Supplementary Materials

Supplementary Table 1. Nucleotide frequencies (%) at different codon sites in a non-redundant set of 19317 human mRNAs.

Nucleotide	Codon site 1	Codon site 2	Codon site 3
A	0.2666	0.3094	0.1908
U	0.1731	0.2633	0.2156
G	0.3132	0.1941	0.2923
C	0.2470	0.2333	0.3013

Supplementary Table 2. Dinucleotide frequencies (%) at three codon sites in abundant and rare human mRNAs. Dinucleotide frequencies were calculated from a non-redundant set of 19317 human mRNAs, 3227 abundant mRNAs (≥108 ESTs in GenBank) and 3639 rare mRNAs (1-15 ESTs in GenBank). Frequencies of CC, UG, CA at codon sites [1,2], GG at sites [2,3] and CC at positions [3,1] are significantly enhanced, while levels of AA and GA at positions [1,2] and UG at sites [3,1] are significantly reduced in highly abundant transcripts relatively to low abundant transcripts.

Dinucleotide	Sites [1,2]			Sites [2,3]			Sites [3,1]		
	Abund	Rare	All	Abund	Rare	All	Abund	Rare	All
CG	3.15	3.35	3.38	2.91	2.44	2.68	4.28	4.22	4.19
GC	6.99	7.30	7.11	7.24	6.22	6.70	7.69	7.05	7.59
UA	2.85	3.07	2.90	2.83	2.66	2.82	3.39	3.72	3.55
AU	6.49	7.05	6.63	5.59	6.43	6.02	2.87	2.60	2.73
CC	6.57	5.86	6.19	8.83	8.20	8.47	8.97	8.00	8.61
GG	6.92	6.88	6.69	5.77	4.86	5.37	9.96	10.21	10.16
UU	5.82	5.61	5.73	5.39	5.83	5.58	3.85	3.47	3.65
AA	8.68	10.20	9.32	6.35	6.88	6.67	4.96	4.86	4.96
UG	4.61	3.30	3.89	10.32	10.25	10.37	8.85	10.61	9.58
CA	7.39	6.77	7.23	5.89	5.83	5.90	10.22	10.77	10.33
AG	5.99	5.04	5.57	10.11	11.46	10.84	6.99	7.76	7.43
CU	8.26	7.34	7.93	6.19	6.43	6.24	7.66	6.57	7.03
AC	5.20	5.16	5.17	7.32	7.68	7.50	3.78	3.59	3.73
GU	5.71	6.21	5.98	3.84	3.78	3.75	3.95	3.92	3.92
GA	10.46	12.41	11.58	3.82	3.72	3.71	7.42	7.77	7.53
UC	5.05	4.57	4.81	7.74	7.47	7.50	4.92	4.67	4.81

Supplementary Table 3. Trinucleotide frequencies in abundant and rare human mRNAs. Trinucleotide frequencies were calculated from a non-redundant set of 19317 human mRNAs (All), 3227 abundant mRNAs (≥108 ESTs in GenBank) and 3639 rare mRNAs (1-15 ESTs in GenBank). mRNA sequences with complete CDSs and at least 30 nt 5'UTRs and 3'UTRs were used in this analysis. Frequencies of codons for histidine, proline, cysteine and tryptophan are significantly enhanced, and codon frequencies for lysine, asparagine, aspartate and glutamate are significantly reduced in highly abundant transcripts, as compared to rare transcripts.

AAA	AAC	AAG	AAT	ACA	ACC	ACG	ACT	AGA	AGC	AGG	AGT	ATA	ATC	ATG	ATT	
0.025	0.019	0.033	0.016	0.014	0.018	0.006	0.013	0.012	0.020	0.012	0.012	0.007	0.021	0.023	0.015	Total (19318)
0.026	0.021	0.038	0.017	0.014	0.019	0.006	0.013	0.011	0.018	0.010	0.011	0.006	0.023	0.024	0.017	Rare (3639)
0.023	0.018	0.030	0.015	0.014	0.019	0.006	0.013	0.013	0.021	0.014	0.012	0.008	0.020	0.023	0.014	Abundant (3228)

