Toward predictable 5'UTRs in *Saccharomyces cerevisiae*: Development of a yUTR calculator

Supporting Information

Thomas Decoene[a], Gert Peters[a], Sofie L. De Maeseneire[b], Marjan De Mey[a,1]

[a] *Centre for Synthetic Biology, Ghent University, Coupure links 653, 9000 Ghent, Belgium*

[b] *Centre for Industrial Biotechnology and Biocatalysis, Ghent University, Coupure links 653, 9000 Ghent, Belgium*

Contents

[1] **Corresponding author:** marjan.demey@ugent.be, Centre for Synthetic Biology, Ghent University, Coupure links 653, 9000 Ghent, Belgium

Table S.1: Overview of all strains and plasmids used in this study. The predicted values of the protein abundances (PPA) are generated by the PLS regression model. PPA values that were predicted via reverse engineering are indicated with an asterisk. The plasmid backbone is a low copy backbone with a *CEN/ARS4* ori, the *URA3* auxotrophic marker and the *TEF2p-mCherry-PGK1t* transcription unit. All yeast strains were derived from the S288c laboratory strain BY4742.

Strain	Genotype	PPA	Plasmid
BY4742	*Matα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*	-	-
sTemplate1	RPL8Ap-nativeRPL8A_UTR-yECitrine-ADH1t	5.50*	pTemplate1
sTemplate2	TEF1coreP-nativeTEF1_UTR-yECitrine-ADH1t	6.42*	pTemplate2
sTemplate3	RPL8Ap-nativeRPL8A_UTR-mTFP1-ADH1t	6.24*	pTemplate3
sTemplate4	TEF1p-nativeTEF1_UTR-*RcTal1*-ADH1t	6.84*	pTemplate4
s_yC^{I}-1	RPL8Ap-UTRa1-yECitrine-ADH1t	5.46	p_yC^{I}-1
s_yC^{I}-2	RPL8Ap-UTRa2-yECitrine-ADH1t	3.38	p_yC^{I}-2
s_yC^{I}-3	RPL8Ap-UTRa3-yECitrine-ADH1t	6.64	p_yC^{I}-3
s_yC^{I}-4	RPL8Ap-UTRa4-yECitrine-ADH1t	4.05	p_yC^{I}-4
s_yC^{I}-5	RPL8Ap-UTRa5-yECitrine-ADH1t	4.96	p_yC^{I}-5
s_yC^{I}-6	RPL8Ap-UTRa6-yECitrine-ADH1t	2.67	p_yC^{I}-6
