Effects of codon optimization on coagulation factor IX translation and structure: Implications for protein and gene therapies

Aikaterini Alexaki [a*], Gaya K. Hettiarachchi [a*], John C. Athey [a], Upendra K. Katneni [a], Vijaya Simhadri[a], Nobuko Hamasaki-Katagiri [a], Puja Nanavaty[b], Brian Lin [a], Kazuyo Takeda[a], Darón Freedberg[a], Dougald Monroe[c], Joseph R. McGill [a], Robert Peters[d], Jacob M. Kames[a], David D. Holcomb[a], Ryan C. Hunt [a], Zuben E. Sauna[a], Amy Gelinas[e], Nebojsa Janjic[e], Michael DiCuccio[f], Haim Bar[g], Anton A. Komar[b], and Chava Kimchi-Sarfaty[a]

[a] Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, MD, USA

[b] Center for Gene Regulation in Health and Disease, Cleveland State University, Cleveland, OH, USA

[c] University of North Carolina at Chapel hill, Chapel hill, NC, USA

[d] Bioverativ, Cambridge, MA, USA

[e] SomaLogic, Inc., Boulder, CO, USA

[f] National Center of Biotechnology Information, National Institutes of Health, Bethesda, MD, USA

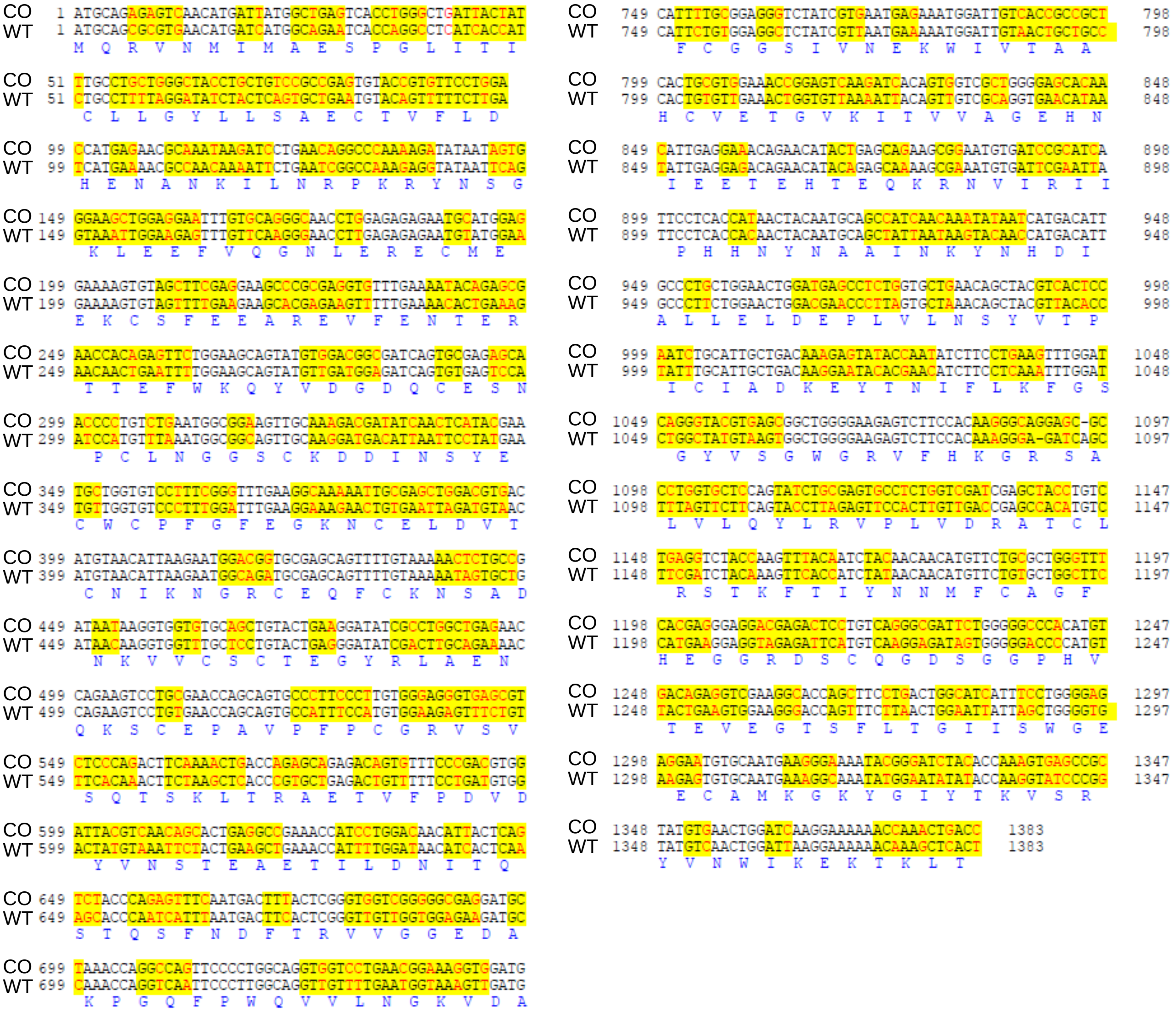
[g] Department of Statistics, University of Connecticut, Storrs, CT, USA

