIS diversity when only AUG is considered the start codon. $P < 0.05$ in all cases. (**C**) Spearman's correlation between gene expression level measured by RNA-seq and TIS diversity. $P < 0.01$ for samples 6, 8, and 9; others are not significant.

Fig. S3. Results in Fig. 1 remain qualitatively unchanged upon the control of the transcript length. (**A**) Spearman's correlations between translational amount and Simpson index of TIS diversity among genes in each of five human and four mouse samples. (**B**) Spearman's correlations between translational amount and Shannon index of TIS diversity among genes in each of five human and four mouse samples. Squares, triangles, and circles show the correlations on the basis of the (i) original data, (ii) down-sampled data, and (iii) down-sampled, transcript length-controlled data, respectively. $P < 0.05$ for all correlations calculated from the down-sampled data and down-sampled, length-controlled data, except for sample #7 of down-sampled, length-controlled data. Sample IDs listed on the x-axis refer to those in Table S1.

Fig. S4. The relatively highly translated paralog tends to have lower TIS diversity than the relatively lowly translated one within a pair of human paralogous genes. Results from

downsampled data are shown. (**A**) Simpson index of TIS diversity for each member of a paralogous gene pair in human HEK293 cells. Each dot represents a paralogous gene pair. Dots above and below the diagonal are colored blue and red, respectively. Numbers of blue and red dots are respectively indicated in the corresponding color. *P*-value is from a binomial test of the null hypothesis of equal numbers of red and blue dots. Some dots overlap one another because the genes have the same fractional uses of TISs in downsampled data. (**B-C**) Fraction of paralogous gene pairs for which the Simpson or Shannon index of TIS diversity of the relatively lowly translated gene exceeds that of the relatively highly translated gene. Panels (A) and (B) are based on down-sampled data of 10 reads per gene whereas panel (C) is based on down-sampled data of 20 reads per gene. Statistical significance is shown by stars/dashes. Sample IDs listed on the X-axis of (B) and (C) refer to those in Table S1.

Fig. S5. Fractional use of each ranked TIS when we consider only TISs with AUG as the start codon. Results from supergene analysis are shown. (**A**) Fractional use of the major TIS increases, but that of each minor TIS examined decreases with translational amount in human HEK293 cells. Each circle represents a supergene. (**B**) Spearman's correlation between fractional use of a ranked TIS and translational amount. Sample IDs listed on the X-axis of (B) refer to those in Table S1. $P < 0.05$ in all cases.

Fig. S6. Fractional use of each minor TIS (rank #5 to #8) decreases with translational amount. (**A**) Fractional use of each minor TISs in the gene decreases with the translational amount of the gene in human HEK293 cells. Each dot represents a gene. (**B**) Fractional use of each minor TIS in a supergene decreases with the median translational amount of all genes belonging to the supergene in HEK293. Each dot represents a supergene. (**C**) Spearman's correlation between the median translational amount of genes belonging to a supergene and the fractional use of each minor TIS. Sample IDs listed on the X-axis of (C) refer to those in Table S1. $P < 0.05$ in all cases except ranks #6 to #8 of sample 7.

Fig. S7. Correlation between total percent use of upstream minor TISs and translational amount or mRNA concentration. Results from supergene analysis are presented. (**A-C**) Correlation between total percent use of upstream minor TISs and translational amount (circle) or mRNA

concentration (estimated from RNA-seq; triangle) when AUG and nine near-cognate codons are considered as start codons (A), when only AUG is considered as the start codon (B), or when the annotated TIS in a gene is considered the main TIS (C). Sample IDs listed on the X-axis refer to those in Table S1. Statistical significance is shown by stars ($P < 0.05$) or dashes ($P \geq 0.05$), where color black and red correspond to results for circles and triangles, respectively. (**D**) Total fractional use of upstream minor TISs decreases with translational amount (when AUG and nine near-cognate codons are considered as start codons) in human HEK293 cells. Each circle represents a supergene.

Fig. S8. Fractional uses of ranked TISs in human paralogous genes. Results from downsampling analysis are shown. (**A**) Fractional uses of TISs of ranks #1 to #4 in the relatively lowly translated and relatively highly translated members of each paralogous gene pair in human HEK293 cells. Each dot represents a paralogous gene pair. Dots above and below the diagonal are colored blue and red, respectively. Numbers of blue and red dots are respectively indicated. *P*-value shown is based on a binomial test of the null hypothesis of equal numbers of blue and red dots. Some dots overlap one another because the genes have the same fractional uses of TISs in downsampled data. (**B**) Proportion of paralogous gene pairs for which the fractional usage of a ranked TIS is higher in the more highly translated paralog. Statistical significance is shown by stars/dashes. Sample IDs listed on the X-axis of (B) refer to those in Table S1.

Fig. S9. Both the number (**A**) and amount of translation initiation (**B**) of out-of-frame minor TISs are much less than those of in-frame minor TISs.

Fig. S10. Rank correlation between the translational amount of a gene and the mean Kozak strength of minor TISs relative to the Kozak strength of the major TIS. Statistical significance is shown by stars/dashes. Sample IDs listed on the X-axis refer to those in Table S1.

Fig. S11. The observed number of upstream AUGs (uAUGs) is lower than expected by chance regardless of the translational amount of the gene. Red triangles and black circles donate the mean number of observed and expected uAUGs per gene in each bin, respectively. Genes are ranked and binned according to their translational amounts in HEK293 cells as in Fig. 1C. The

observed number of uAUGs in an annotated gene is the mean number of uAUGs in the 5'UTR regions of all transcripts of the gene, whereas the chance expectation is the mean number of uAUGs in the randomized nucleotide sequences in the same 5'UTR regions.