s_yC^{I}-7	RPL8Ap-UTRa7-yECitrine-ADH1t	5.83	p_yC^{I}-7
s_yC^{I}-8	RPL8Ap-UTRa8-yECitrine-ADH1t	3.08	p_yC^{I}-8
s_yC^{I}-9	RPL8Ap-UTRa9-yECitrine-ADH1t	5.82	p_yC^{I}-9
s_yC^{I}-10	RPL8Ap-UTRa10-yECitrine-ADH1t	6.62	p_yC^{I}-10
s_yC^{I}-11	RPL8Ap-UTRa11-yECitrine-ADH1t	5.12	p_yC^{I}-11
s_yC^{I}-12	RPL8Ap-UTRa12-yECitrine-ADH1t	5.72	p_yC^{I}-12
s_yC^{I}-13	RPL8Ap-UTRa13-yECitrine-ADH1t	4.11	p_yC^{I}-13
s_yC^{I}-14	RPL8Ap-UTRa14-yECitrine-ADH1t	4.64	p_yC^{I}-14
s_yC^{I}-15	RPL8Ap-UTRa15-yECitrine-ADH1t	2.69	p_yC^{I}-15
s_yC^{I}-16	RPL8Ap-UTRa16-yECitrine-ADH1t	3.40	p_yC^{I}-16
s_yC^{II}-1	TEF1coreP-UTRa1-yECitrine-ADH1t	5.46	p_yC^{II}-1
s_yC^{II}-2	TEF1coreP-UTRa2-yECitrine-ADH1t	3.38	p_yC^{II}-2
s_yC^{II}-3	TEF1coreP-UTRa3-yECitrine-ADH1t	6.64	p_yC^{II}-3
s_yC^{II}-4	TEF1coreP-UTRa4-yECitrine-ADH1t	4.05	p_yC^{II}-4
s_yC^{II}-5	TEF1coreP-UTRa5-yECitrine-ADH1t	4.96	p_yC^{II}-5
s_yC^{II}-6	TEF1coreP-UTRa6-yECitrine-ADH1t	2.67	p_yC^{II}-6
s_yC^{II}-7	TEF1coreP-UTRa7-yECitrine-ADH1t	5.83	p_yC^{II}-7
s_yC^{II}-8	TEF1coreP-UTRa8-yECitrine-ADH1t	3.08	p_yC^{II}-8
s_yC^{II}-9	TEF1coreP-UTRa9-yECitrine-ADH1t	5.82	p_yC^{II}-9
s_yC^{II}-10	TEF1coreP-UTRa10-yECitrine-ADH1t	6.62	p_yC^{II}-10
s_yC^{II}-11	TEF1coreP-UTRa11-yECitrine-ADH1t	5.12	p_yC^{II}-11
s_yC^{II}-12	TEF1coreP-UTRa12-yECitrine-ADH1t	5.72	p_yC^{II}-12
s_yC^{II}-13	TEF1coreP-UTRa13-yECitrine-ADH1t	4.11	p_yC^{II}-13
s_yC^{II}-14	TEF1coreP-UTRa14-yECitrine-ADH1t	4.64	p_yC^{II}-14
s_yC^{II}-15	TEF1coreP-UTRa15-yECitrine-ADH1t	2.69	p_yC^{II}-15
s_yC^{II}-16	TEF1coreP-UTRa16-yECitrine-ADH1t	3.40	p_yC^{II}-16
s_yC^{III}-1	RPL8Ap-UTRb1-mTFP1-ADH1t	7.25	p_yC^{III}-1
s_yC^{III}-2	RPL8Ap-UTRb2-mTFP1-ADH1t	4.57	p_yC^{III}-2
s_yC^{III}-3	RPL8Ap-UTRb3-mTFP1-ADH1t	5.88	p_yC^{III}-3