Corresponding author:

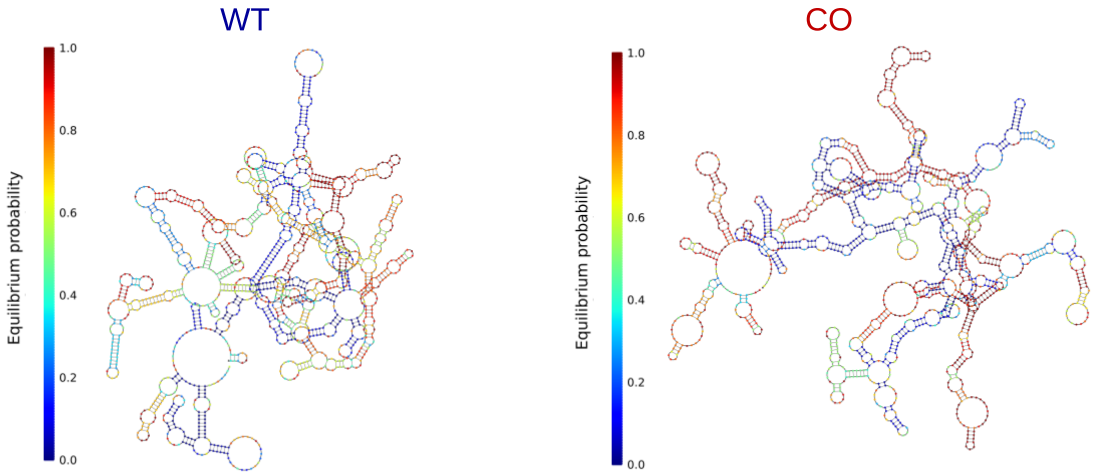
Chava.kimchi-sarfaty@fda.hhs.gov; Division of Plasma Protein Therapeutics, Office of Tissue and Advanced Therapies, Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, USA

*Authors AA and GKH contributed equally

Supplemental Figure S1. Comparison of the nucleotide sequences of codon-optimized and wild type *F9* variants. Nucleotide changes between the two sequences are indicated in red, while codon changes are highlighted in yellow.



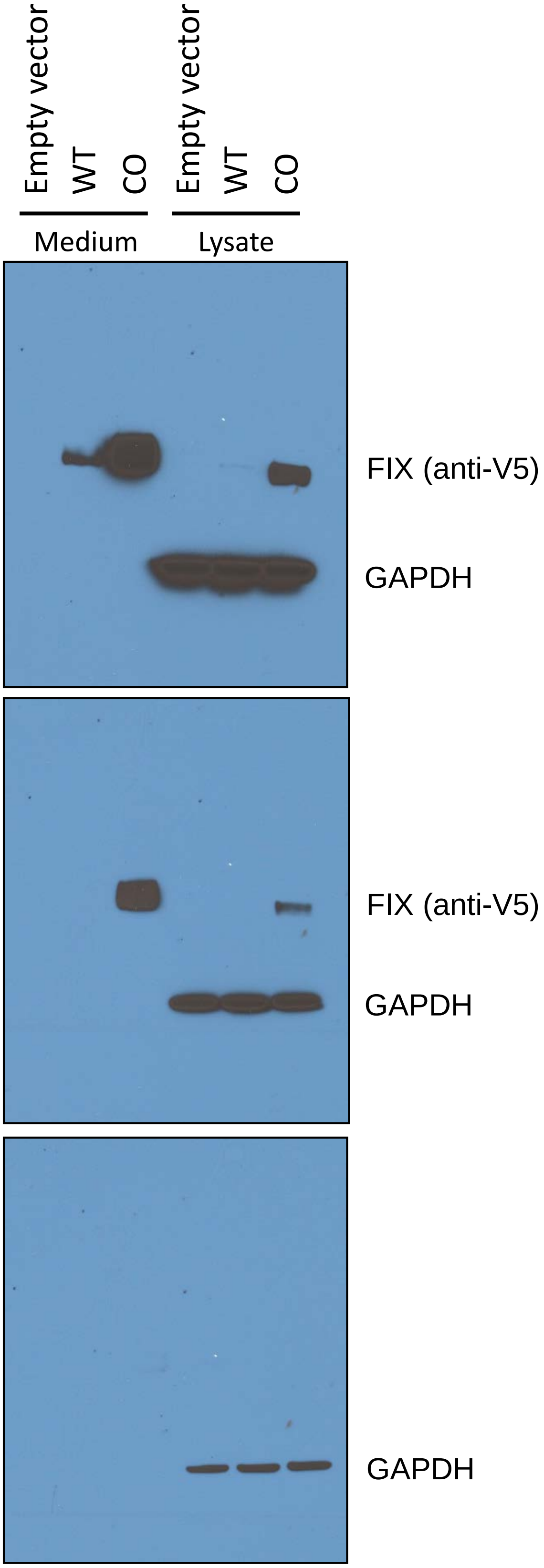
Supplemental Figure S2. In silico analyses of mRNA folding and stability for the WT and CO F9 coding sequence. Predicted MFE values and structures are shown for 37°C based on NUPACK algorithm (www.nupack.org)