CAA	CAC	CAG	CAT	CCA	CCC	CCG	CCT	CGA	CGC	CGG	CGT	CTA	CTC	CTG	CTT	
0.012	0.015	0.034	0.011	0.017	0.020	0.008	0.017	0.006	0.011	0.012	0.005	0.007	0.020	0.040	0.013	Total (19318)
0.011	0.013	0.034	0.010	0.017	0.018	0.007	0.017	0.007	0.010	0.011	0.005	0.006	0.017	0.037	0.012	Rare (3639)
0.013	0.017	0.033	0.011	0.018	0.022	0.009	0.018	0.005	0.011	0.011	0.004	0.007	0.021	0.041	0.013	Abundant (3228)

GAA	GAC	GAG	GAT	GCA	GCC	GCG	GCT	GGA	GGC	GGG	GGT	GTA	GTC	GTG	GTT	
0.029	0.025	0.040	0.021	0.016	0.028	0.008	0.018	0.017	0.023	0.017	0.011	0.007	0.014	0.028	0.011	Total (19318)
0.031	0.027	0.042	0.025	0.017	0.029	0.007	0.020	0.018	0.023	0.015	0.012	0.007	0.014	0.029	0.012	Rare (3639)
0.026	0.023	0.037	0.018	0.015	0.029	0.009	0.017	0.017	0.024	0.018	0.010	0.006	0.015	0.027	0.010	Abundant (3228)

TAA	TAC	TAG	TAT	TCA	TCC	TCG	TCT	TGA	TGC	TGG	TGT	TTA	TTC	TTG	TTT	
0.001	0.015	0.001	0.012	0.012	0.017	0.005	0.014	0.002	0.014	0.013	0.011	0.007	0.020	0.013	0.017	Total (19318)
0.001	0.016	0.001	0.013	0.011	0.016	0.005	0.014	0.002	0.011	0.012	0.009	0.006	0.020	0.012	0.017	Rare (3639)
0.001	0.015	0.001	0.012	0.012	0.019	0.005	0.014	0.002	0.017	0.015	0.013	0.007	0.022	0.012	0.017	Abundant (3228)

Supplementary Table 4. Frequencies of trinucleotides (*F*). complementary trinucleotides (F_c), base paired trinucleotides (*f*), and base paired complementary trinucleotides (f_c) at codon sites [2,3,1]. The 3rd four-fold degenerate codon positions are boldfaced.