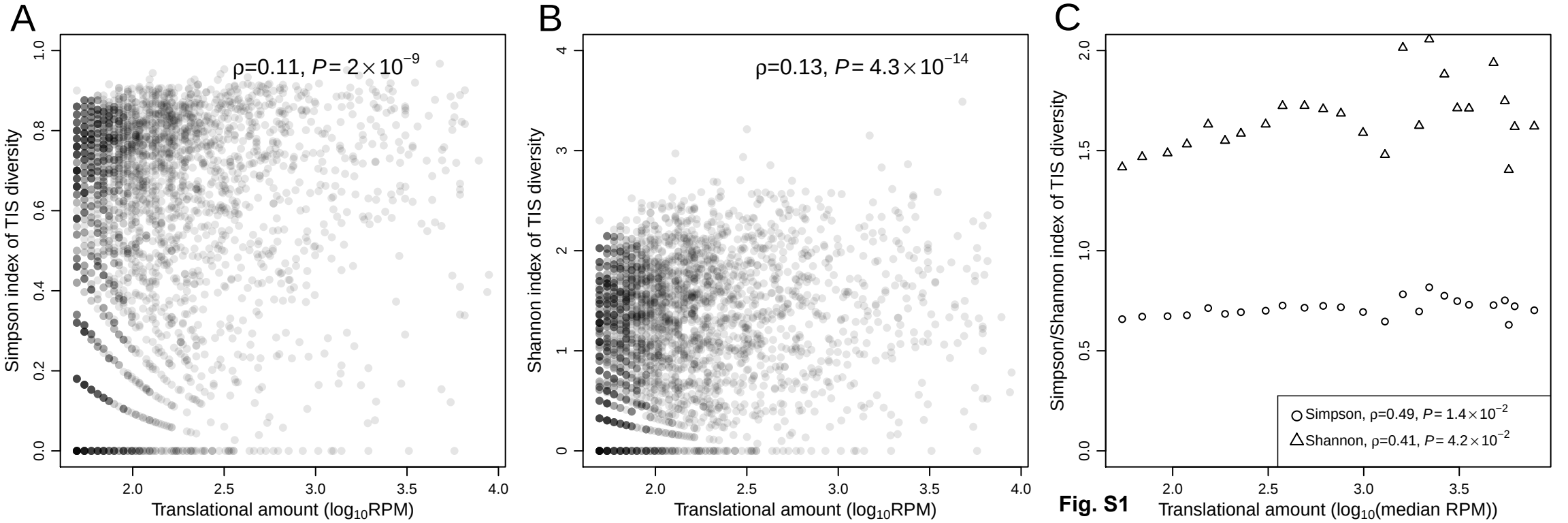
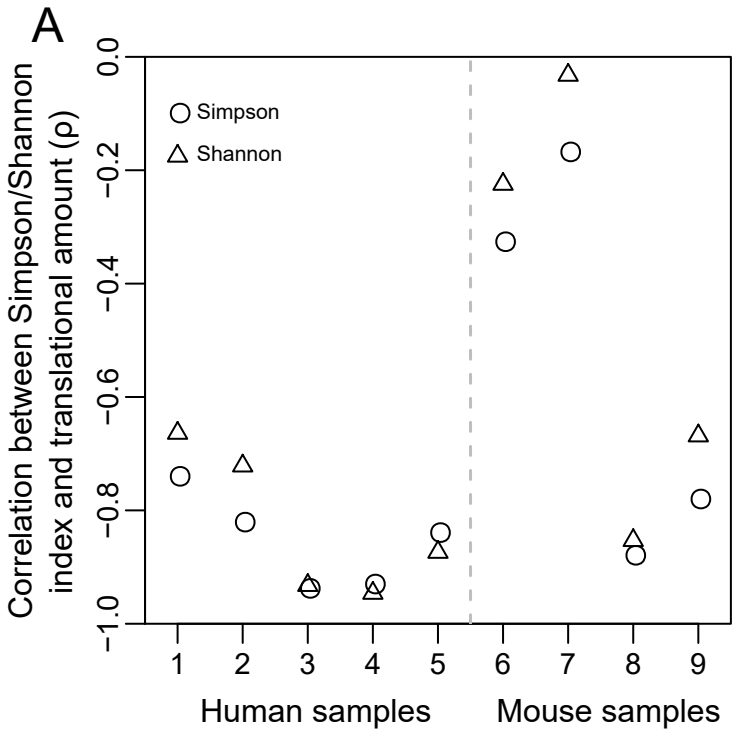
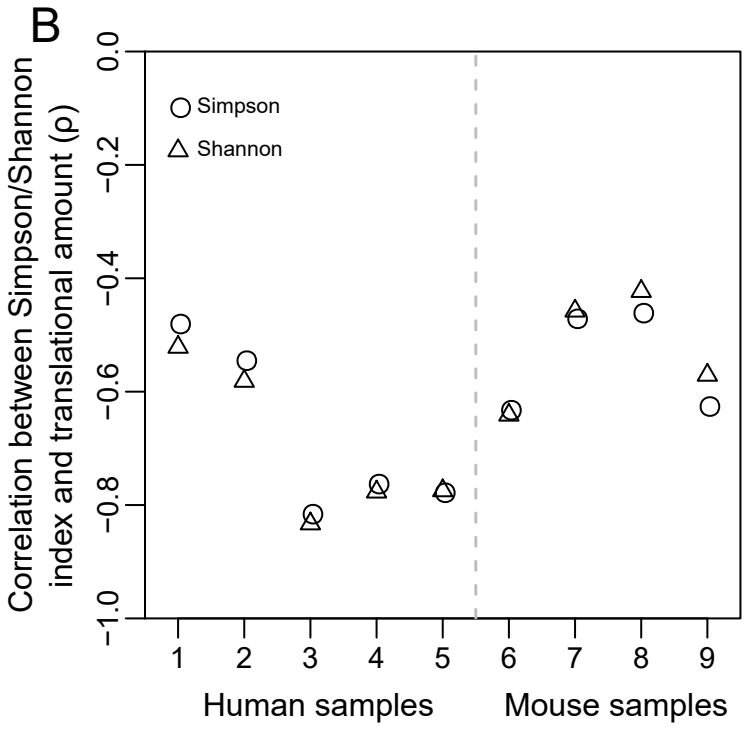

A
ρ=0.11, $P = 2 \times 10^{-9}$
Simpson index of TIS diversity
Translational amount ($\log_{10}$RPM)
B
ρ=0.13, $P = 4.3 \times 10^{-14}$
Shannon index of TIS diversity
Translational amount ($\log_{10}$RPM)
C
Simpson/Shannon index of TIS diversity
Simpson, ρ=0.49, $P = 1.4 \times 10^{-2}$
Shannon, ρ=0.41, $P = 4.2 \times 10^{-2}$
Translational amount ($\log_{10}$(median RPM))

Fig. S1


A
Correlation between Simpson/Shannon index and translational amount (ρ)
Simpson
Shannon
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
1
2
3
4
5
6
7
8
9
Human samples
Mouse samples


B
Correlation between Simpson/Shannon index and translational amount (ρ)
Simpson
Shannon
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
1
2
3
4
5
6
7
8
9
Human samples
Mouse samples