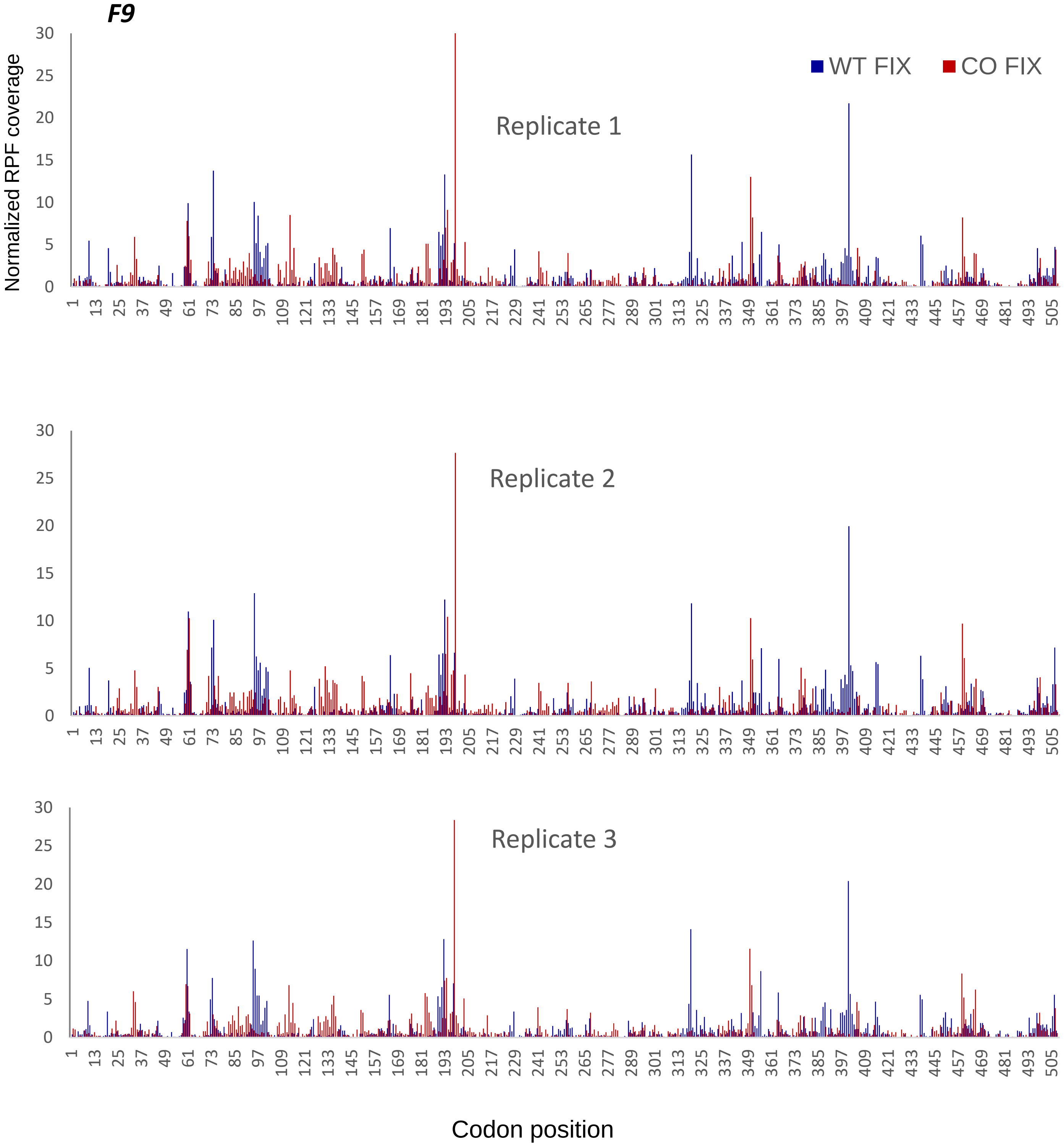
MFE: -339.90 kcal/mol

MFE: -410.50 kcal/mol

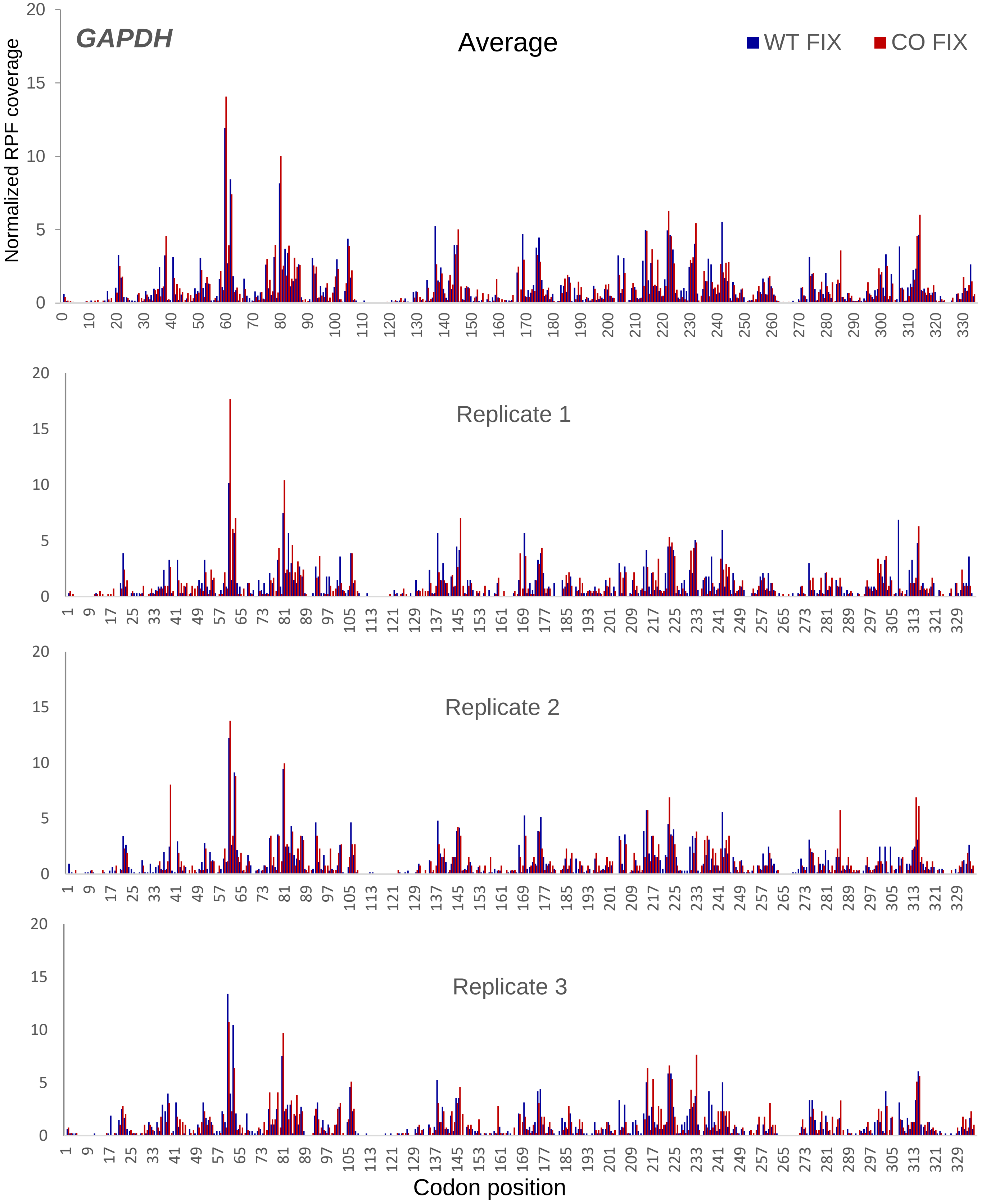
Supplemental Figure S3. FIX expression is higher in HEK293T cells stably transfected with wild-type vs. codon-optimized *F9.* Cell supernatant (concentrated medium) and lysate was collected and analyzed by immunoblotting. Membranes were stained with anti-V5 or anti-GAPDH. Multiple exposures of film are shown to demonstrate FIX (top image) and GAPDH (bottom image) at appropriate contrast. In the medium lanes, 5 μg of protein were loaded per lane. In the lysate lanes, 10 μg of protein were loaded per lane. Portions of these images were used to create Figure 2b.



