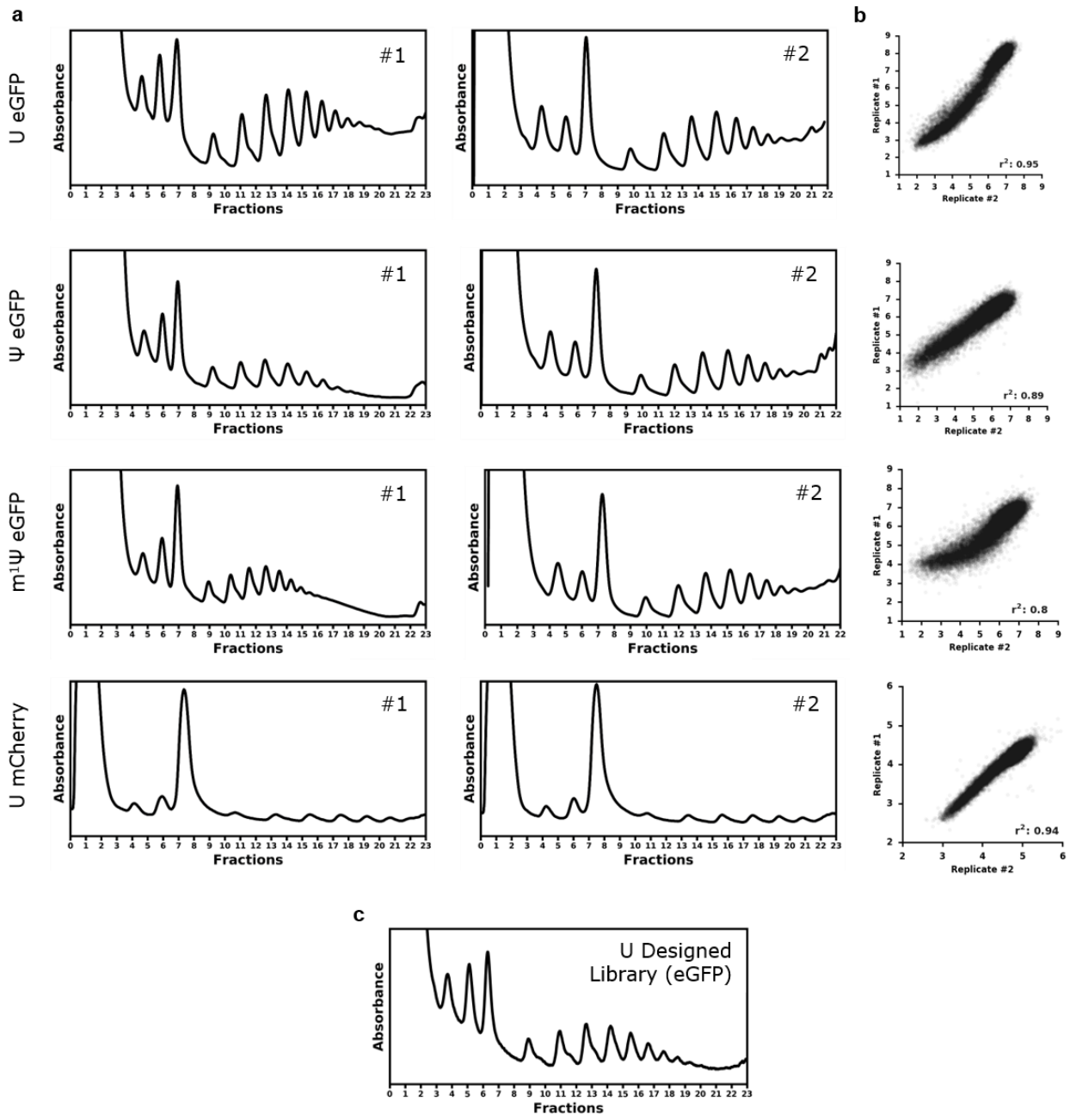
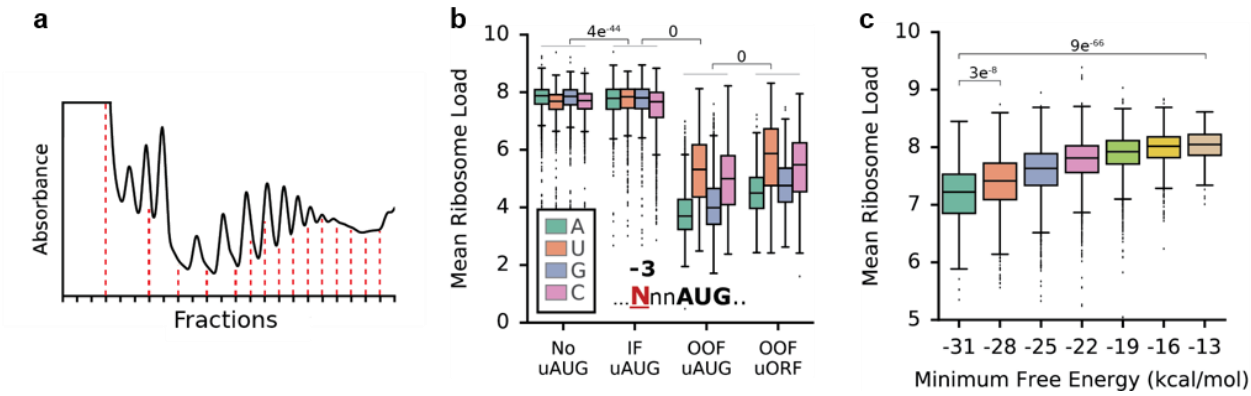


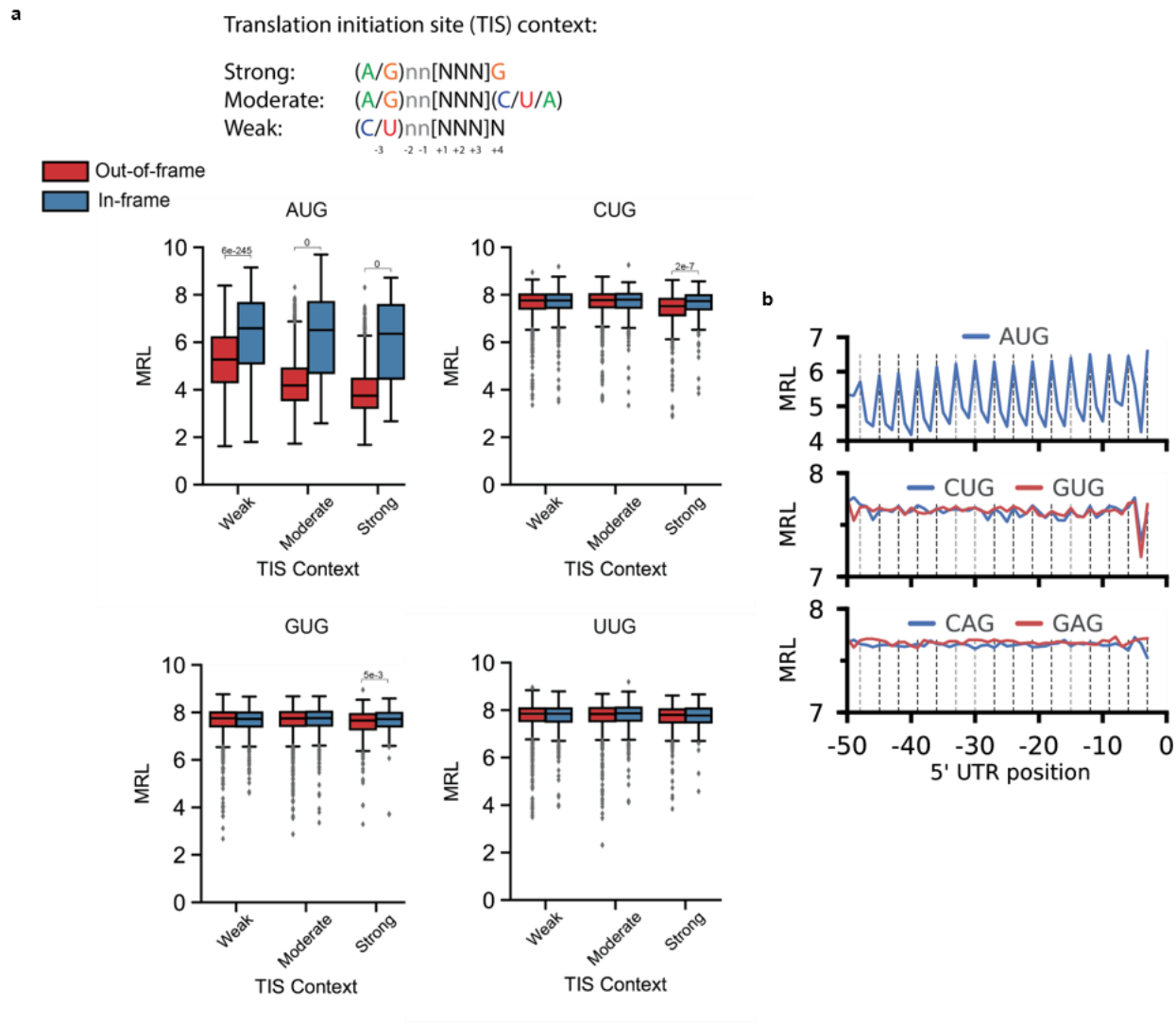
Supplementary Fig. 1

(a) Polysome profiles of two independent experiments for unmodified (U) RNA eGFP, pseudouridine (Ψ) eGFP, 1-methyl-pseudouridine ($m^1\Psi$) eGFP, and unmodified (U) RNA mCherry libraries. The library IVT mRNA were transfected into two independent cell plates, and polysome profiles were harvested independently. The polysome profiling protocol for the mCherry experiments differed from the eGFP experiments (**Online Methods**) **(b)** Mean ribosome load (MRL) correlation between experiments (n = 20,000). **(c)** Polysome profile of the designed library which includes the sub-libraries: human UTRs and SNVs, target ribosome loading, step-wise evolution, eGFP controls, 2,000 mCherry library UTRs, and additional random UTRs.