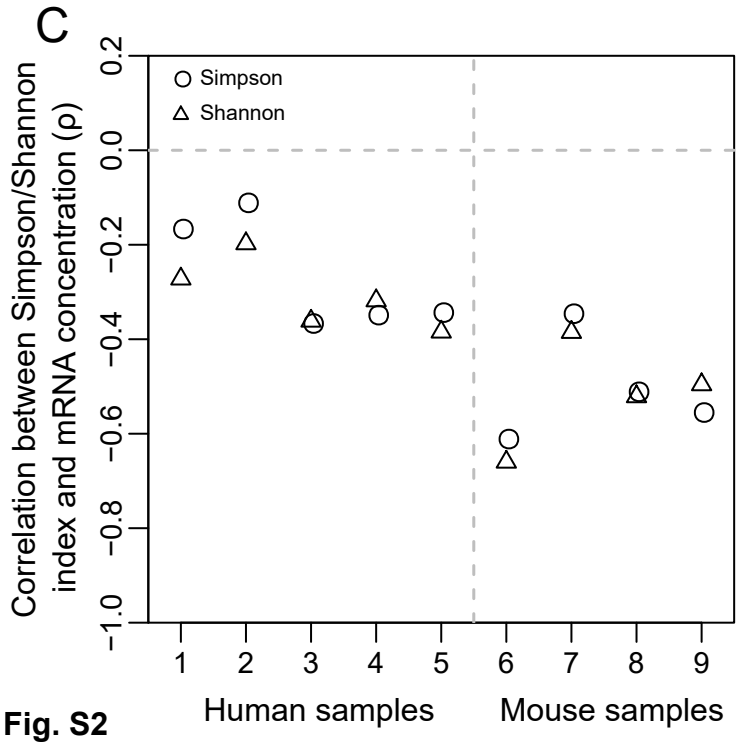

C
Correlation between Simpson/Shannon index and mRNA concentration (ρ)
Simpson
Shannon
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
1
2
3
4
5
6
7
8
9
Human samples
Mouse samples

Fig. S2

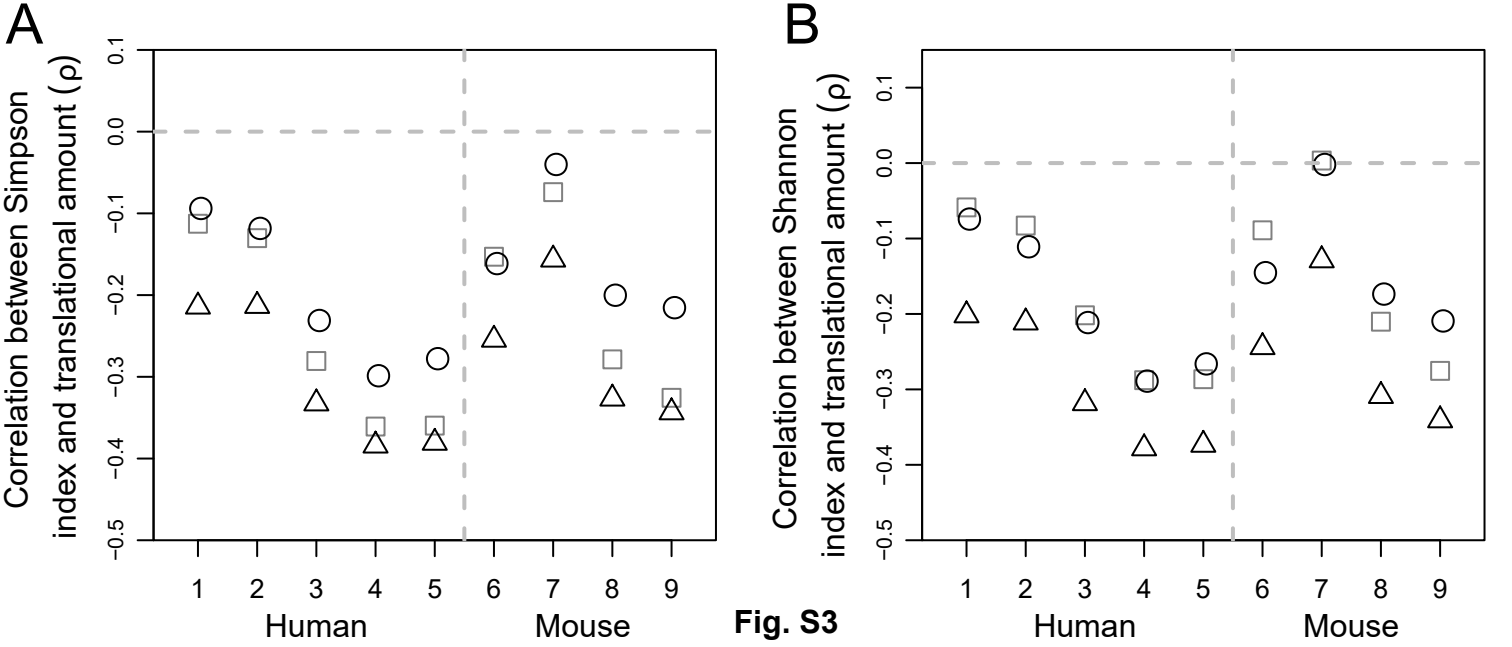

A
Correlation between Simpson index and translational amount (ρ)
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
1
2
3
4
5
6
7
8
9
Human
Mouse
B
Correlation between Shannon index and translational amount (ρ)
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
1
2
3
4
5
6
7
8
9
Human
Mouse

Fig. S3

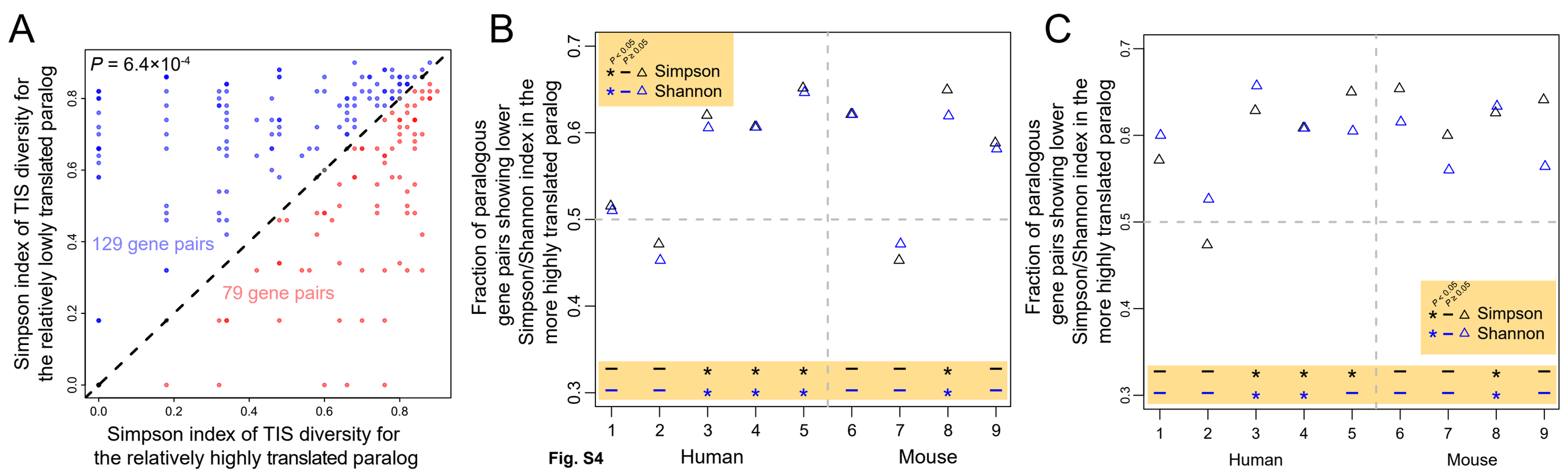

A
P = 6.4×10-4
129 gene pairs
79 gene pairs
Simpson index of TIS diversity for the relatively lowly translated paralog
Simpson index of TIS diversity for the relatively highly translated paralog
0.0
0.2
0.4
0.6
0.8
B
C
Fraction of paralogous gene pairs showing lower Simpson/Shannon index in the more highly translated paralog
0.3
0.4
0.5
0.6
0.7
P < 0.05
P ≥ 0.05
Simpson
Shannon
1
2
3
4
5
6
7
8
9
Human
Mouse