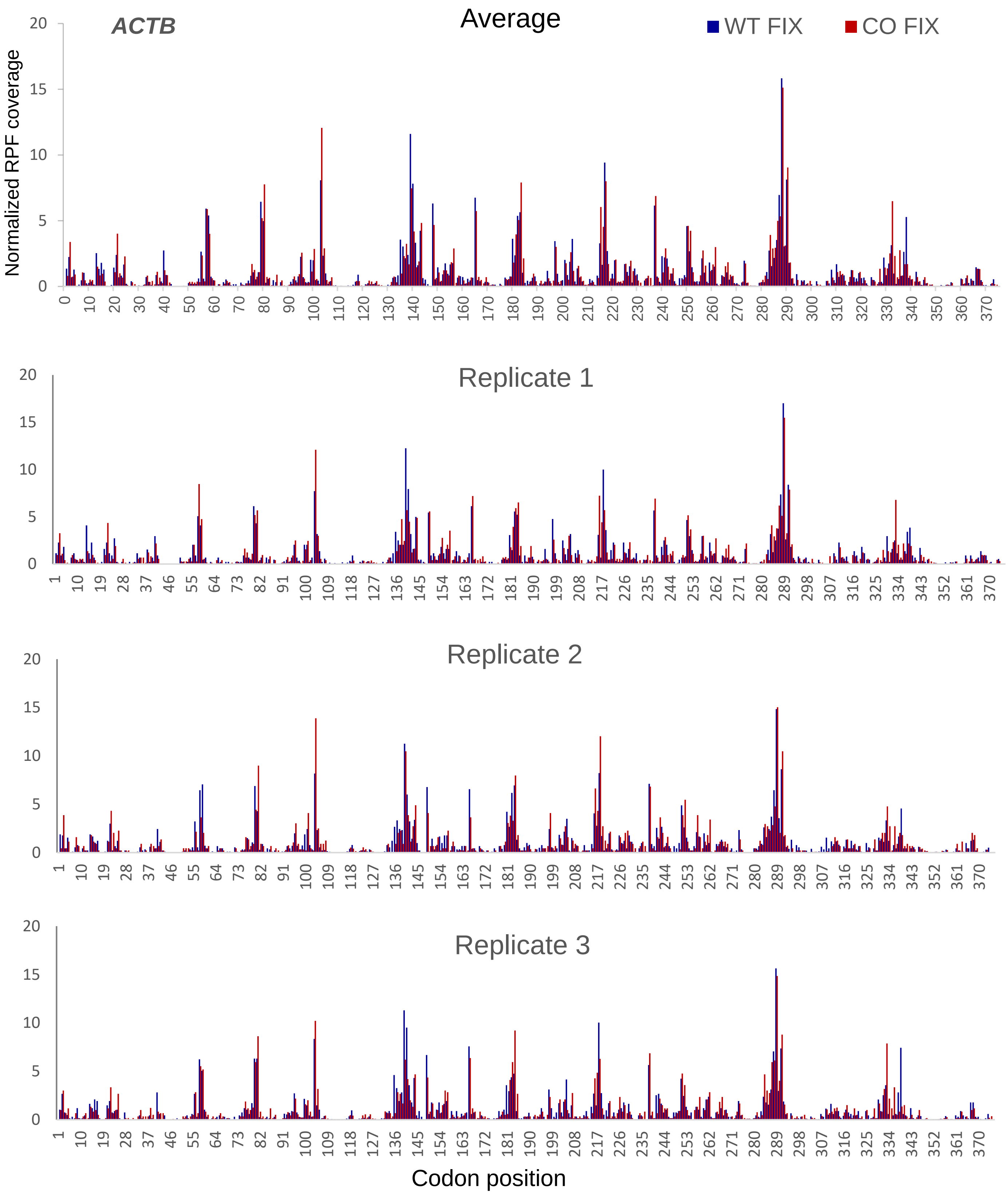
Supplemental Figure S4. RPF densities along the WT and CO *F9* mRNA for three replicates show reproducibility of results.