Supplementary Fig. 2

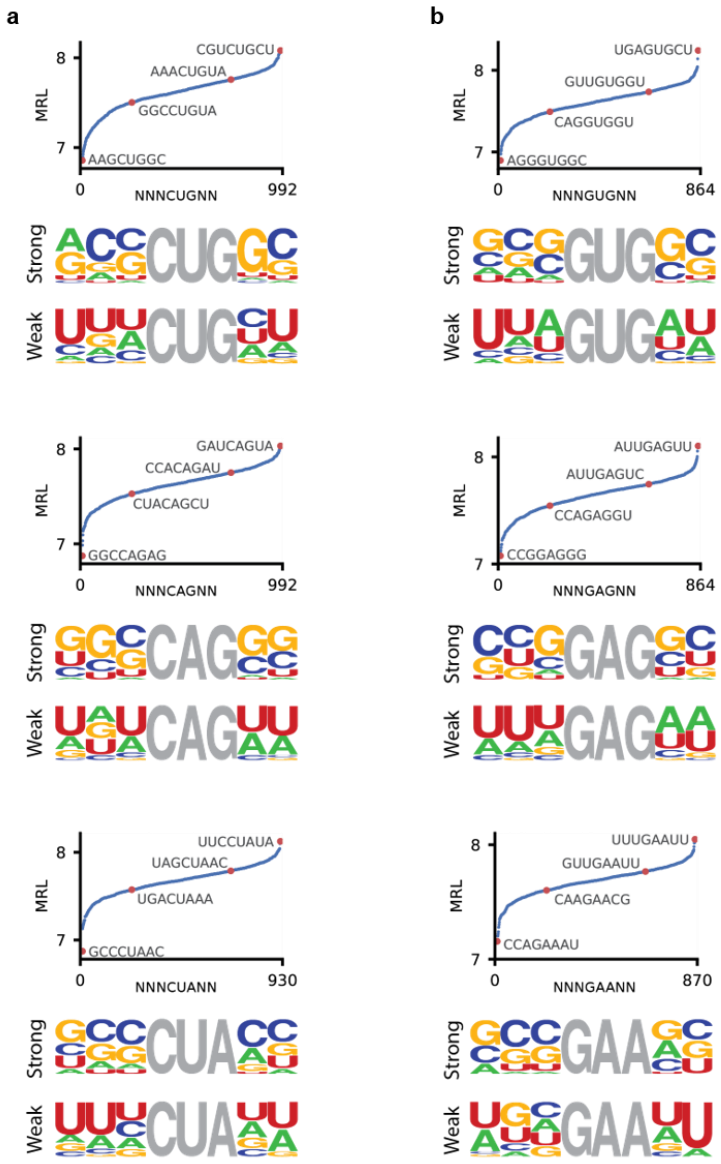
(a) The polysome profile and collected fractions of the eGFP unmodified RNA experiment #1 from two independent polysome profiling experiments. For all polysome profiles, the median number of processed fractions was 13. Some fractions were pooled to accommodate individual polysome peaks. **(b)** Out-of-frame upstream open reading frames (OOF uORFs) and upstream AUGs (uAUGs) cause reduced loading of ribosomes, compared to sequences without uAUGs (No uAUG). In-frame uAUGs (IF uAUGs) do not show a reduction in ribosome loading (p-values from two-sided t-test are shown). Purines at the -3 position of the Kozak show increased loading at the intended start codon, but reduced loading when preceding OOF uORFs and uAUGs. P-values from two-sided t-test of impact of purines vs. pyrimidines at -3 position in four groups were computed; No uAUG: $5e^{-103}$, IF uAUG: $8e^{-29}$, OOF uAUG: 0, OOF uORF: 0. For 16 box plots from left to right: (n = 7,008; min., 2.677; quartile (Q)1, 7.592; median, 7.876; Q3, 8.092; max., 9.253), (3,004; 2.941; 7.406; 7.687; 7.916; 9.389), (4,706; 2.871; 7.563; 7.855; 8.087; 9.214), (4,049; 3.238; 7.425; 7.712; 7.951; 8.824), (1,967; 2.678; 7.415; 7.79; 8.102; 9.403), (1,006; 3.011; 7.453; 7.846; 8.107; 8.727), (1,443; 2.694; 7.43; 7.808; 8.108; 8.95), (1,883; 2.862; 7.129; 7.671; 7.998; 8.835), (4,683; 0.479; 3.238; 3.702; 4.278; 6.988), (3,040; 2.484; 4.362; 5.319; 6.174; 8.126), (4,041; 1.71; 3.416; 3.993; 4.657; 7.592), (3,436; 2.379; 4.098; 4.999; 5.796; 8.224), (2,854; 2.427; 3.97; 4.501; 5.04; 7.215), (1,873; 2.419; 4.758; 5.875; 6.729; 8.316), (2,270; 2.625; 4.186; 4.76; 5.363; 7.48), (2,747; 1.605; 4.546; 5.479; 6.242; 7.948). **(c)** Boxplots showing the relationship between mean ribosome load and minimum free energy (MFE) were calculated by finding the MFE of 20,000 UTRs using Nupack. Ribosome loading increases as the predicted MFE increases (p-values from two-sided t-test were shown). For 7 box plots from left to right: (n = 406; min., 3.634; quartile (Q)1, 6.851; median, 7.221; Q3, 7.531; max., 8.449), (1,649; 2.871; 7.09; 7.415; 7.722; 8.744), (4,248; 2.677; 7.336; 7.634; 7.887; 8.948), (6,519; 3.273; 7.56; 7.811; 8.024; 9.386), (5,185; 3.469; 7.708; 7.921; 8.115; 9.032), (1,680; 3.118; 7.819; 8.017; 8.18; 8.837), (215; 7.011; 7.859; 8.049; 8.222; 8.612).



Supplementary Fig. 3

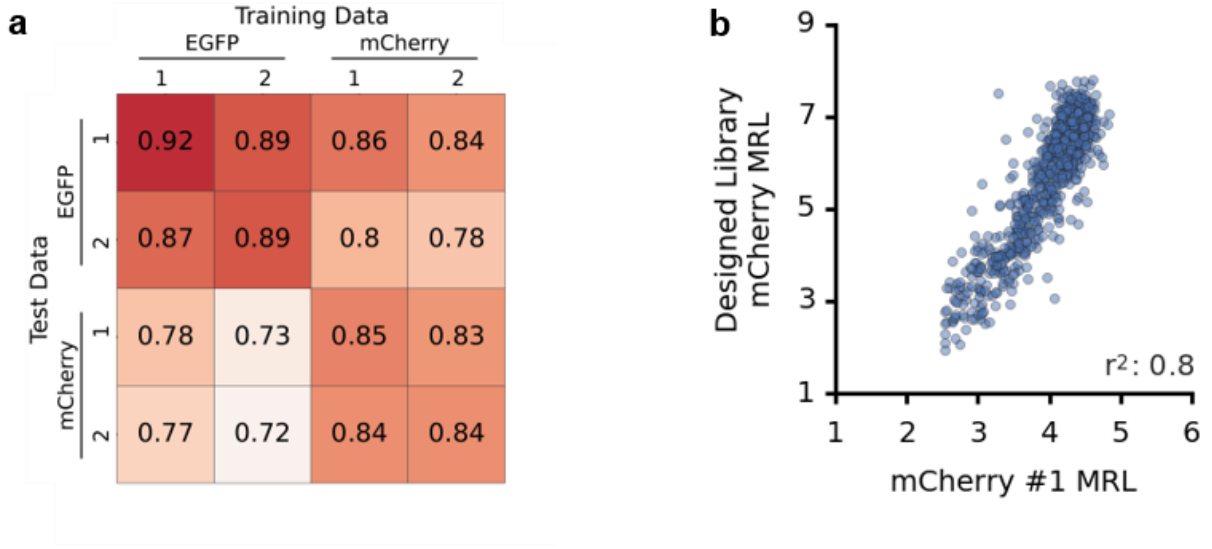
Effect of non-AUG translation initiation sites (TIS) on ribosome loading. When analyzing non-AUG TIS, all sequences with AUG were removed. (**a)** UTRs were grouped based on the presence of AUG, CUG, GUG, or UUG between positions -21 through -8 and by the TIS context in which they were found. All AUG TISs dramatically reduce ribosome loading when in an out-of-frame position and the extent of repression is dependent on the strength of the TIS context. For CUG and GUG, there is a minor reduction in ribosome loading when in the out-of-frame position and within a strong TIS context (p-values from two-sided t-test are shown). UUG shows no effect. For 6 box plots in AUG from left to right: (n = 6,693; min., 1.622; quartile (Q)1, 4.32; median, 5.271; Q3, 6.205; max., 8.384), (2,490; 1.802; 5.115; 6.591; 7.648; 9.15), (7,535; 1.725; 3.562; 4.177; 4.889; 8.316), (2,390; 2.586; 4.708; 6.515; 7.689; 9.695), (3,069; 1.678; 3.24; 3.752; 4.46; 8.308), (908; 2.667; 4.467; 6.353; 7.566; 8.718). For 6 box plots in CUG from left to right: (2,330; 3.355; 7.415; 7.755; 8.009; 8.945), (931; 3.491; 7.45; 7.755; 8.013; 9.189), (2,150; 3.549; 7.473; 7.772; 8.022; 8.763), (757; 3.335; 7.455; 7.792; 8.035; 9.253), (796; 2.871; 7.138; 7.515; 7.827; 8.621), (349; 3.846; 7.384; 7.73; 7.978; 8.564). For 6 box plots in GUG from left to right: (2,392; 2.677; 7.416; 7.755; 8.004; 8.763), (972; 4.61; 7.403; 7.73; 7.994; 8.664), (2,079; 2.871; 7.431; 7.744; 8.009;

8.678), (816; 3.355; 7.446; 7.763; 8.021; 8.681), (654; 3.29; 7.286; 7.66; 7.93; 8.952), (299; 3.702; 7.413; 7.72; 7.975; 8.588). For 6 box plots in UUG from left to right: (2,297; 3.491; 7.542; 7.842; 8.073; 8.952), (934; 3.946; 7.51; 7.846; 8.066; 8.788), (2,000; 2.314; 7.532; 7.834; 8.082; 8.681), (783; 4.109; 7.555; 7.861; 8.104; 9.189), (779; 3.84; 7.482; 7.802; 8.028; 8.613), (295; 4.576; 7.48; 7.772; 8.063; 8.66). (**b)** Presence of out-of-frame CAG and GAG, single base mismatches of CUG and GUG, does not cause reduced mean MRLs.