Fig. S4

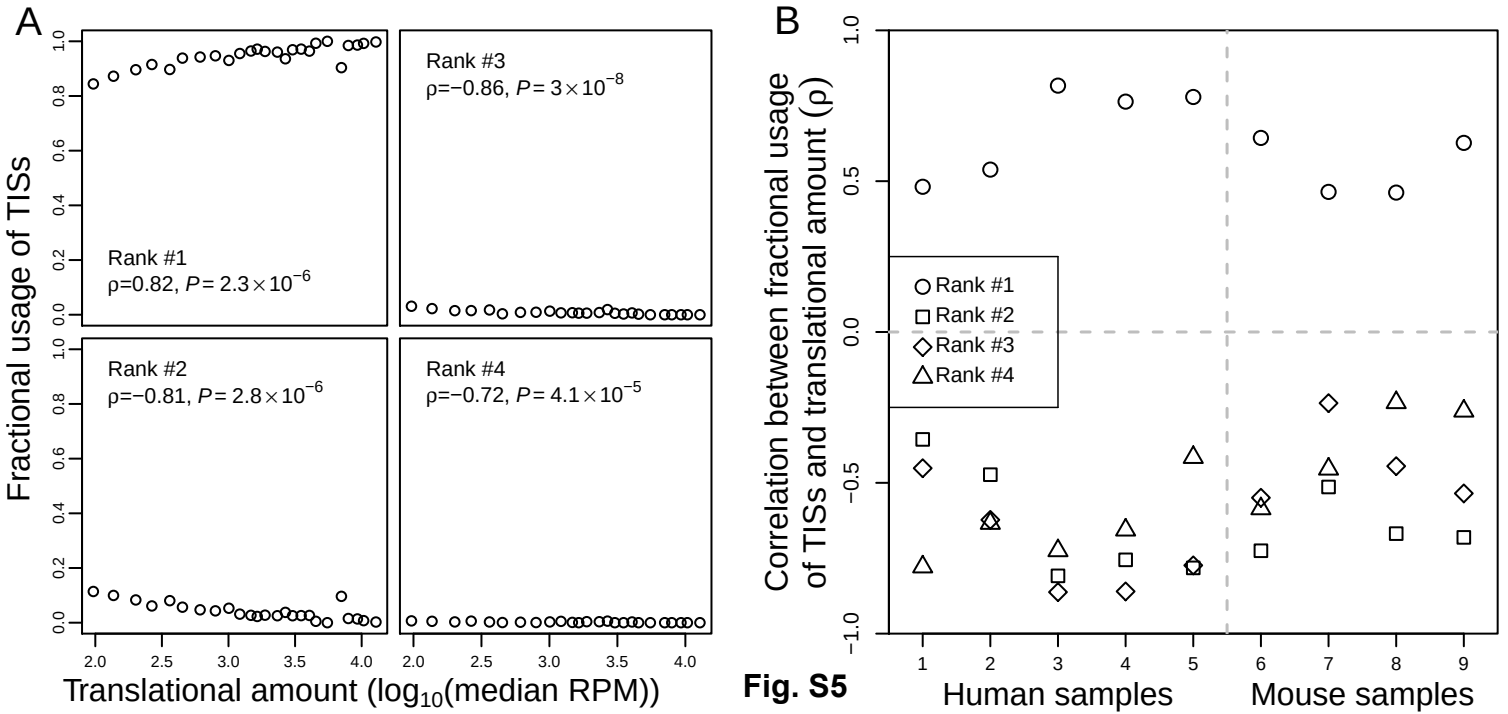
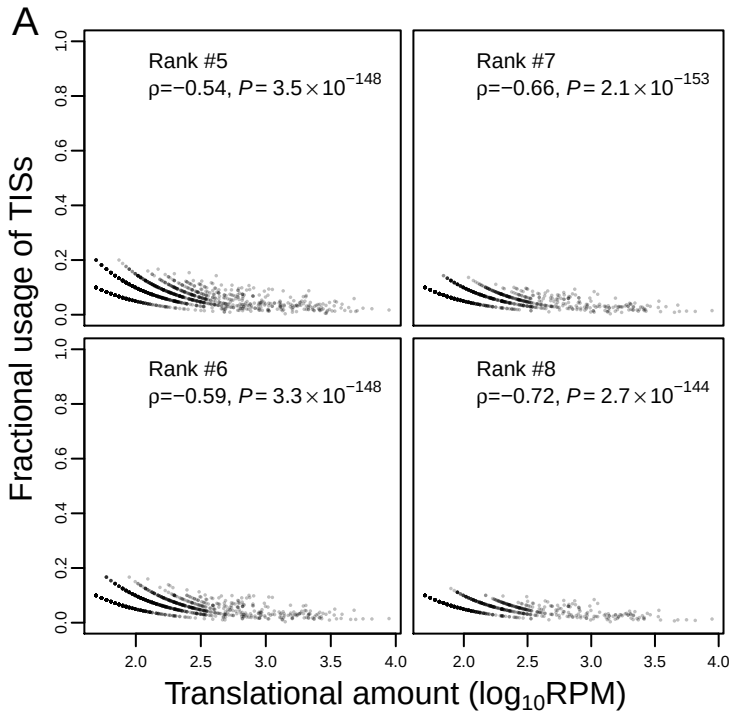
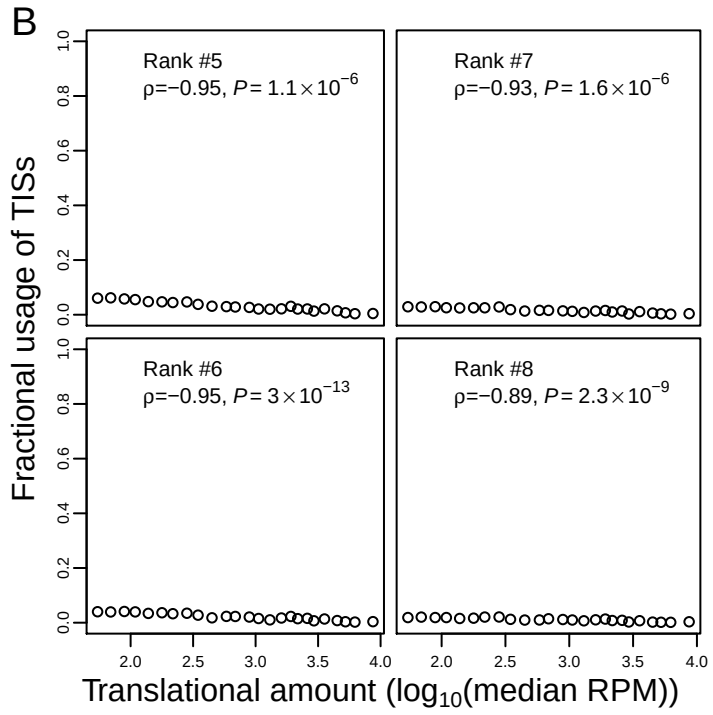