s_yC^{III}-4	RPL8Ap-UTRb4-mTFP1-ADH1t	3.21	p_yC^{III}-4
s_yC^{III}-5	RPL8Ap-UTRb5-mTFP1-ADH1t	6.74	p_yC^{III}-5
s_yC^{III}-6	RPL8Ap-UTRb6-mTFP1-ADH1t	4.11	p_yC^{III}-6
s_yC^{III}-7	RPL8Ap-UTRb7-mTFP1-ADH1t	5.38	p_yC^{III}-7
s_yC^{III}-8	RPL8Ap-UTRb8-mTFP1-ADH1t	2.74	p_yC^{III}-8
s_yC^{III}-9	RPL8Ap-UTRb9-mTFP1-ADH1t	7.26	p_yC^{III}-9
s_yC^{III}-10	RPL8Ap-UTRb10-mTFP1-ADH1t	4.57	p_yC^{III}-10
s_yC^{III}-11	RPL8Ap-UTRb11-mTFP1-ADH1t	6.69	p_yC^{III}-11
s_yC^{III}-12	RPL8Ap-UTRb12-mTFP1-ADH1t	3.97	p_yC^{III}-12
s_yC^{III}-13	RPL8Ap-UTRb13-mTFP1-ADH1t	6.14	p_yC^{III}-13
s_yC^{III}-14	RPL8Ap-UTRb14-mTFP1-ADH1t	3.42	p_yC^{III}-14
s_yC^{III}-15	RPL8Ap-UTRb15-mTFP1-ADH1t	5.67	p_yC^{III}-15
s_yC^{III}-16	RPL8Ap-UTRb16-mTFP1-ADH1t	2.75	p_yC^{III}-16
s_yC^{IV}-1	TEF1coreP-UTRc1-yECitrine-ADH1t	5.58	p_yC^{IV}-1
s_yC^{IV}-2	TEF1coreP-UTRc2-yECitrine-ADH1t	6.55	p_yC^{IV}-2
s_yC^{IV}-3	TEF1coreP-UTRc3-yECitrine-ADH1t	3.89	p_yC^{IV}-3
s_yC^{IV}-4	TEF1coreP-UTRc4-yECitrine-ADH1t	4.63	p_yC^{IV}-4
s_yC^{IV}-5	TEF1coreP-UTRc5-yECitrine-ADH1t	4.93	p_yC^{IV}-5
s_yC^{IV}-6	TEF1coreP-UTRc6-yECitrine-ADH1t	5.78	p_yC^{IV}-6
s_yC^{IV}-7	TEF1coreP-UTRc7-yECitrine-ADH1t	2.56	p_yC^{IV}-7
s_yC^{IV}-8	TEF1coreP-UTRc8-yECitrine-ADH1t	3.16	p_yC^{IV}-8
s_yC^{IV}-9	TEF1coreP-UTRc9-yECitrine-ADH1t	5.43	p_yC^{IV}-9
s_yC^{IV}-10	TEF1coreP-UTRc10-yECitrine-ADH1t	6.51	p_yC^{IV}-10
s_yC^{IV}-11	TEF1coreP-UTRc11-yECitrine-ADH1t	3.36	p_yC^{IV}-11
s_yC^{IV}-12	TEF1coreP-UTRc12-yECitrine-ADH1t	4.52	p_yC^{IV}-12
s_yC^{IV}-13	TEF1coreP-UTRc13-yECitrine-ADH1t	4.78	p_yC^{IV}-13
s_yC^{IV}-14	TEF1coreP-UTRc14-yECitrine-ADH1t	5.92	p_yC^{IV}-14
s_yC^{IV}-15	TEF1coreP-UTRc15-yECitrine-ADH1t	2.75	p_yC^{IV}-15
s_yC^{IV}-16	TEF1coreP-UTRc16-yECitrine-ADH1t	4.02	p_yC^{IV}-16
s_yC^{V}-1	RPL8Ap-UTRa1-mTFP1-ADH1t	5.87*	p_yC^{V}-1
s_yC^{V}-2	RPL8Ap-UTRa2-mTFP1-ADH1t	3.58*	p_yC^{V}-2
s_yC^{V}-3	RPL8Ap-UTRa3-mTFP1-ADH1t	6.91*	p_yC^{V}-3
s_yC^{V}-4	RPL8Ap-UTRa4-mTFP1-ADH1t	4.28*	p_yC^{V}-4
s_yC^{V}-5	RPL8Ap-UTRa5-mTFP1-ADH1t	5.13*	p_yC^{V}-5
s_yC^{V}-6	RPL8Ap-UTRa6-mTFP1-ADH1t	2.43*	p_yC^{V}-6
s_yC^{V}-7	RPL8Ap-UTRa7-mTFP1-ADH1t	5.97*	p_yC^{V}-7
s_yC^{V}-8	RPL8Ap-UTRa8-mTFP1-ADH1t	3.08*	p_yC^{V}-8
s_yC^{V}-9	RPL8Ap-UTRa9-mTFP1-ADH1t	7.02*	p_yC^{V}-9
s_yC^{V}-10	RPL8Ap-UTRa10-mTFP1-ADH1t	7.05*	p_yC^{V}-10
s_yC^{V}-11	RPL8Ap-UTRa11-mTFP1-ADH1t	6.64*	p_yC^{V}-11
s_yC^{V}-12	RPL8Ap-UTRa12-mTFP1-ADH1t	6.76*	p_yC^{V}-12
s_yC^{V}-13	RPL8Ap-UTRa13-mTFP1-ADH1t	5.16*	p_yC^{V}-13
s_yC^{V}-14	RPL8Ap-UTRa14-mTFP1-ADH1t	5.18*	p_yC^{V}-14
s_yC^{V}-15	RPL8Ap-UTRa15-mTFP1-ADH1t	4.77*	p_yC^{V}-15
s_yC^{V}-16	RPL8Ap-UTRa16-mTFP1-ADH1t	4.89*	p_yC^{V}-16
s_yC^{VI}-1	TEF1p-UTRt1-*RcTal1*-ADH1t	2.71	p_yC^{VI}-1
s_yC^{VI}-2	TEF1p-UTRt2-*RcTal1*-ADH1t	3.43	p_yC^{VI}-2