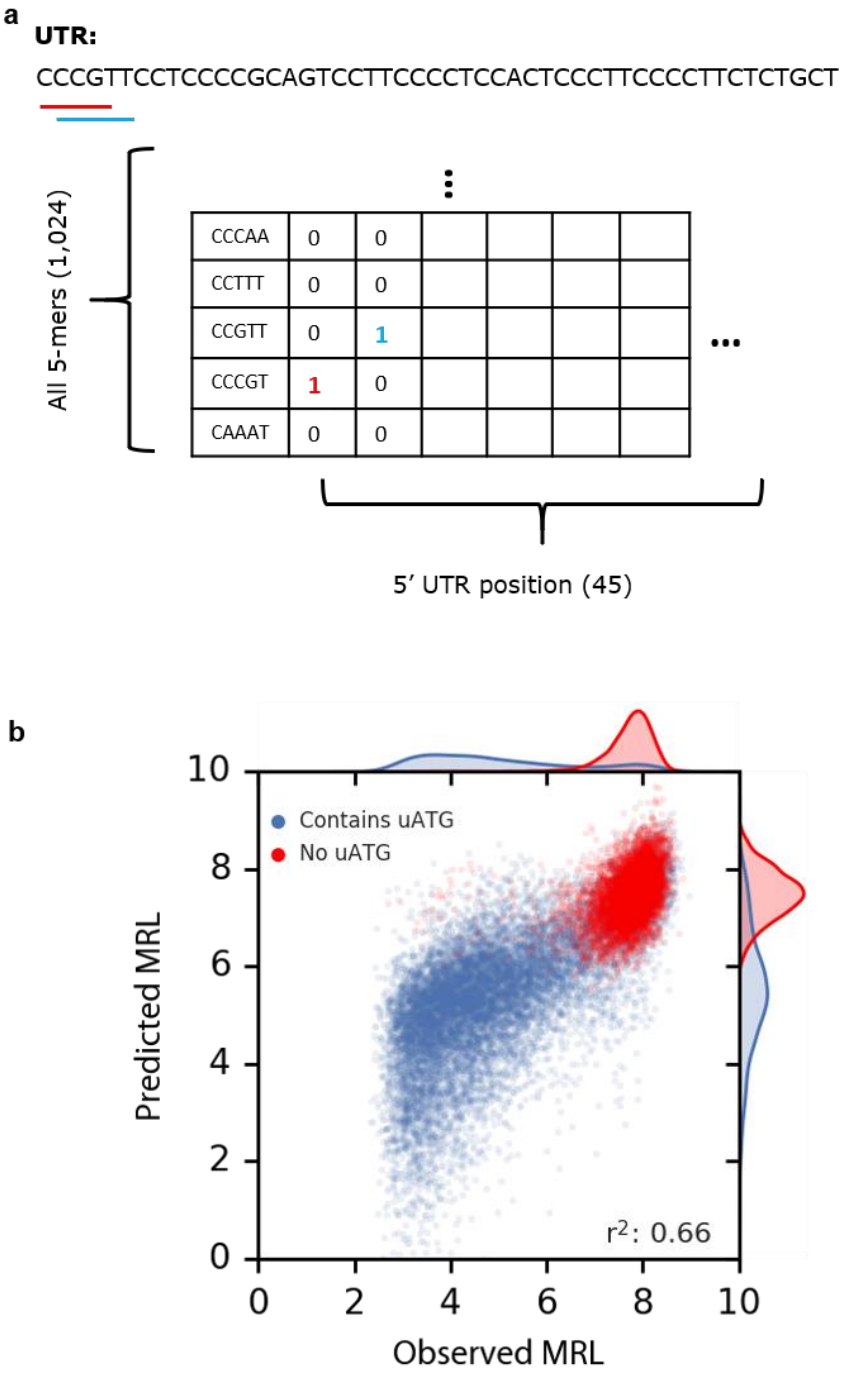
Supplementary Fig. 4

(a) The repressive strength of all out-of-frame variations of NNNCUGNN. The single-mismatch 3-mers CAG and CUA were included as controls. The "strong" (most repressive) TIS consensus sequence matches that of AUG – A/G at -3, CC at -2 and -1, and GC at +4 and +5. The "strong" and "weak" TIS sequences of CAG and CUA are GC-rich and AU-rich respectively, reflecting the repressive nature of GC-rich sequences rather than likelihood of translation initiation. When analyzing non-AUG TIS, all sequences with AUG were removed. **(b)** The repressive strength of all out-of-frame variations of NNNGUGNN. The 3-mers GAG and GAA were included as controls. The "strong" TIS consensus sequence does not match the pattern of the consensus sequence of AUG and CUG. Like the control 3-mers, the "strong" and "weak" sequences are simply GC-rich and AU-rich, respectively. When analyzing non-AUG TIS, all sequences with AUG were removed.



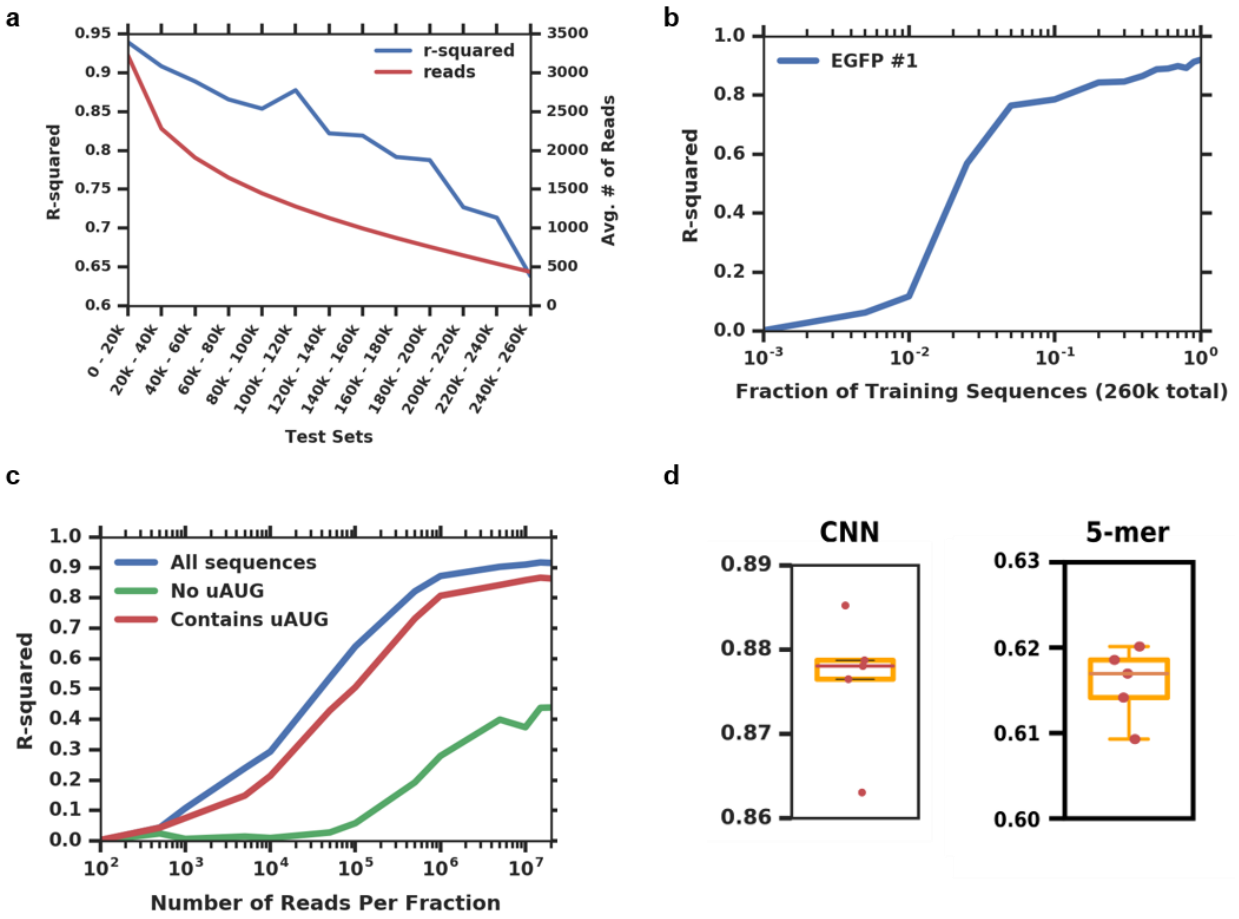
Supplementary Fig. 5

Model generalization between coding sequences. **(a)** A similar randomized 5′ UTR library was made but with eGFP replaced with mCherry (200k unique sequences – no 5′ UTRs are shared between the two libraries). Two independent experiments were evaluated via polysome profiling and modeling. Model performance – r^2 within each box of the heatmap - was cross-tested between all CDS / replicate test data. The mCherry models generalize well with eGFP, while the eGFP models do to a lesser extent. This may be due to differences between polysome profiling experimental conditions (**Online Methods**, *n* = 20,000). **(b)** Polysome profiling of 2,000 UTRs from the eGFP library but with eGFP replaced with mCherry. The MRL between the eGFP and mCherry data are highly correlated (r^2: 0.8, *n* = 2,000).



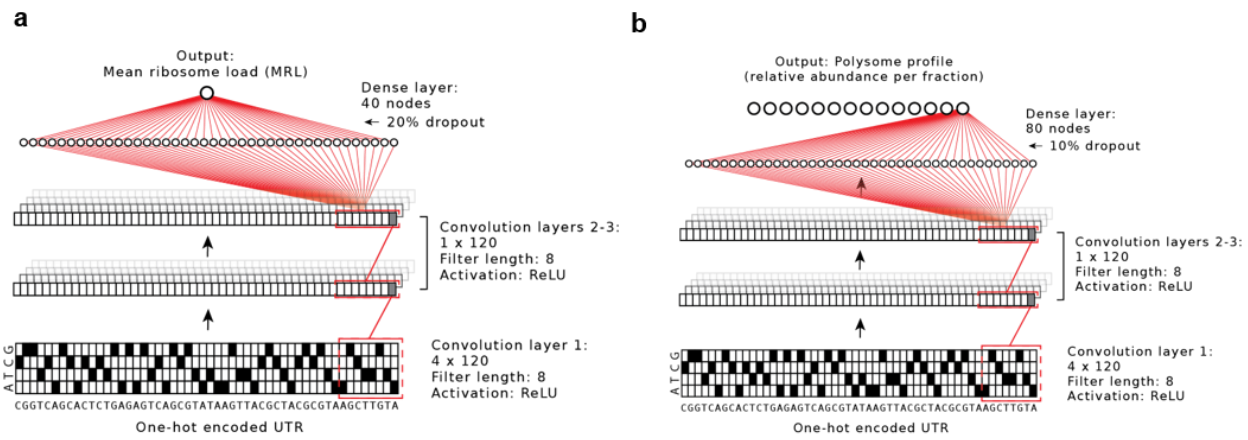
Supplementary Fig. 6

(a) Position-specific 5-mer linear model. UTRs were encoded such that 5-mers and the position in which they occur in a sequence serve as features for linear regression. Position information is especially important for uATGs where placement relative to the CDS determines whether they are in-frame or out-of-frame. **(b)** k-mer models ranging from 1 - 7 were tested and the 5-mer model showed the best performance of the set (r^2: 0.66, *n* = 20,000).



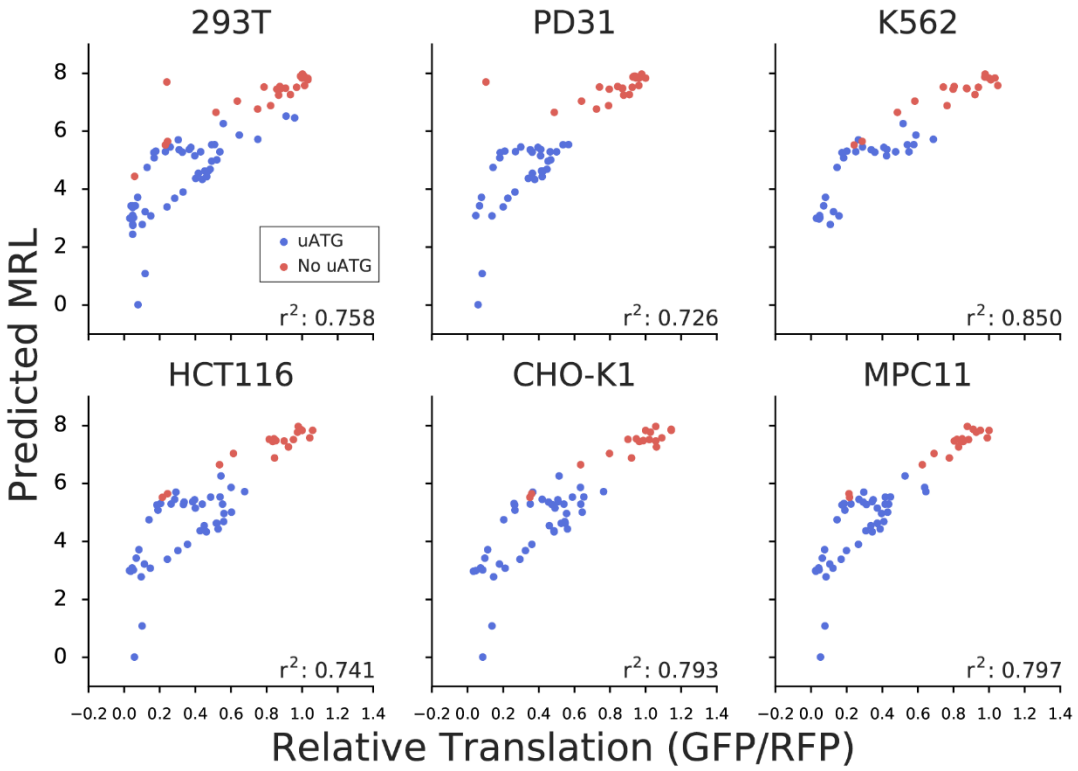
Supplementary Fig. 7

(a) The number of sequencing reads per UTR in the test set affects the observed model performance. Sequences were sorted high to low by the total number of sequencing reads across all polysome fractions. A sliding window of 20k sequences were used as test sets and the remaining sequences were used for training (n = 20,000). **(b)** Model quality decreases as fewer UTRs are used for training. The full set consists of 260,000 unique 5′ UTRs. A dramatic rise in model performance occurs from training on 6,500 (2.5%) to 26,000 (10%) sequences, which is likely due to the model learning the rules of uAUGs and uORFs (n = 20,000). **(c)** Evaluation of model performance as a function of the number of reads per fractions. Fraction reads were down-sampled and used for training and testing. Initial read counts per fraction ranged from 21M to 33M except for fraction 13 which had 10M. Sequences without uAUGs require more training examples than those with uAUGs (n = 20,000). **(d)** Five models were trained and tested independently through CNN (left) and 5-mer linear model (right) using randomly selected subsets of sequences. The median r-squared for CNN is 0.878 and the median r-squared for 5-mer linear model is 0.617. For 2 box plots from left to right: (n = 5; min., 0.863; quartile (Q)1, 0.876; median, 0.878; Q3, 0.879; max., 0.885), (5; 0.604; 0.612; 0.613; 0.614; 0.616).