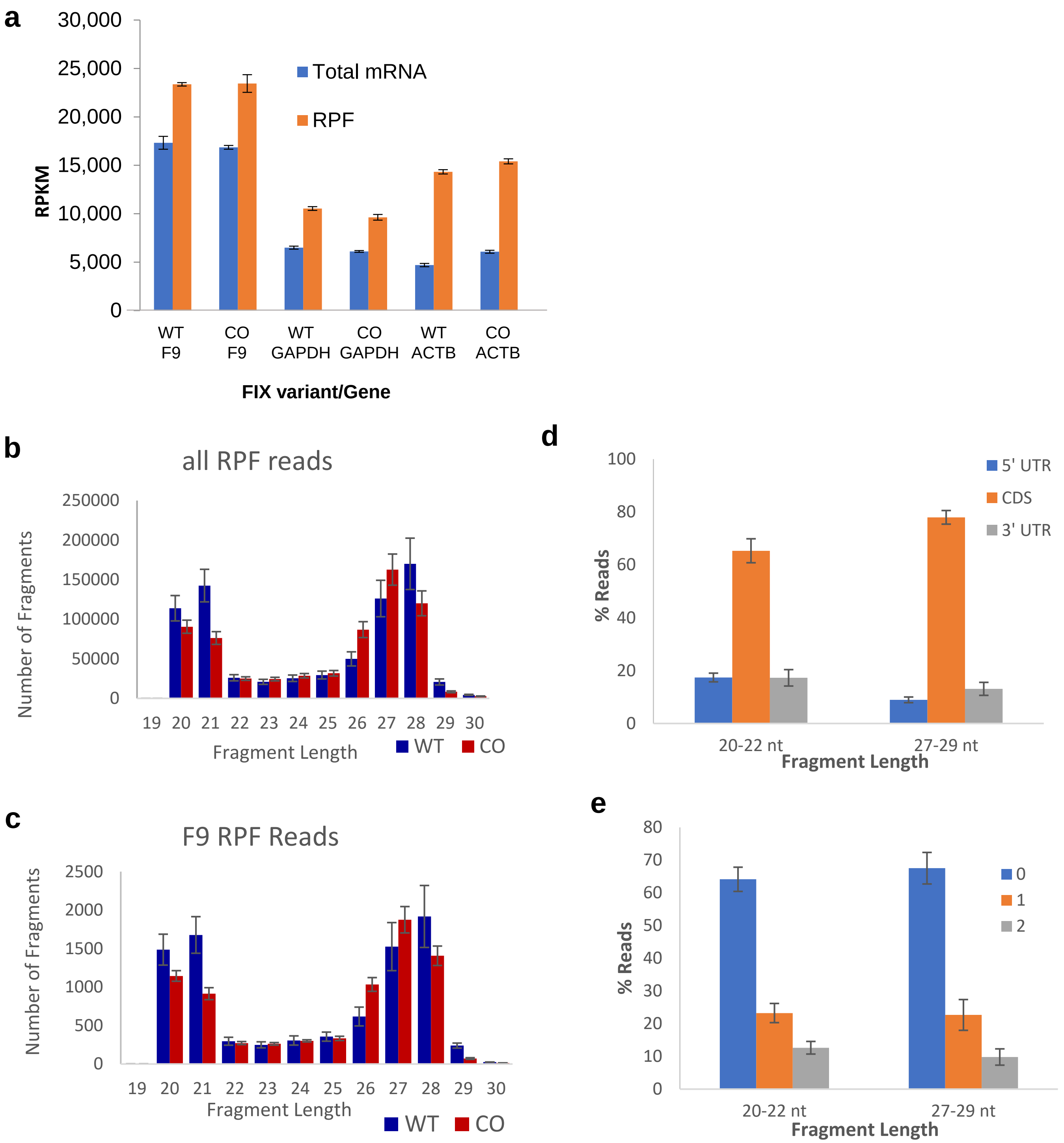
Supplemental Figure S5. *GAPDH* is translated with very similar kinetics in the wild type and codon-optimized expressing cell lines. RPF densities along the G*APDH* mRNA are shown as average of 3 experiments and in each individual ribosome profiling experiment.