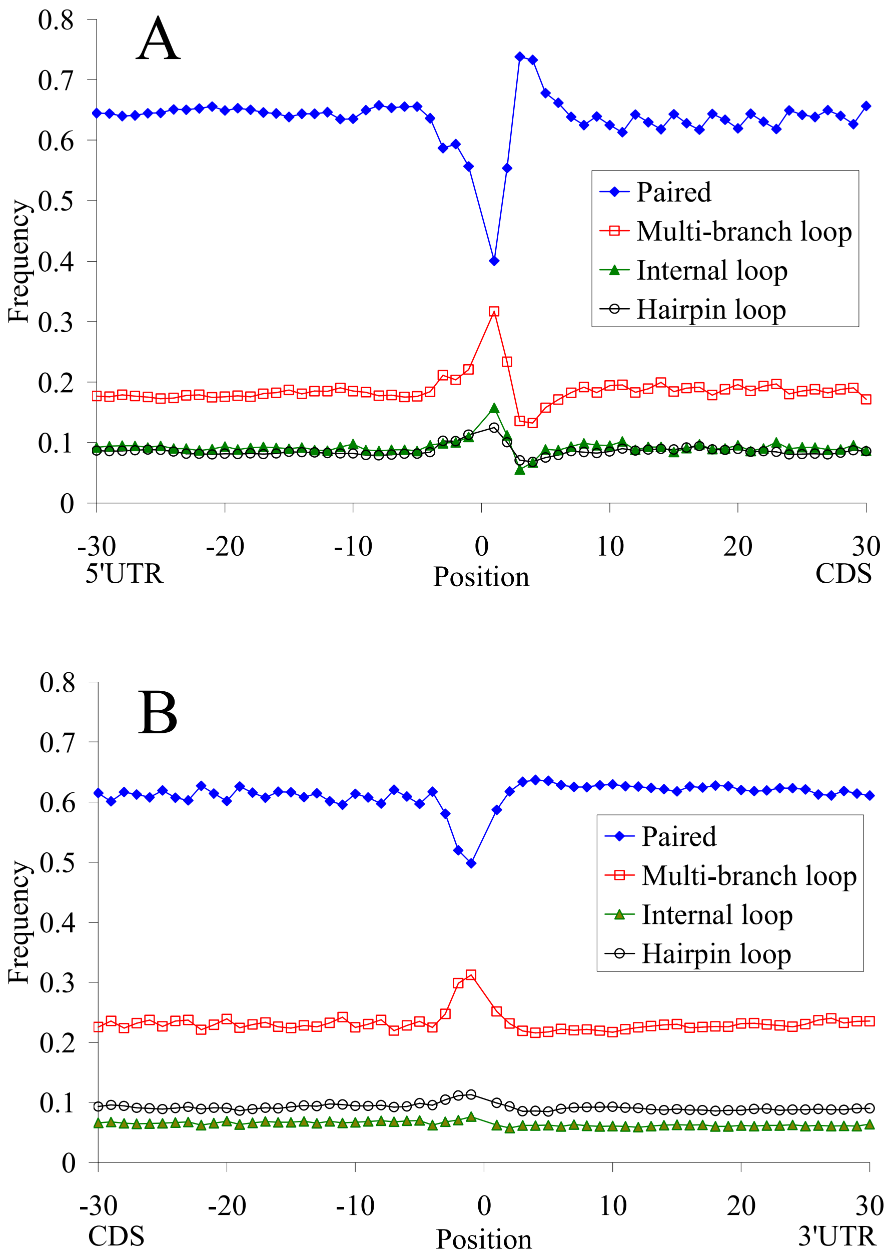
Trinucleotides			Complementary trinucleotides		
Sequence	F	f	Sequence	F_c	f_c
C**C**A	0.542	0.676	T**G**G	0.618	0.723
C**C**C	0.385	0.375	G**G**G	0.312	0.417
C**C**G	0.144	0.127	C**G**G	0.132	0.125
G**C**A	0.376	0.436	T**G**C	0.470	0.577
G**C**C	0.372	0.313	G**G**C	0.316	0.446
G**C**G	0.158	0.145	C**G**C	0.131	0.161
T**C**A	0.352	0.371	T**G**A	0.413	0.470
T**C**C	0.246	0.209	G**G**A	0.282	0.334
T**C**G	0.068	0.037	C**G**A	0.060	0.040
x**C**x	0.293	0.299	x**G**x	0.304	0.366
C**T**A	0.164	0.131	T**A**G	0.119	0.096
C**T**C	0.268	0.256	G**A**G	0.303	0.263
C**T**G	0.376	0.381	C**A**G	0.349	0.367
G**T**A	0.086	0.067	T**A**C	0.090	0.059
G**T**C	0.137	0.115	G**A**C	0.175	0.124
G**T**G	0.226	0.175	C**A**C	0.216	0.208
T**T**A	0.122	0.095	T**A**A	0.113	0.065
T**T**C	0.194	0.155	G**A**A	0.257	0.163
T**T**G	0.196	0.144	C**A**A	0.235	0.153
x**T**x	0.197	0.169	x**A**x	0.206	0.166

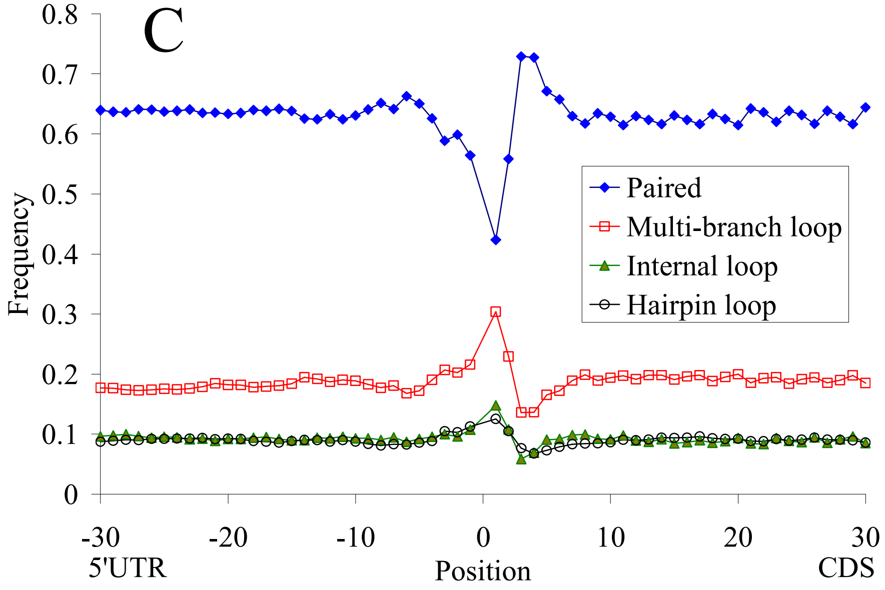
Supplementary Table 5. Differences in base pairing levels (Δf) between codon sites [1,2], [2.3], and [3,1] in the human mRNAs and in randomized sequences.