Supplementary Fig. 8

(a) CNN architecture to predict mean ribosome load (MRL) from 50-mer 5′ UTR sequences. **(b)** CNN architecture for predicting the polysome profile of a given 5′ UTR. The model differs from the MRL model in the number of nodes in the dense layer (80 rather than 40), the dropout from the dense layer (10% rather than 20%), and the single final linear node is replaced by 14 linear nodes corresponding to each fraction collected during polysome profiling. The output values sum to 1 and represent the relative abundance per fraction.



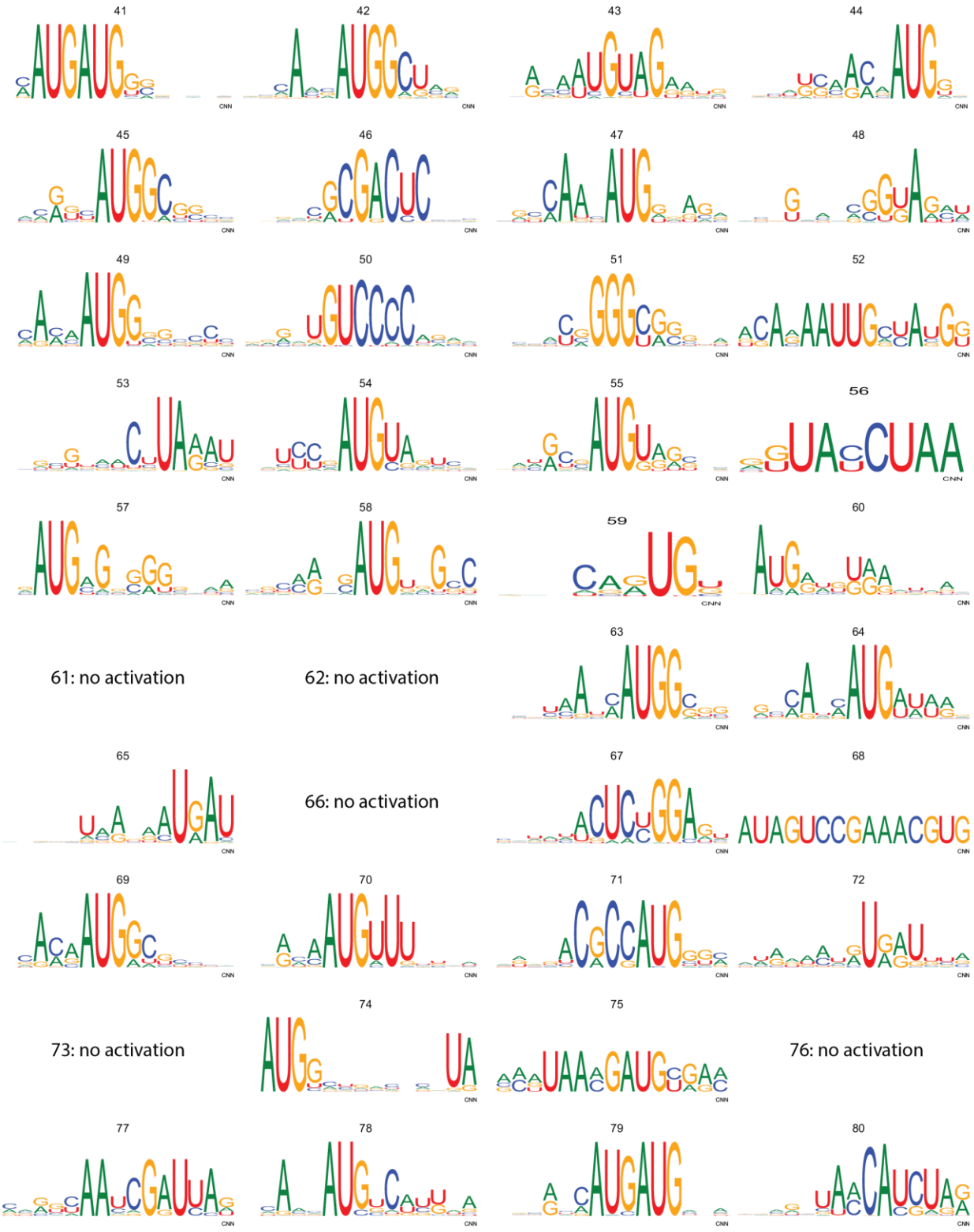
Supplementary Fig. 9

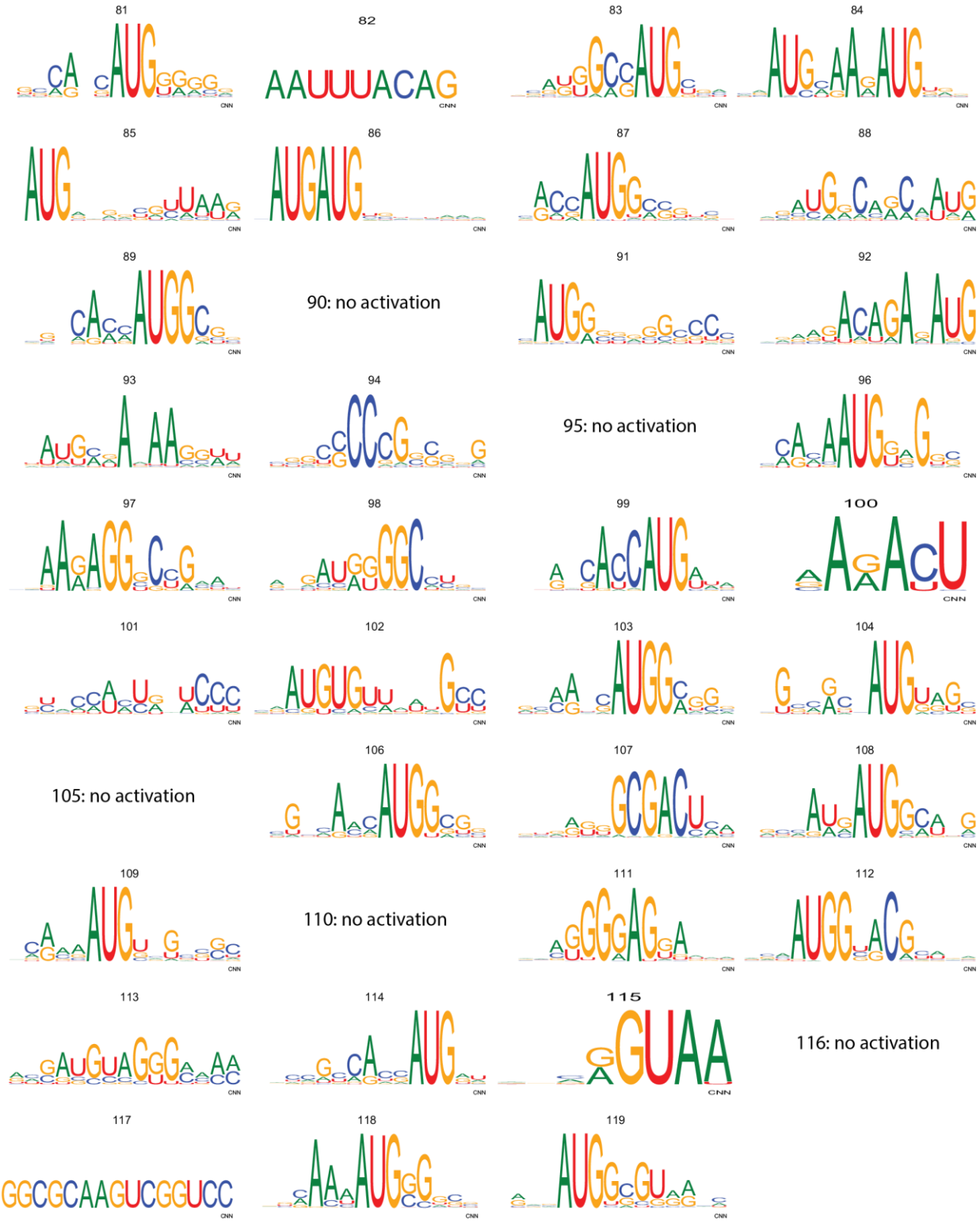
Model performance on 77 5' UTRs designed and experimentally characterized by Ferreira et al. (Proceedings of the National Academy of Sciences **110**, 11284-11289 (2013)). Flow cytometry was performed after delivery of reporter constructs to six different cell lines, including human embryonic kidney cells (293T), mouse pre-B lymphocytes (PD31), human chronic myelogenous leukemia cells (K562), human colon cancer cells (HCT116), Chinese hamster ovary cells (CHO-K1) and mouse plasmacytoma (MPC11). Only 5' UTRs upstream of coding sequence begin with 'ATGG' were used for model testing. Sequences were dropped if fluorescence data was missing for some cell lines. UTR length ranged from 3 nt to 47 nt, and sequences were zero-padded on the 5' end to generate 50 nt input sequences for our model. UTRs with no upstream ATG (No uATG) are shown in red and UTRs with upstream ATG (uATG) are shown in blue. Our model's prediction on mean ribosome load and the translation level measured by fluorescence assay were highly correlated across all cell lines (r^2: 0.73-0.85, 293T: n = 77, PD31: n = 55, K562: n = 46, HCT116: n = 46, CHO-K1: n = 46, MPC11: n = 46).