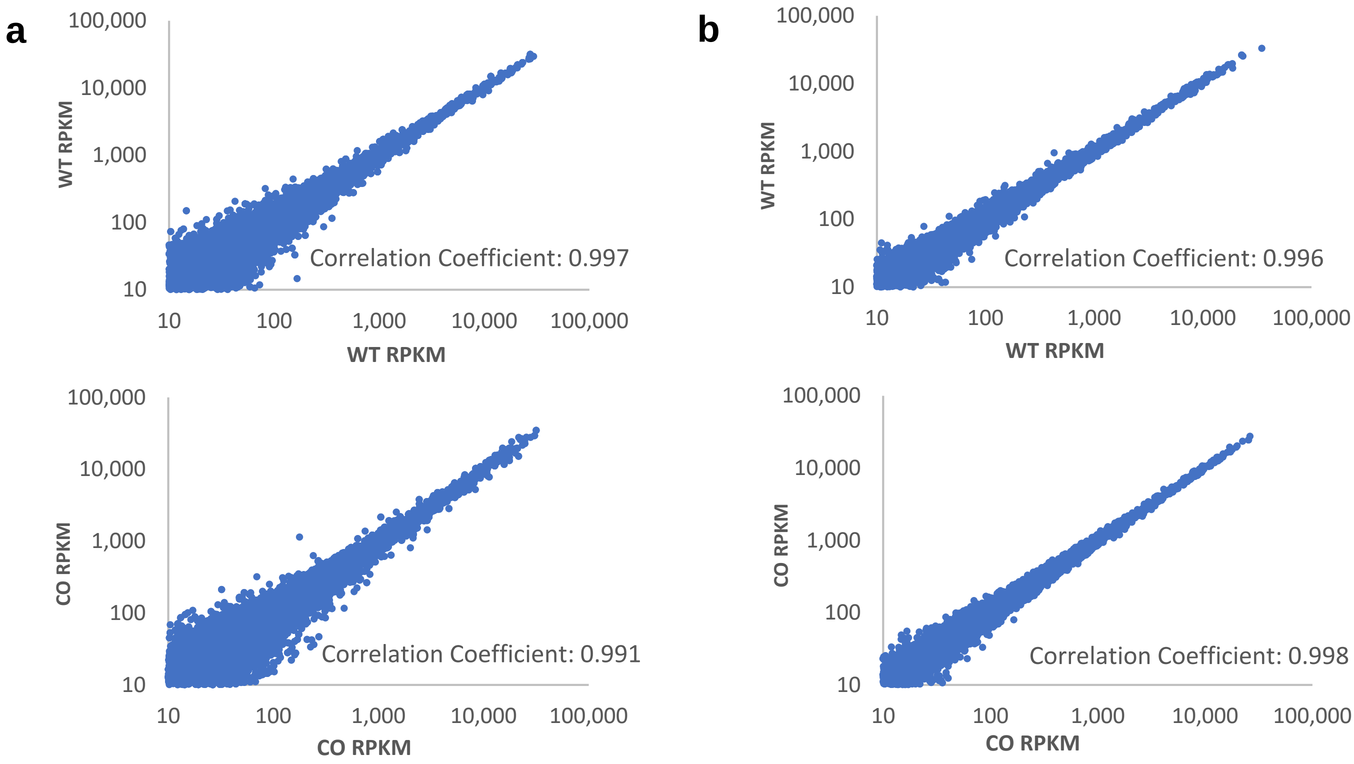
Supplemental Figure S6. *ACTB* is translated with very similar kinetics in the wild type and codon-optimized expressing cell lines. RPF densities along the *ACTB* mRNA are shown as average of 3 experiments and in each individual ribosome profiling experiment.



Supplemental Figure S7. Multiple sets of controls show consistency in ribosome profiling experiments and analysis (a) Transcript and RPF abundance for *F9*, *GAPDH* and *ACTB* in the WT and CO FIX expressing cell lines. The average of 3 experiments was plotted. (b) RPF size distribution of all reads in the WT and CO FIX expressing cell lines. (c) *F9* RPF size distribution in the WT and CO FIX expressing cell lines. (d) RPF sizes 20-22 and 27-29nt mapped 5'UTR, CDS and 3'UTR regions, (c) triplet periodicity of RPF fragment sizes 20-22 and 27-29 nucleotides, the average of 3 experiments was plotted, s.e.m. are shown.