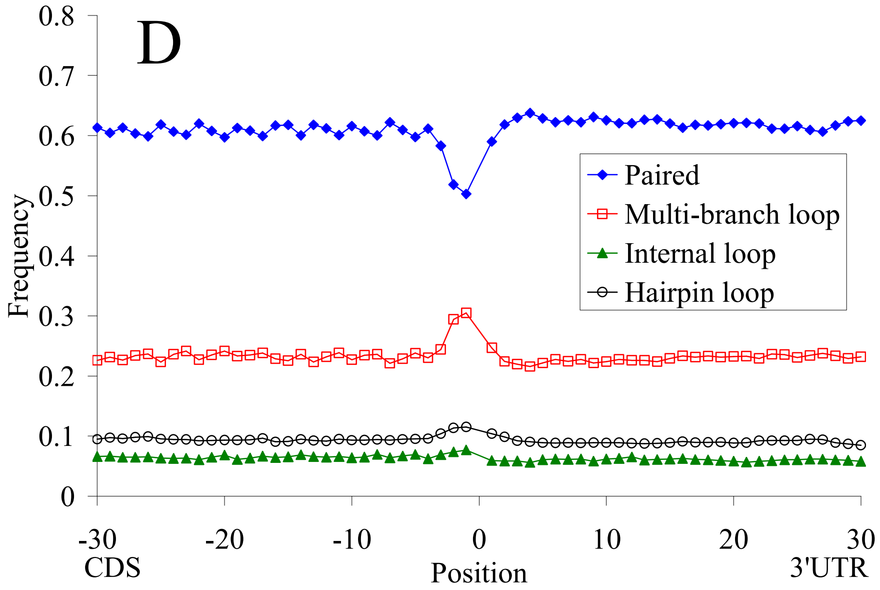
Sequences	Codon sites [1,2]		Codon sites [2,3]		Codon sites [3,1]	
	$\Delta f_{1,2}$	*P*-value	$\Delta f_{2,3}$	*P*-value	$\Delta f_{3,1}$	*P*-value
Real mRNAs	.01365±.00047		.02272±.00047		.00907±.00047	
Random CC	.01112±.00045	10^{-1}	.00537±.00044	10^{-124}	-.00576±.00044	10^{-120}
Random NS	.00042±.00046	10^{-82}	.00051±.00046	10^{-255}	.00019±.00046	10^{-45}
Random CSx4	.01128±.00047	10^{-3}	.01609±.00047	10^{-20}	.00480±.00047	10^{-9}
Random CCx4	.01101±.00046	10^{-3}	.01199±.00045	10^{-43}	.00097±.00045	10^{-29}
Random DCS	.01307±.00045	0.2	.02141±.00045	0.03	.00834±.00046	0.12
Random DNS	.00026±.00045	10^{-89}	.00007±.00045	10^{-272}	.00019±.00045	10^{-44}

Abbreviations: Real mRNAs, coding sequences of 19,317 native human mRNA; Random CC, mRNA sequences with randomly chosen synonymous codons; Random NS, sequences with randomly shuffled nucleotides and the same nucleotide composition as native mRNAs; Random CSx4, sequences with shuffled 4-fold degenerate synonymous codons; Random CCx4, sequences with randomly chosen 4-fold degenerate synonymous codons; Random DCS, dicodone shuffling that preserved dinucleotide frequencies, encoded amino acid sequence, and codon usage of native mRNAs; Random DNS, random shuffling of all dinucleotides that retained nucleotide composition of native mRNAs. Start and stop codons were excluded from this analysis.