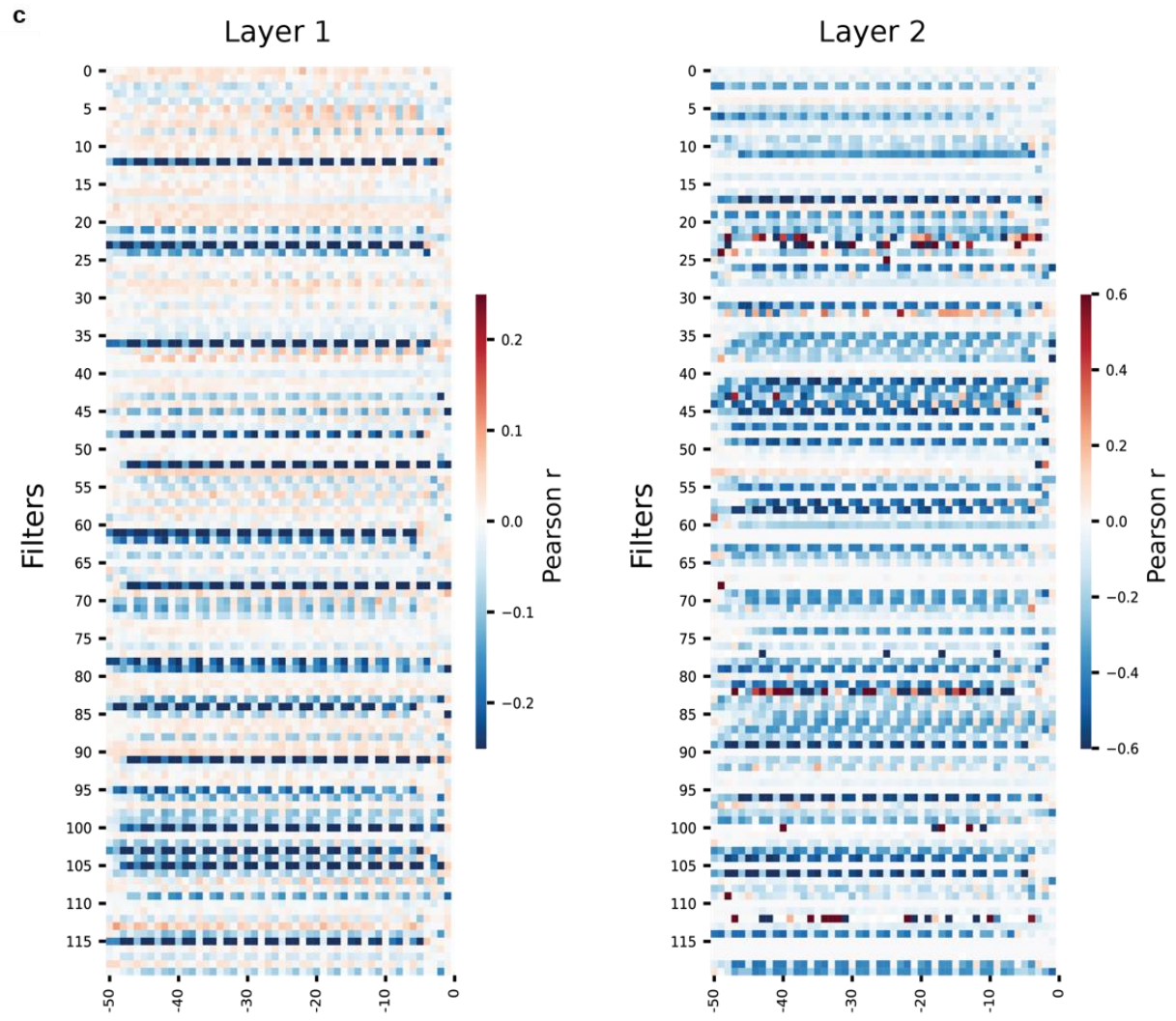

a
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
bits
CNN


60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
bits
CNN


b
0
1
2
4
5
AAAAAAUAU
6
7
8
9
10
11
12
13
ACAGAAUACCGAGUA
14
15
16
17
18
19
20
21
22
23
UAAGAUGCCAUCGUG
24
25
26
27
28
29: no activation
30
31
32
33
34
35
36
37
38
39
40
CNN

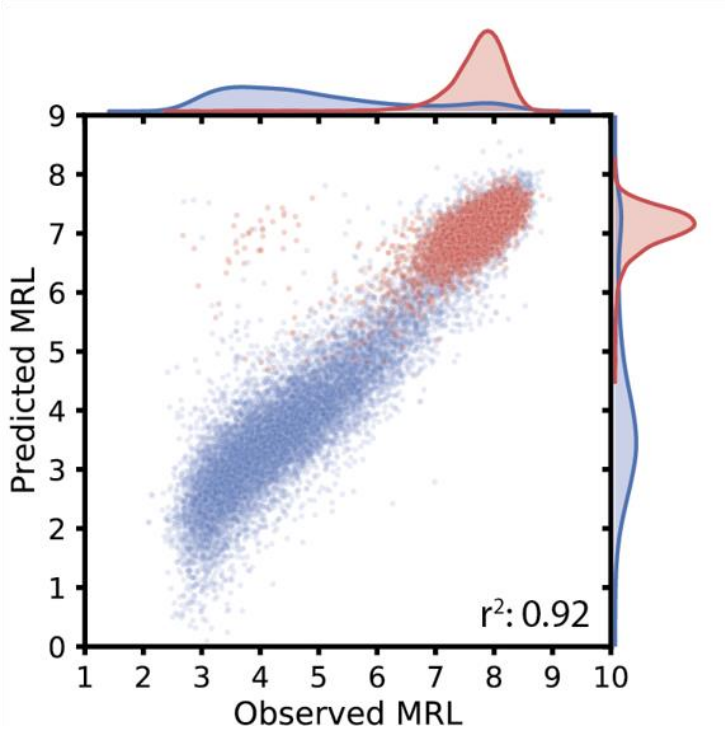

41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61: no activation
62: no activation
63
64
65
66: no activation
67
68
AUAGUCCGAAACGUG
69
70
71
72
73: no activation
74
75
76: no activation
77
78
79
80
CNN


81
82
AAUUUACAG
83
84
85
86
87
88
89
90: no activation
91
92
93
94
95: no activation
96
97
98
99
100
101
102
103
104
105: no activation
106
107
108
109
110: no activation
111
112
113
114
115
116: no activation
117
GGCGCAAGUCGGUCC
118
119



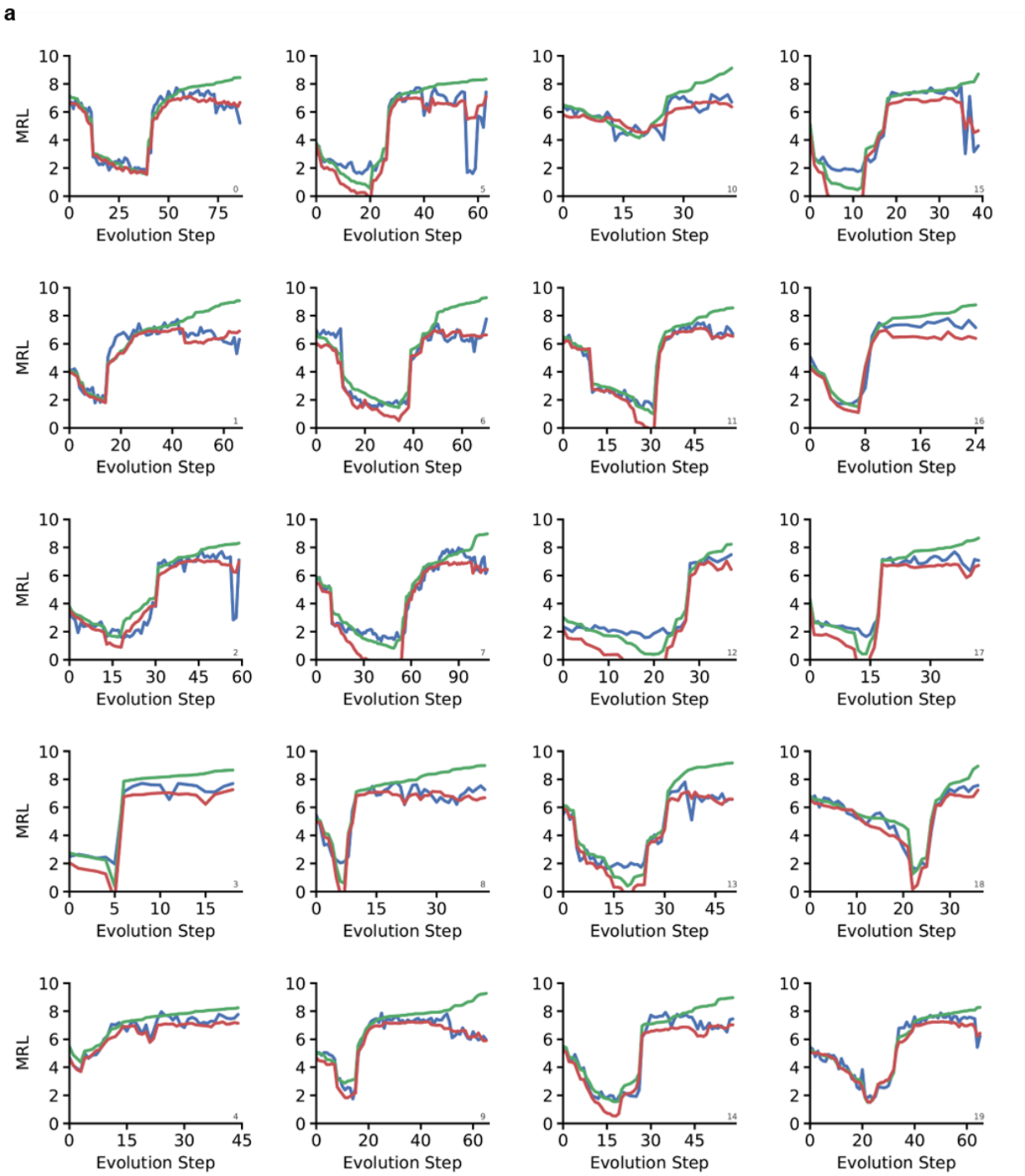
Supplementary Fig. 10

(a) Visualization of filters from the first convolution layer of the eGFP model (before retraining) (Figure 2). For each filter we collected the top 2,000 8-mers in the eGFP library that showed maximal activation. These were then used to calculate position weight matrices and visualized as sequence logos. Some filters had fewer than 2,000 8-mers that showed activation. **(b)** Visualization of filters from the second convolution layer of the eGFP model (before retraining) (Figure 2). For each filter we collected the top 2,000 8-mers in the eGFP library that showed maximal activation. These were then used to calculate position weight matrices and visualized as sequence logos. Some filters had fewer than 2,000 8-mers that showed activation or no activation at all. **(c)** For convolution layers one and two, the correlation between filter activation and MRL for each filter at each position of the 5′ UTRs. If UTRs that show high filter activation have low MRLs then the two are negatively correlated. This shows the importance of each filter at each position for predicting MRL.