A
Fractional usage of TISs
Translational amount (log10(median RPM))
Rank #1
ρ=0.82, P= 2.3×10−6
Rank #2
ρ=−0.81, P= 2.8×10−6
Rank #3
ρ=−0.86, P= 3×10−8
Rank #4
ρ=−0.72, P= 4.1×10−5
B
Correlation between fractional usage of TISs and translational amount (ρ)
Rank #1
Rank #2
Rank #3
Rank #4
Human samples
Mouse samples

Fig. S5


A
Rank #5
ρ=−0.54, P = 3.5×10−148
Rank #7
ρ=−0.66, P = 2.1×10−153
Rank #6
ρ=−0.59, P = 3.3×10−148
Rank #8
ρ=−0.72, P = 2.7×10−144
Fractional usage of TISs
Translational amount (log10RPM)


B
Rank #5
ρ=−0.95, P = 1.1×10−6
Rank #7
ρ=−0.93, P = 1.6×10−6
Rank #6
ρ=−0.95, P = 3×10−13
Rank #8
ρ=−0.89, P = 2.3×10−9
Fractional usage of TISs
Translational amount (log10(median RPM))

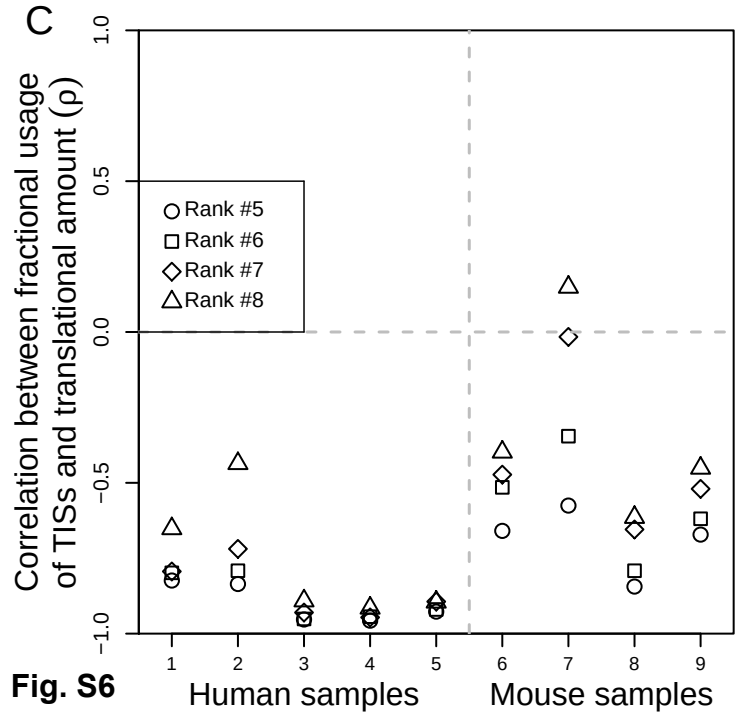

C
Rank #5
Rank #6
Rank #7
Rank #8
Correlation between fractional usage of TISs and translational amount (ρ)
Human samples
Mouse samples