Supplemental Figure S8. Reproducibility of ribosome profiling experiments. Pearson's correlation for (a) RPFs and (b) total mRNA between two representative experiments.



Supplemental Table 1. In a multiple linear regression model, the ribosome occupancy of adjacent codons is the only consistently significant predictor of ribosome occupancy of a codon.

	Estimate	Std. Error	t-value	Pr(>|t|)	Significance
WT *F9*					
(Intercept)	-1.05664	0.50251	-2.103	0.036	*
RSCU_Genome	0.03087	0.89716	0.034	0.9726	
RSCU_HEK_SRR	0.55596	1.06783	0.521	0.6028	
RSCU_HEK_ERX	-0.44406	1.23987	-0.358	0.7204	
RSCPU_Genome	-0.49032	0.30636	-1.6	0.1101	
RSCPU_HEK_SRR	-0.39351	0.40322	-0.976	0.3296	
RSCPU_HEK_ERX	0.73974	0.4682	1.58	0.1148	
tRNA	-4.70107	3.00585	-1.564	0.1185	
MFE_151	-0.02097	0.0105	-1.997	0.0464	*
Pair-prob	0.22075	0.16498	1.338	0.1815	
Lag3	0.05156	0.04475	1.152	0.2498	
Lag2	0.03957	0.04945	0.8	0.4239	
Lag1	0.47339	0.04486	10.554	<2e-16	***
CO *F9*					
(Intercept)	-0.428561	0.470664	-0.911	0.362978	
RSCU_Genome	0.987203	0.686076	1.439	0.150808	
RSCU_HEK_SRR	-0.510807	0.884671	-0.577	0.563933	
RSCU_HEK_ERX	-0.562106	1.060021	-0.53	0.596157	
RSCPU_Genome	-0.100063	0.230617	-0.434	0.664555	
RSCPU_HEK_SRR	0.012383	0.335691	0.037	0.970589	
RSCPU_HEK_ERX	0.119879	0.396397	0.302	0.762458	
tRNA	-3.330872	2.315622	-1.438	0.150943	
MFE_151	-0.007955	0.008367	-0.951	0.342211	
Pair-prob	-0.130687	0.200669	-0.651	0.515183	
Lag3	-0.02156	0.045746	-0.471	0.637641	
Lag2	0.177367	0.048099	3.688	0.000252	***
Lag1	0.390892	0.045033	8.68	<2e-16	***
ACTB					
(Intercept)	-1.404031	0.545822	-2.572	0.0105	*
RSCU_Genome	-0.310705	0.916587	-0.339	0.7348	
RSCU_HEK_SRR	-0.796966	1.242992	-0.641	0.5218	
RSCU_HEK_ERX	1.025136	1.514433	0.677	0.4989	
RSCPU_Genome	-0.100309	0.265983	-0.377	0.7063	
RSCPU_HEK_SRR	-0.075492	0.43977	-0.172	0.8638	
RSPCU_HEK_ERX	0.273948	0.478317	0.573	0.5672	
tRNA	1.554097	2.632387	0.59	0.5553	
MFE_151	-0.008856	0.007942	-1.115	0.2655	
Pair-prob	0.642673	0.26481	2.427	0.0157	*
Lag3	0.107027	0.052224	2.049	0.0411	*
Lag2	0.05047	0.057602	0.876	0.3815	
Lag1	0.442478	0.052295	8.461	6.64E-16	***
GAPDH					
(Intercept)	-0.149095	0.462448	-0.322	0.74736	
RSCU_Genome	1.409989	0.828001	1.703	0.08955	.
RSCU_HEK_SRR	-1.172997	1.265043	-0.927	0.3545	
RSCU_HEK_ERX	-0.093217	1.421276	-0.066	0.94775	
RSCPU_Genome	-0.652895	0.287641	-2.27	0.02388	*
RSCPU_HEK_SRR	-0.4377	0.438042	-0.999	0.31844	
RSCPU_HEK_ERX	1.113346	0.497467	2.238	0.0259	*
tRNA	-2.032467	2.679425	-0.759	0.44868	
MFE_151	0.008537	0.006816	1.252	0.2113	
Pair-prob	0.061488	0.22621	0.272	0.78594	
Lag3	0.031727	0.05551	0.572	0.56802	
Lag2	0.179942	0.058943	3.053	0.00246	**
Lag1	0.371884	0.056404	6.593	1.76E-10	***