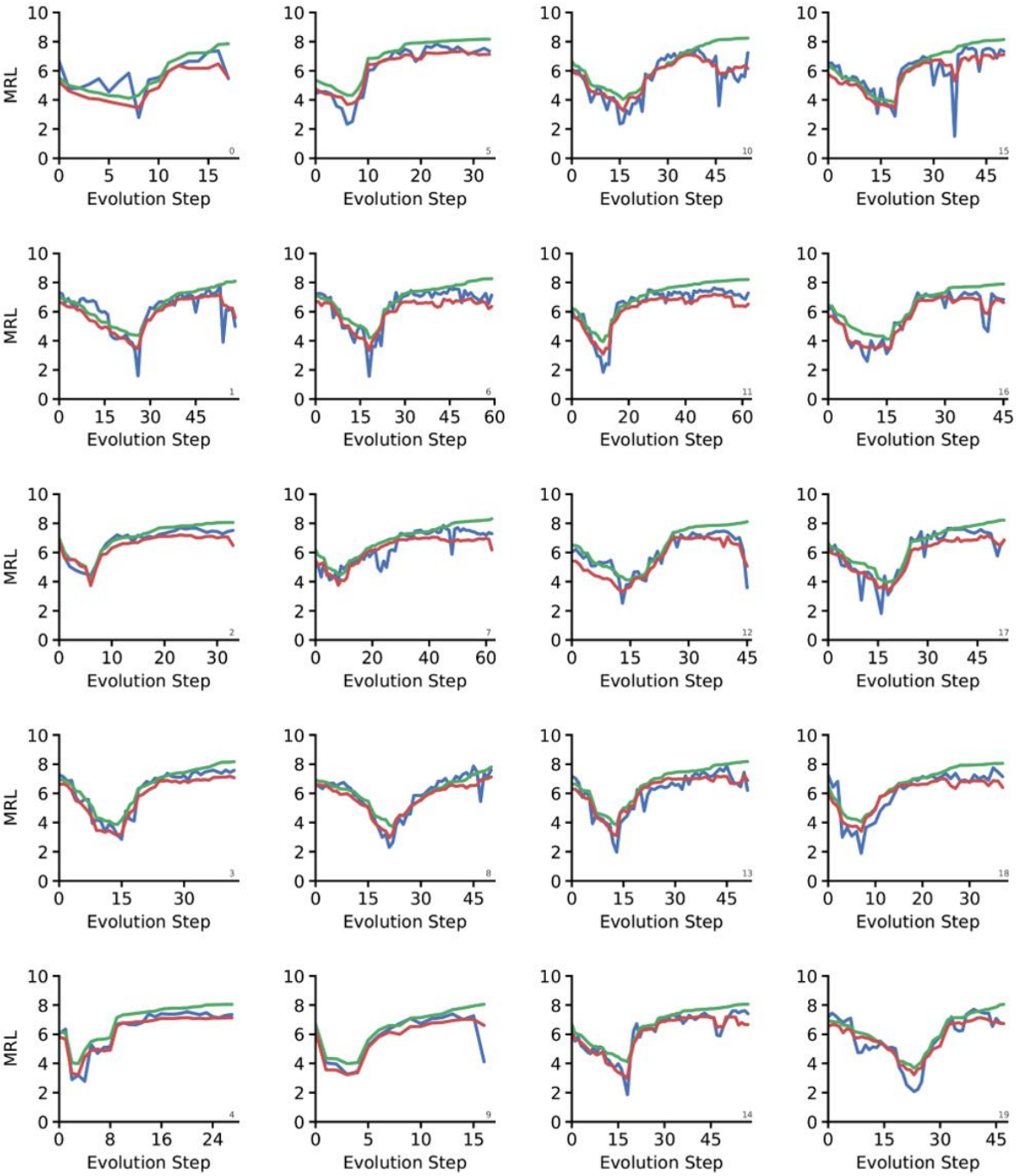


Supplementary Fig. 11

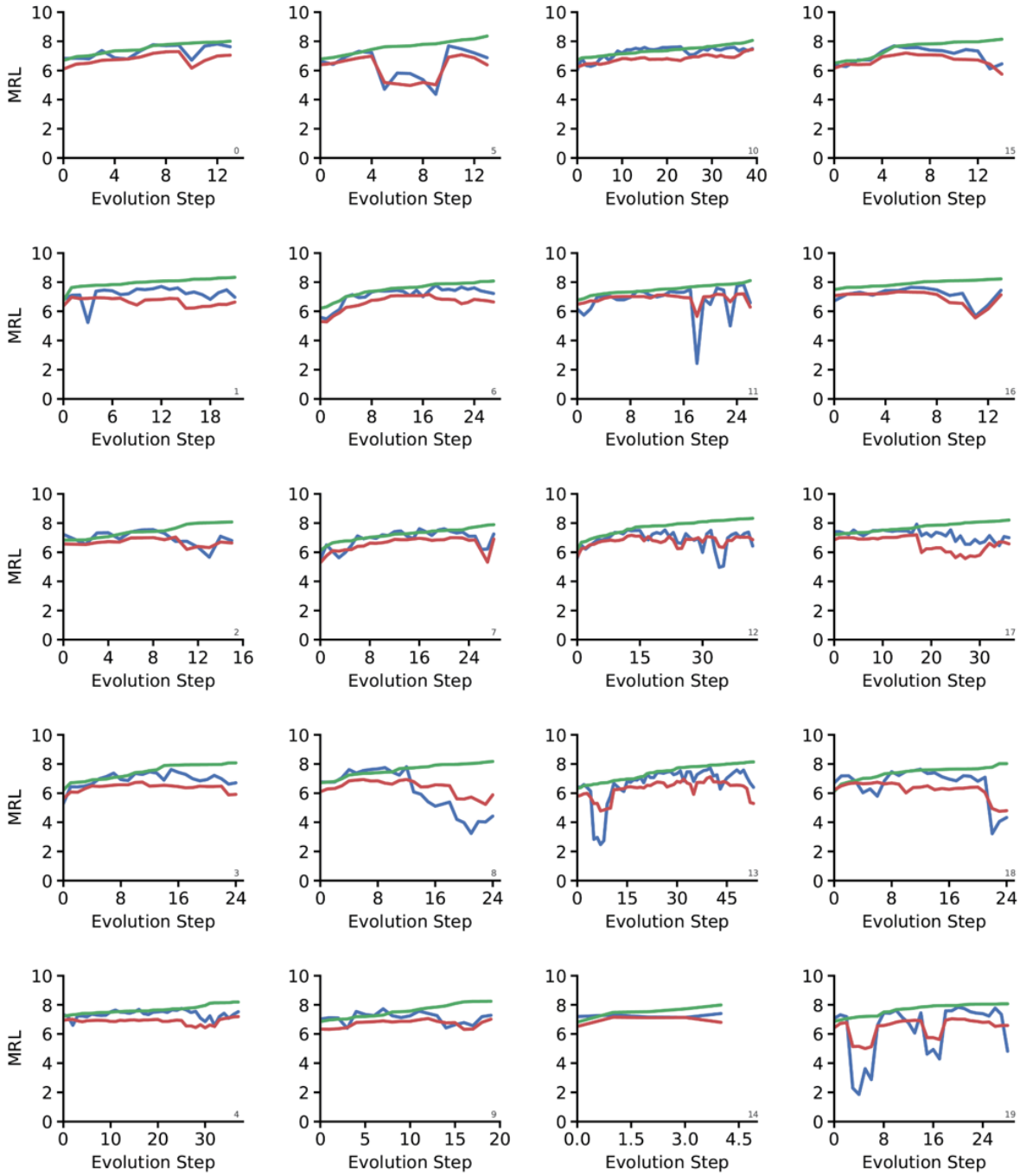
Performance of an early iteration of a mean ribosome load model that used a slightly different network architecture (**Online Methods**) relative to the main MRL model presented in **Figure 2.** This model was used with the genetic algorithm to design sequences for targeted expression (*n* = 20,000).