Fig. S6

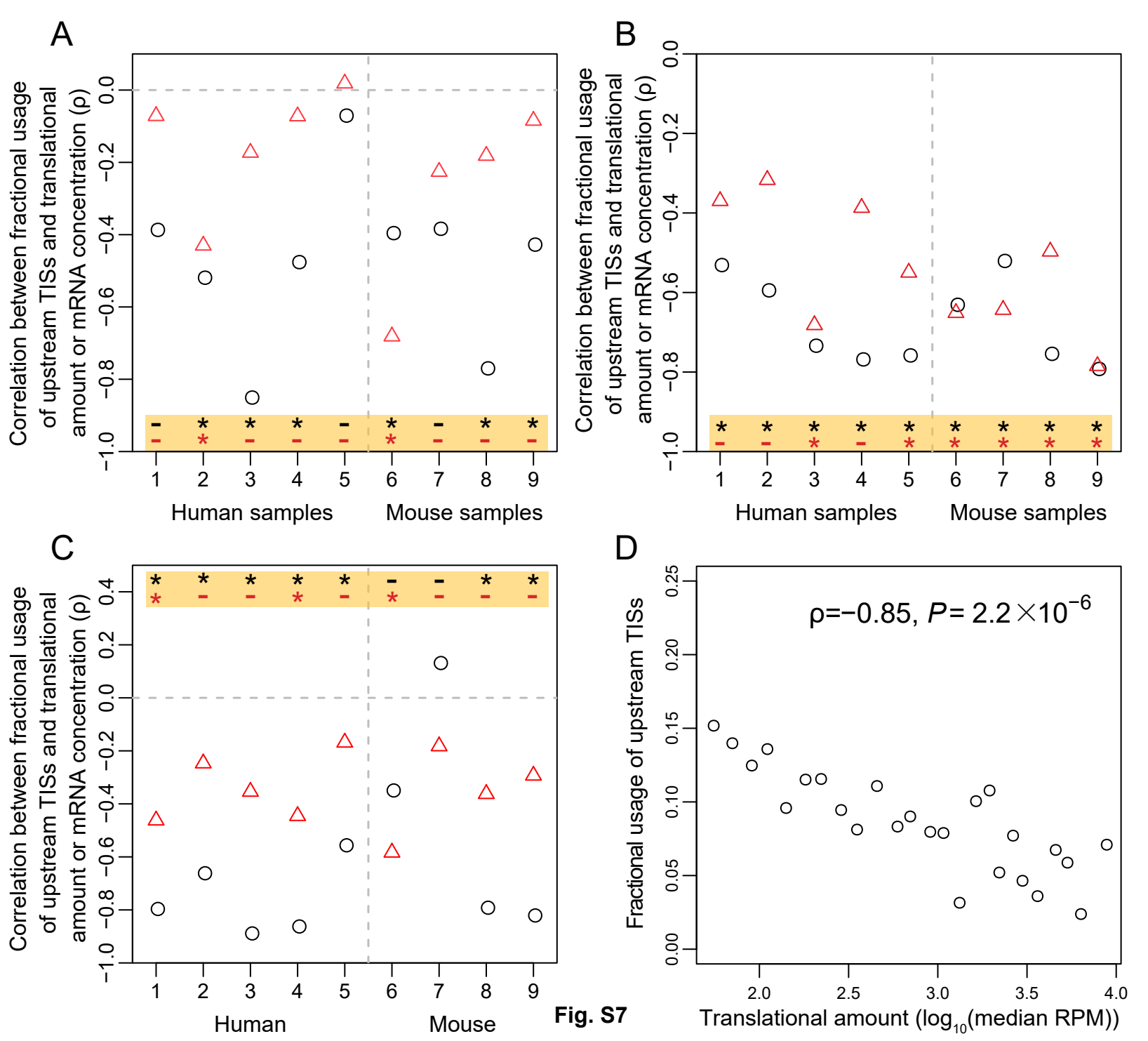
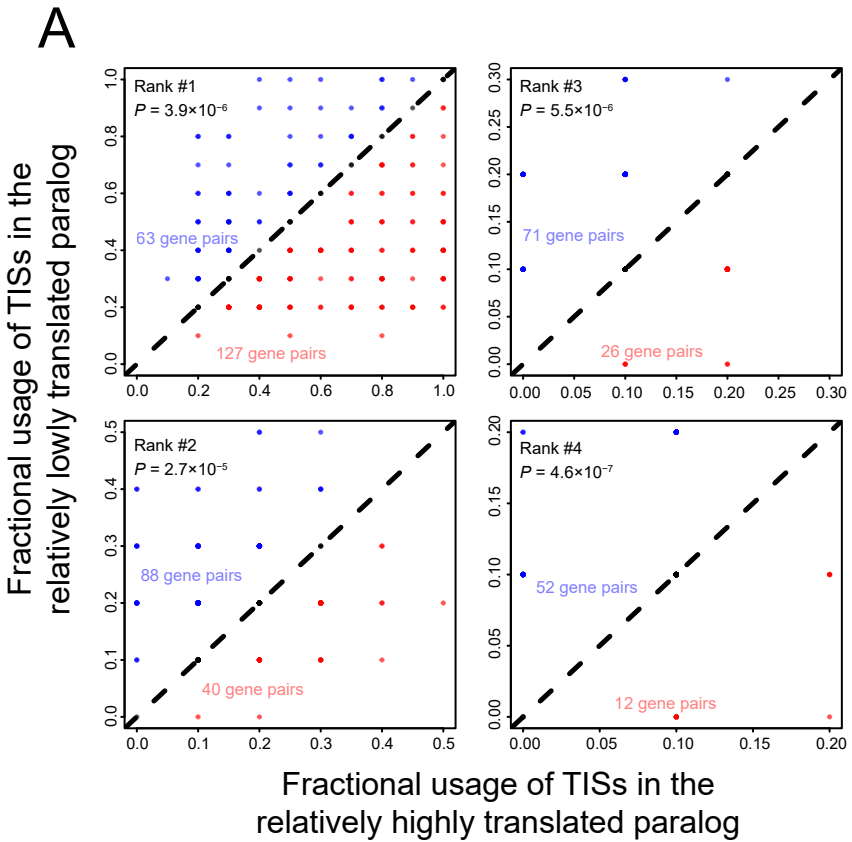

A
Correlation between fractional usage of upstream TISs and translational amount or mRNA concentration (ρ)
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
1
2
3
4
5
6
7
8
9
Human samples
Mouse samples
B
Correlation between fractional usage of upstream TISs and translational amount or mRNA concentration (ρ)
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
1
2
3
4
5
6
7
8
9
Human samples
Mouse samples
C
Correlation between fractional usage of upstream TISs and translational amount or mRNA concentration (ρ)
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
1
2
3
4
5
6
7
8
9
Human
Mouse
D
Fractional usage of upstream TISs
0.25
0.20
0.15
0.10
0.05
0.00
ρ=−0.85, P= 2.2×10−6
2.0
2.5
3.0
3.5
4.0
Translational amount (log10(median RPM))

Fig. S7


A
Rank #1
P = 3.9×10⁻⁶
63 gene pairs
127 gene pairs
Rank #3
P = 5.5×10⁻⁶
71 gene pairs
26 gene pairs
Rank #2
P = 2.7×10⁻⁵
88 gene pairs
40 gene pairs
Rank #4
P = 4.6×10⁻⁷
52 gene pairs
12 gene pairs
Fractional usage of TISs in the relatively lowly translated paralog
Fractional usage of TISs in the relatively highly translated paralog

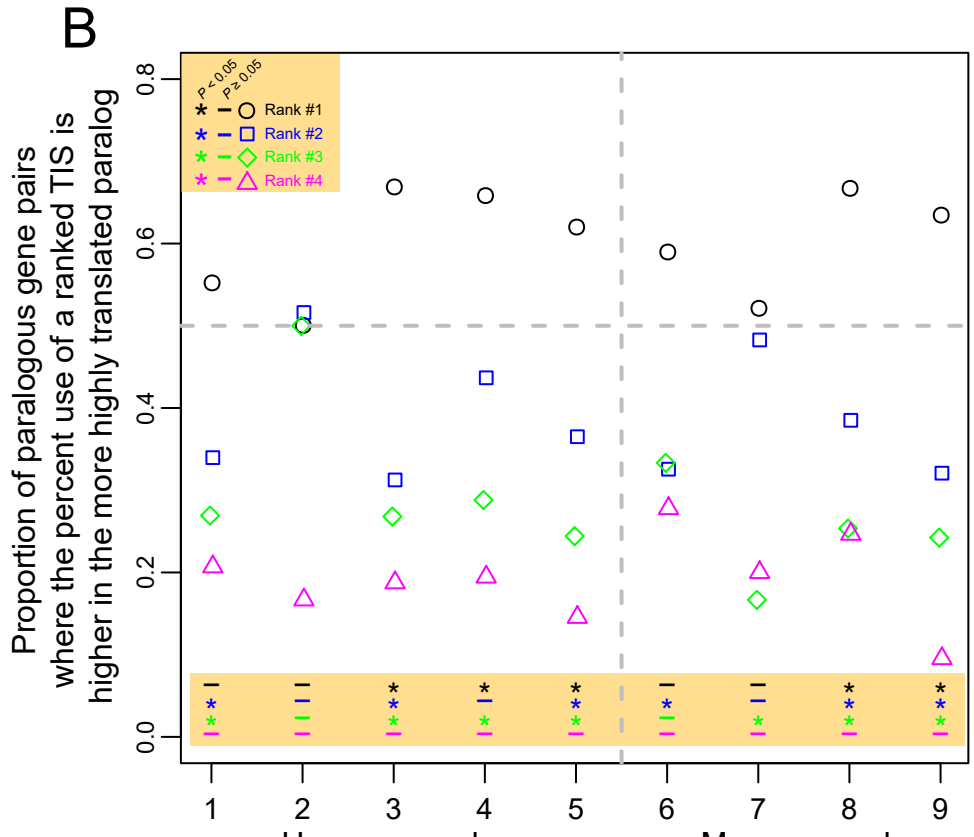

B
P < 0.05
P ≥ 0.05
Rank #1
Rank #2
Rank #3
Rank #4
Proportion of paralogous gene pairs where the percent use of a ranked TIS is higher in the more highly translated paralog
Human samples
Mouse samples