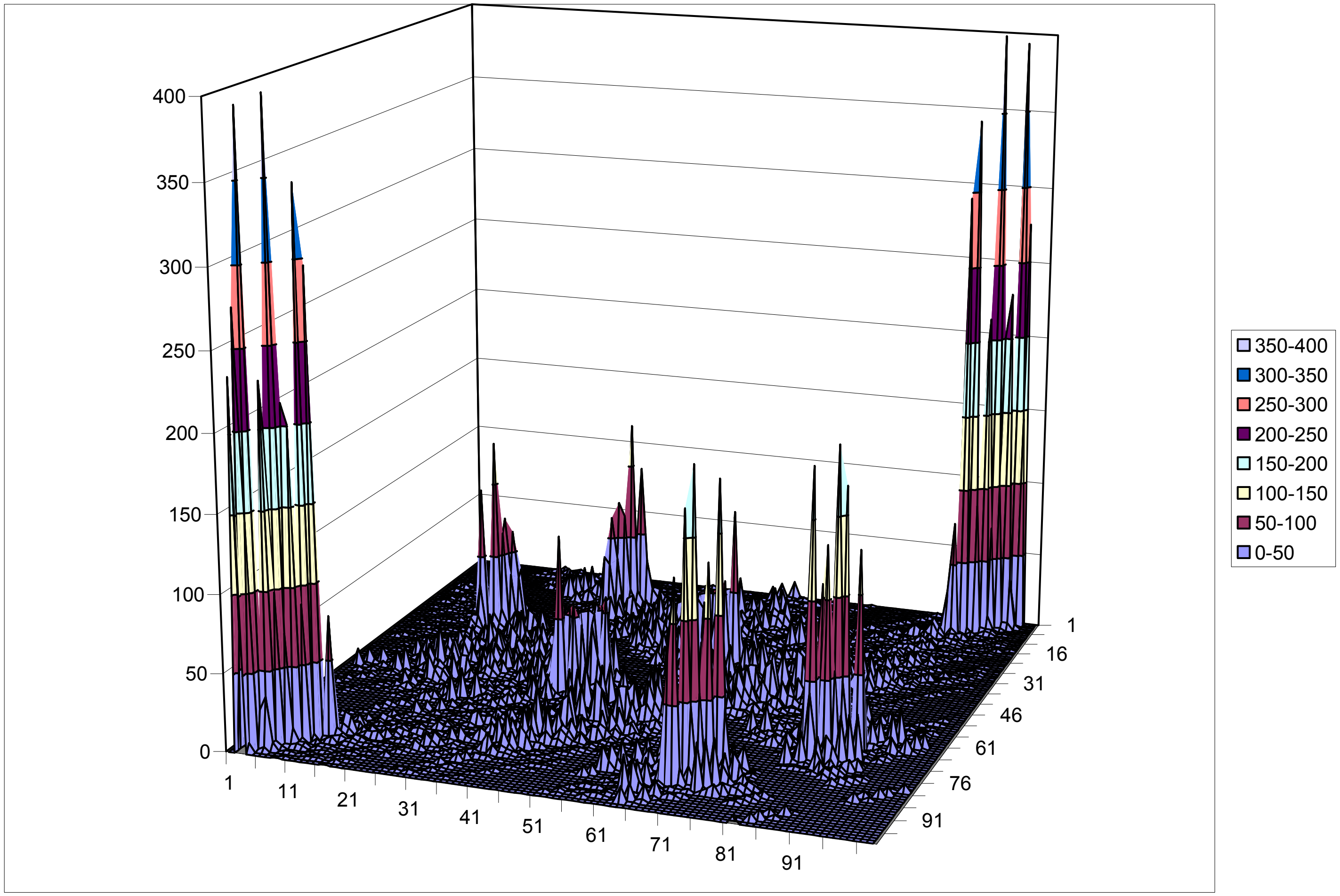
Supplementary Figure 1. Profiles of nucleotide involvement in different secondary structure elements around the start codon (**A** and **C**) and the stop codon (**B** and **D**) in in 19,317 human mRNAs (**A** and **B**) and 20,892 mouse mRNAs (**C** and **D**). Positions from -30 to -1 correspond to 5'UTRs and positions from 1 to 30 correspond to CDSs. Blue, base paired nucleotides; red, nucleotides in multi-branch loops; green, nucleotides in internal loops; black, nucleotides in hairpins.




C
Frequency
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-30
-20
-10
0
10
20
30
5'UTR
Position
CDS
Paired
Multi-branch loop
Internal loop
Hairpin loop


D
Frequency
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-30
-20
-10
0
10
20
30
CDS
Position
3'UTR
Paired
Multi-branch loop
Internal loop
Hairpin loop

Supplementary Figure 2. Summarized histogram of nucleotide base pairing for 622 human tRNAs. Sequences of tRNAs were computationally folded as described in Materials and Methods. Eight major peaks on the hisogram correspond to eight canonical tRNA stem structures.


400
350
300
250
200
150
100
50
0
1
11
21
31
41
51
61
71
81
91
1
16
31
46
61
76
91
350-400
300-350
250-300
200-250
150-200
100-150
50-100
0-50