a
MRL
Evolution Step

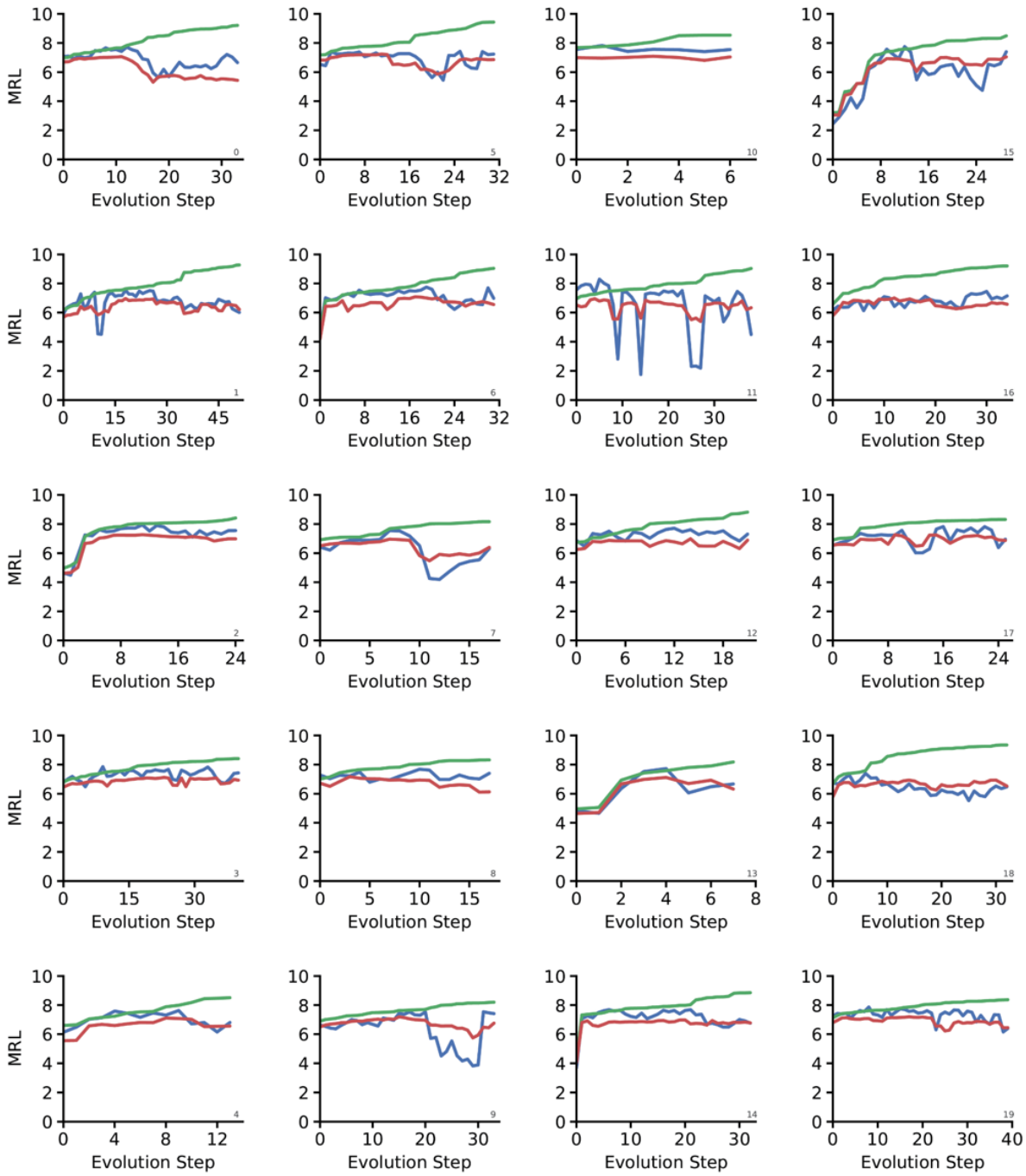
b

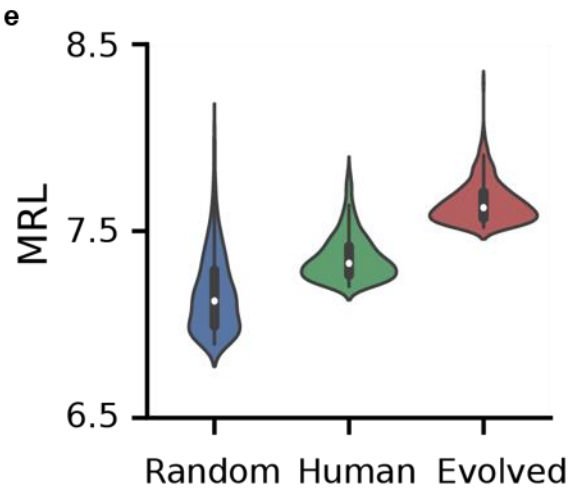


c



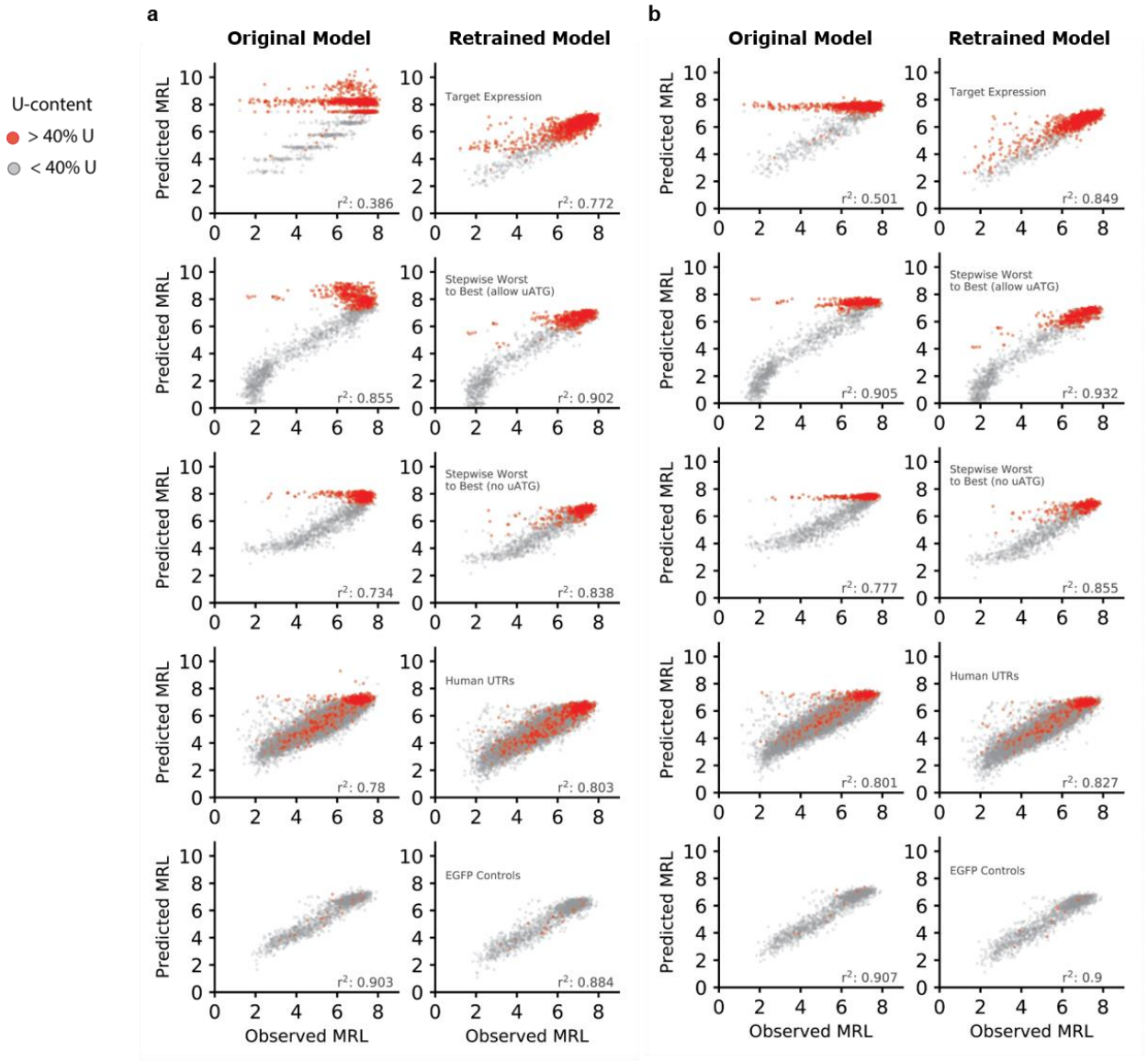
d





Supplementary Fig. 12

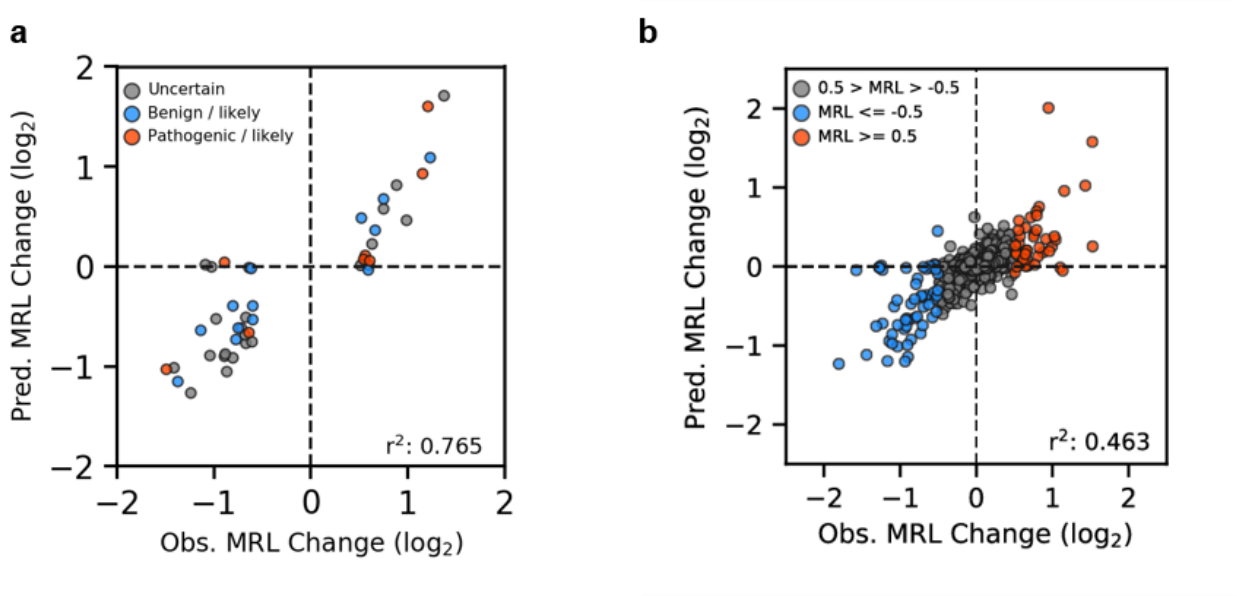
(a-d) Following the sequence evolution of our genetic algorithm. Randomly initialized 50-mers were first selected for low ribosome loading (800 iterations) followed by selection for high ribosome loading (800 iterations) – UTRs were either permitted to use uATGs **(a)** or not **(b)**. Two other subsets were randomly initialized but only evolved for high ribosome loading – UTRs were either permitted to use uATGs **(c)** or not **(d)**. All unique sequences along a UTR's evolution were synthesized (sub-libraries in the designed library) and tested via polysome profiling. The retrained model (red) closely matches the observed MRL (blue) and performs significantly better than the original model (green) that was used for evolving the sequences. **(e)** The observed MRLs of the top 500 UTRs in terms of MRL from the random library, human library, and the evolved library. Overall, sequences from the evolved library show higher ribosome loading. For 3 violin plots from left to right: (n = 1,000; min., 6.468; quartile (Q)1, 6.702; median, 6.895; Q3, 7.126; max., 8.061, s.d., 0.292), (1,000; 7.012; 7.102; 7.203; 7.327; 7.825; 0.16), (1,000; 7.362; 7.434; 7.52; 7.626; 8.292; 0.137).



Supplementary Fig. 13

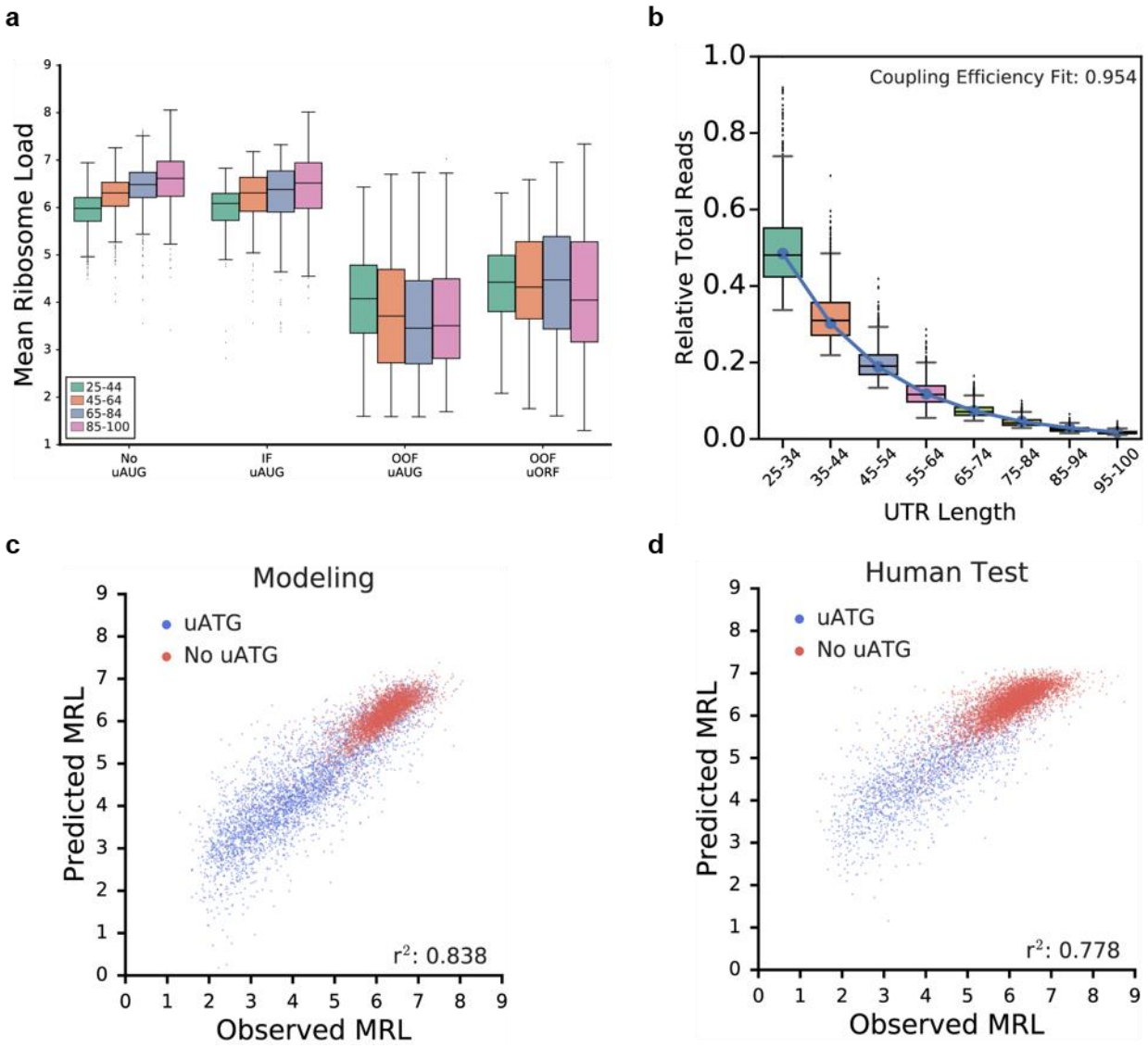
(a) Comparing the performance of the original model used for evolving new sequences (left) to the retrained model performance (right) on subsets of the designed library. The three evolved UTR groups showed the most dramatic improvement with the retrained model. Notably, the original model made predictions of high MRL for many sequences that, when measured, showed moderate to low ribosome loading. These sequences are uridine-rich (red). The retrained model was able to predict MRL more accurately for these sequences. The retrained model also showed a small increase in performance when predicting MRL for truncated human UTRs (r^2 improvement of 0.023). For the control UTRs from the eGFP library, the retrained model showed a slight performance decrease (-0.019). These results indicate that the original model had difficulty predicting MRLs for UTRs with sequences that are unlikely to be found in a randomized library, such as UTRs with high poly-U content. **(b)** The model from Figure 2, which has an improved network architecture and shows higher accuracy than the model used for sequence evolution, was retrained in the same manner as the evolution model in Figure 3. The model performance

significantly increased with all sub-libraries except for the eGFP control UTRs which decreased minimally (-0.007). (Sample sizes are identical for 4 columns in a and b. Target Expression: n = 2,146, Stepwise Worst to Best (allow uATG): n = 1,987, Stepwise Worst to Best (no uATG): n = 1,681, Human UTRs: n = 14,128, EGFP Controls: n = 1,626.)