Fig. S8

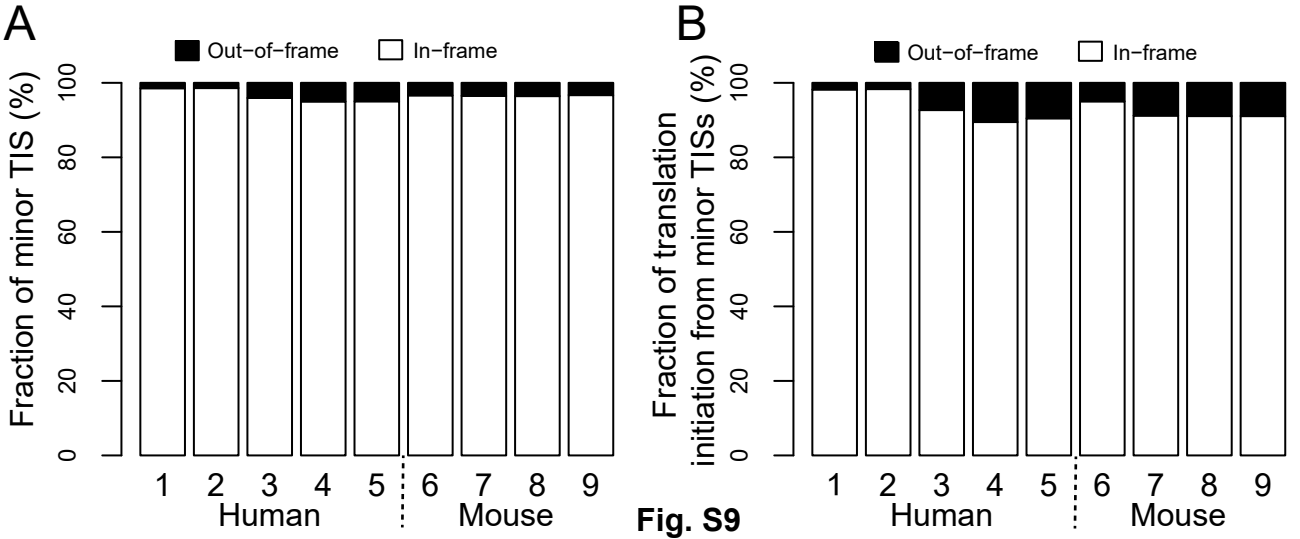

A
Out−of−frame
In−frame
Fraction of minor TIS (%)
0
20
40
60
80
100
1
2
3
4
5
6
7
8
9
Human
Mouse
B
Out−of−frame
In−frame
Fraction of translation initiation from minor TISs (%)
0
20
40
60
80
100
1
2
3
4
5
6
7
8
9
Human
Mouse

Fig. S9

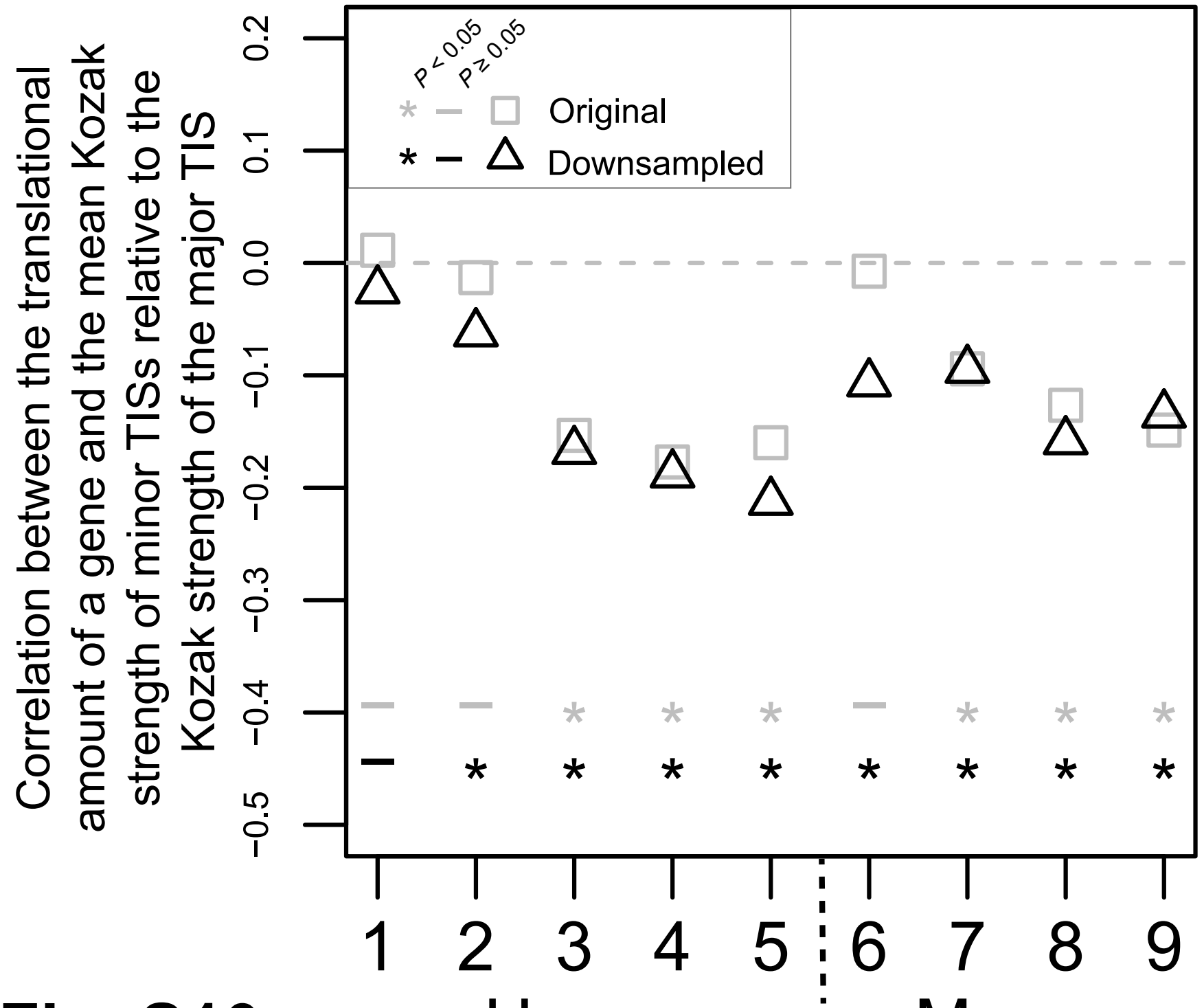

Correlation between the translational amount of a gene and the mean Kozak strength of minor TISs relative to the Kozak strength of the major TIS
P < 0.05
P ≥ 0.05
Original
Downsampled
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
1
2
3
4
5
6
7
8
9
Human
Mouse