Supplementary Fig. 14

(a) Predicting the change in mean ribosome load (MRL) between an SNV and its common sequence with $\log_2$ differences greater than 0.5 or less than -0.5 (n = 45). **(b)** Model prediction of 'spontaneous' UTR variants created by oligo synthesis errors. 103 out of 2,308 showed MRL changes greater than 0.5 or less than -0.5 ($\log_2$) (n = 2,308).



Supplementary Fig. 15

(a) Out-of-frame upstream open reading frames (OOF uORFs) and upstream AUGs (uAUGs) cause reduced loading of ribosomes while in-frame (IF) do not. When sequences do not contain any uAUGs (No uAUGs) or have in-frame uAUGs (IF uAUGs), longer UTR sequences show increased ribosome loading. For 16 box plots from left to right: (n = 1,256; min., 4.494; quartile (Q)1, 5.71; median, 5.982; Q3, 6.209; max., 6.944), (997; 4.015; 6.027; 6.308; 6.532; 7.261), (686; 3.549; 6.211; 6.485; 6.741; 7.643), (392; 3.417; 6.235; 6.613; 6.974; 8.056), (145; 2.82; 5.728; 6.086; 6.299; 6.829), (167; 3.558; 5.921; 6.307; 6.638; 7.179), (214; 3.386; 5.905; 6.379; 6.77; 7.324), (166; 3.37; 5.983; 6.518; 6.943; 8.011), (389; 1.595; 3.353; 4.078; 4.786; 6.433), (449; 1.591; 2.724; 3.71; 4.696; 6.704), (520; 1.588; 2.704; 3.457; 4.456; 6.737), (438; 1.695; 2.819; 3.509; 4.498; 7.025), (210; 2.082; 3.807; 4.425; 4.992; 6.306), (387; 1.76; 3.651; 4.32; 5.281; 6.587), (580; 1.606; 3.439; 4.475; 5.389; 6.954), (604; 1.298; 3.163; 4.048; 5.278; 7.342). **(b)** RNA sequencing read depth decreases with longer UTRs. All read counts are normalized to the maximum read count observed. All oligos used for 5'UTR construction were synthesized on the same array and the observed decrease is consistent with a coupling efficiency c < 1. A fit to the

expression N^c where N is the length and c the coupling efficiency results in c = 0.954 as a reasonable estimate of the coupling efficiency in oligo array synthesis (blue line). For 8 box plots from left to right: (*n* = 1,000; min., 0.337; quartile (Q)1, 0.424; median, 0.841; Q3, 0.552; max., 1), (1,000; 0.219; 0.271; 0.31; 0.357; 0.688), (1,000; 0.134; 0.168; 0.191; 0.22; 0.419), (1,000; 0.055; 0.097; 0.116; 0.139; 0.286), (1,000; 0.048; 0.062; 0.071; 0.083; 0.165), (1,000; 0.029; 0.036; 0.042; 0.05; 0.1), (1,000; 0.015; 0.021; 0.025; 0.029; 0.065), (600; 0.011; 0.014; 0.016; 0.02; 0.046). **(c)** 7,600 random 5'UTR sequences with lengths from 25 to 100 nucleotides (100 sequences at each length) were tested using the generalized model. The model was trained on a library of 76,319 random sequences spanning the same length range (r^2: 0.838, *n* = 7,600). **(d)** 7,600 human native 5'UTR sequences with varying lengths from 25 to 100 nucleotides (100 sequences at each length) are tested using the generalized model (r^2: 0.778, *n* = 7,600).

k-mer	r^2
1	0.1273
2	0.3524
3	0.6283
4	0.6397
5	0.6439
6	0.5606

Supplementary Table 1.

Linear regression modeling with position-specific k-mers. 1-mers to 6-mers were evaluated using the same training and test sets that were used in building the CNN model. 5-fold cross validation found the best regularization (ridge) parameters to limit overfitting (n = 20,000).

Gene	Phenotype	Significance	Obs. Diff. (log_2)	dbSNP	uAUG Effects
GNPAT	Rhizomelic chondrodysplasia punctata	Uncertain	-1.41	rs201907247	Yes
CSTB	Unverricht-Lundborg syndrome	Uncertain	0.56	rs776181852	No
GALT	Deficiency of UDPglucose-hexose-1-phosphate uridylyltransferase	Pathogenic / Likely	0.56	rs111033654	No
MLH1	not specified	Uncertain	-0.9	rs1016433173	Yes
MSH6	not specified	Uncertain	-0.81	rs1064793670	Yes
BBS7	Bardet-Biedl syndrome	Uncertain	0.62	rs757523715	No
GALT	Deficiency of UDPglucose-hexose-1-phosphate uridylyltransferase	Pathogenic / Likely	0.54	rs111033656	No
GALT	Deficiency of UDPglucose-hexose-1-phosphate uridylyltransferase	Pathogenic / Likely	0.6	rs111033656	No
COA6	not specified	Benign / Likely	-0.62	rs73099933	No
PDE6C	Achromatopsia, Cone-Rod Dystrophy, Recessive	Uncertain	-0.72	rs374900090	Yes
TCTN3	not specified	Benign / Likely	-0.81	rs41291572	Yes
HSPB1	Charcot-Marie-Tooth, Type 2, Distal hereditary motor neuronopathy	Uncertain	0.59	rs372833436	No
TARS2	not specified	Benign / Likely	1.23	rs201336268	Yes
ZMPSTE24	Lethal tight skin contracture syndrome, Mandibuloacral dysplasia	Uncertain	-0.67	rs200527699	Yes
SMAD4	not specified	Benign / Likely	-0.61	rs1057523754	Yes
CHRNA4	Autosomal dominant nocturnal frontal lobe epilepsy	Benign / Likely	-0.77	rs200259564	Yes
TMEM127	Pheochromocytoma	Pathogenic / Likely	-1.5	rs121908813	Yes
PNPO	Pyridoxal 5'-phosphate-dependent epilepsy	Uncertain	-0.98	rs886053100	Yes
SMAD3	Loeys-Dietz syndrome 3	Pathogenic / Likely	1.21	rs587776882	No
LZTR1	not specified	Benign / Likely	-1.14	rs370616172	Yes
CTSA	Combined deficiency of sialidase AND beta galactosidase	Benign / Likely	-1.38	rs116893852	Yes
TP53	Sarcoma	Uncertain	-0.64	rs137852791	No
CPOX	Hereditary coproporphyria	Uncertain	-0.89	rs867711777	Yes
MPDU1	Congenital disorder of glycosylation	Uncertain	0.75	rs370389790	No
PEX12	not specified	Uncertain	-1.04	rs727504080	Yes
MLH1	Hereditary cancer-predisposing syndrome, Lynch syndrome	Uncertain	-0.68	rs587779001	Yes
SMAD3	not specified	Uncertain	1.37	rs863223756	Yes
SLX4	Fanconi anemia	Uncertain	-0.61	rs113023461	Yes
POLE	not specified	Uncertain	-0.68	rs1064796567	Yes
MAP2K2	not specified	Benign / Likely	0.66	rs1057520422	Yes
NBN	not specified	Benign / Likely	-0.75	rs730881843	Yes
SYNE2	Emery-Dreifuss muscular dystrophy	Benign / Likely	0.52	rs199566869	Yes
UQCRB	not specified	Benign / Likely	0.75	rs373747569	Yes
PDHX	Pyruvate dehydrogenase complex deficiency	Benign / Likely	-0.6	rs2956112	Yes
HSPB1	Charcot-Marie-Tooth, Type 2, Distal hereditary motor neuronopathy	Uncertain	0.51	rs199602956	No
RPL5	Diamond-Blackfan anemia	Uncertain	-0.87	rs376208311	Yes
ETHE1	Ethylmalonic encephalopathy, not specified	Uncertain	0.98	rs138958351	Yes
FOXRED1	not specified	Uncertain	-1.24	rs778239850	Yes
MKKS	Bardet-Biedl syndrome, McKusick Kaufman syndrome	Uncertain	-1.03	rs886056499	No
TTC19	not provided	Pathogenic / Likely	-0.9	rs769078093	No
PHEX	Familial X-linked hypophosphatemic vitamin D refractory rickets	Uncertain	-1.09	rs1057515841	No
C19orf12	Neurodegeneration with brain iron accumulation 4	Uncertain	0.88	rs186970109	Yes
SMPD1	Niemann-Pick disease, type A	Pathogenic / Likely	-0.64	rs875989837	Yes
PRPH2	Choroidal Dystrophy, Cone-Rod Dystrophy, Fundus albipunctatus,Vitelliform macular dystrophy	Benign / Likely	0.59	rs114062933	No
ACADM	Medium-chain acyl-coenzyme A dehydrogenase deficiency	Pathogenic / Likely	1.15	rs1057516778	Yes

Supplementary Table 4.

45 ClinVar variants with MRL changes greater than 0.5 or less than -0.5 (log_2). 30 variants cause significant MRL changes due to insertion or deletion of uAUG while the other 15 variants impact ribosome loading through other mechanisms.

Primer #	Sequence (5′ to 3′)
220	gacgtgtgctcttccgatctNNNNNNNNNNgtctgggtgccctcgta
252	atagggacatcgtagagagtcgtacttaNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNatgcctcccgagaagaagatc
253	cacgctcttgatcttcttctcgggaggcat
254	ttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttcaaacaacagatggctggca
282	atagggacatcgtagagagtcgtacttaNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNatgggcgaattaagtaagggc
283	acagctcctcgcccttacttaattcgccca
289	gacgtgtgctcttccgatctNNNNNNNNNNagatgaacttcagggtcagc

Supplementary Table 5.
Primers used for library construction, IVT mRNA template, and reverse transcription.

Supplementary Note 1 – MRL Calculation

In order to deal with the artifacts that different total reads in each fraction for each UTR may bring, we need to normalize the reads to compute out MRL for each UTR variant. Each UTR variant will have multiple reads depend on how many fractions are collected. For each UTR-n, it has reads from R_{n0}, R_{n1}, …, to R_{nm} etc., where m represents the number of ribosomes presented in that peak in the polysome profile. Firstly, relative reads for each UTR in the fraction (Relative-R_{nm}) will be computed in order to normalize differences in total read counts between fractions (**Eq. 1**). Then, the normalized reads (Normalized-R_{nm}) for each UTR will be computed based on Relative – R_{nm} in order to normalize differences in total read counts between UTRs (**Eq. 2**). Finally, MRL will be computed as the sum of product of normalized-R_{nm} with its corresponding number of ribosomes m **(Eq. 3**).

$$\text{Relative-R}_{nm} = \frac{\text{R}_{nm}}{\sum_n \text{R}_{nm}} \quad \textbf{(Eq. 1)}$$

$$\text{Normalized-R}_{nm} = \frac{\text{Relative-R}_{nm}}{\sum_m \text{Relative-R}_{nm}} = \frac{\frac{\text{R}_{nm}}{\sum_n \text{R}_{nm}}}{\sum_m \frac{\text{R}_{nm}}{\sum_n \text{R}_{nm}}} \quad \textbf{(Eq. 2)}$$

$$\text{MRL}_n = \sum_m (\text{Normalized-R}_{nm} \times m) = \sum_m \left(\frac{\frac{\text{R}_{nm}}{\sum_n \text{R}_{nm}}}{\sum_m \frac{\text{R}_{nm}}{\sum_n \text{R}_{nm}}} \times m \right) \quad \textbf{(Eq. 3)}$$