Fig. S10

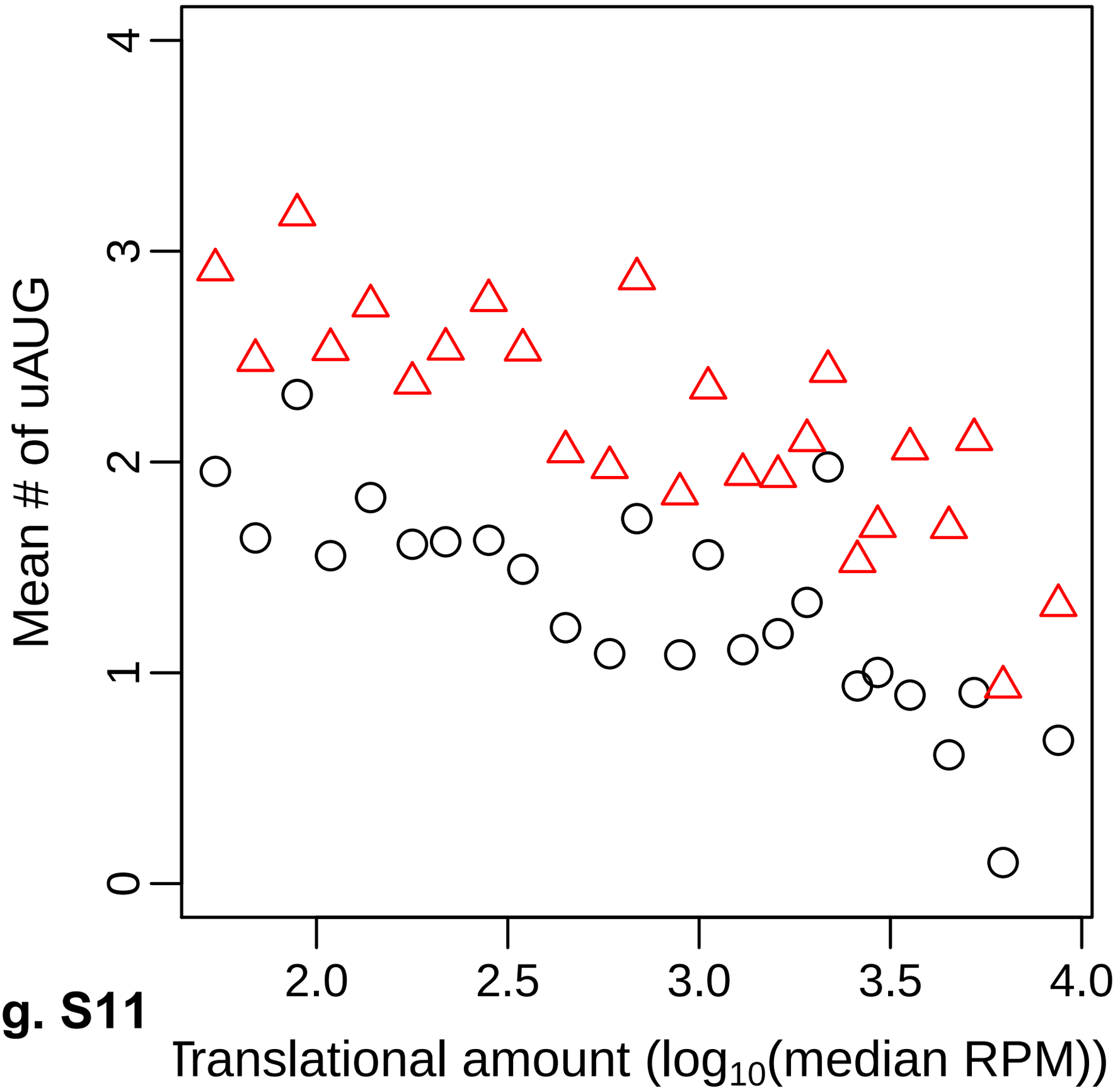

Mean # of uAUG
0
1
2
3
4
2.0
2.5
3.0
3.5
4.0
Translational amount (log_{10}(median RPM))

Fig. S11

Table S1. Datasets used in this study.

Sample ID	Species	Type	Technique	Description	SRA ID	References
1	Human	Cell line	QTI-seq	HeLa cells	NA	https://www.ncbi.nlm.nih.gov/pubmed/?term=26046440
	Human	Cell line	RNA-Seq	HeLa cells	SRR1661492	
2	Human	Cell line	QTI-seq	HeLa cells (YTHDF1 mutant)	NA	
	Human	Cell line	RNA-Seq	HeLa cells (YTHDF1 mutant)	SRR1661494	
3	Human	Cell line	QTI-seq	HEK293 cells	SRR1630828	
	Human	Cell line	RNA-Seq	HEK293 cells	SRR1630838	
4	Human	Cell line	QTI-seq	HEK293 cells (eIF2α(S51D) mutant)	SRR1630829	
	Human	Cell line	RNA-Seq	HEK293 cells (eIF2α(S51D) mutant)	SRR1630839	
5	Human	Cell line	QTI-seq	HEK293 cells (amino acid starvation)	SRR1630830	
	Human	Cell line	RNA-Seq	HEK293 cells (amino acid starvation)	SRR1630840	
6	Mouse	Tissue	QTI-seq	Liver	SRR1630809	https://www.ncbi.nlm.nih.gov/pubmed/?term=25486063
	Mouse	Tissue	RNA-Seq	Liver	SRR1630834	
7	Mouse	Tissue	QTI-seq	Liver (fasting)	SRR1630811	
	Mouse	Tissue	RNA-Seq	Liver (fasting)	SRR1630835	
8	Mouse	Cell line	QTI-seq	MEF cells	SRR1630814	
	Mouse	Cell line	RNA-Seq	MEF cells	SRR1630836	
9	Mouse	Cell line	QTI-seq	MEF cells (amino acid starvation)	SRR1630815	
	Mouse	Cell line	RNA-Seq	MEF cells (amino acid starvation)	SRR1630837	