s_yC^{VI}-3	TEF1p-UTRt3-*RcTal1*-ADH1t	4.70	p_yC^{VI}-3
s_yC^{VI}-4	TEF1p-UTRt4-*RcTal1*-ADH1t	6.89	p_yC^{VI}-4

Table S.2: Overview of the 5'UTR sequences generated by the PLS regression model and used in this study. UTR_RPL8A and UTR_TEF1 represent the native 5'UTRs of the *RPL8A* and *TEF1* gene respectively. The altered 10 bp parts of the 5'UTRs are presented in bold.

5'UTR name	5'UTR sequence
UTR_RPL8A	AAAACAACTAATTCGAA
UTRa1	AAAACAA**ACGCCTCAAA**
UTRa2	AAAACAA**ATGCCTCAAA**
UTRa3	AAAACAA**ACGCGTCAAA**
UTRa4	AAAACAA**ATGCGTCAAA**
UTRa5	AAAACAA**ACGCCTCACA**
UTRa6	AAAACAA**ATGCCTCACA**
UTRa7	AAAACAA**ACGCGTCACA**
UTRa8	AAAACAA**ATGCGTCACA**
UTRa9	AAAACAA**TCAACGAAAA**
UTRa10	AAAACAA**TCTACGAAAA**
UTRa11	AAAACAA**TCAAGGAAAA**
UTRa12	AAAACAA**TCTAGGAAAA**
UTRa13	AAAACAA**TCAACGATAA**
UTRa14	AAAACAA**TCTACGATAA**
UTRa15	AAAACAA**TCAAGGATAA**
UTRa16	AAAACAA**TCTAGGATAA**
UTRb1	AAAACAA**AGATCTAAAA**
UTRb2	AAAACAA**AGATGTAAAA**
UTRb3	AAAACAA**AGATCTACAA**
UTRb4	AAAACAA**AGATGTACAA**
UTRb5	AAAACAA**AGATCTAAAG**
UTRb6	AAAACAA**AGATGTAAAG**
UTRb7	AAAACAA**AGATCTACAG**
UTRb8	AAAACAA**AGATGTACAG**
UTRb9	AAAACAA**TAAGTGTAAA**
UTRb10	AAAACAA**TGAGTGTAAA**
UTRb11	AAAACAA**TAAGTGTACA**
UTRb12	AAAACAA**TGAGTGTACA**
UTRb13	AAAACAA**TAAGTGTAAC**
UTRb14	AAAACAA**TGAGTGTAAC**
UTRb15	AAAACAA**TAAGTGTACC**
UTRb16	AAAACAA**TGAGTGTACC**
UTR_TEF1	GCATAGCAATCTAATCTAAGTTTTAATTACAAA
UTRc1	GCATAGCAATCTAATCTAAGTTT**ACGGTATAAA**
UTRc2	GCATAGCAATCTAATCTAAGTTT**ACTGTATAAA**
UTRc3	GCATAGCAATCTAATCTAAGTTT**ACGGTATTAA**
UTRc4	GCATAGCAATCTAATCTAAGTTT**ACTGTATTAA**
UTRc5	GCATAGCAATCTAATCTAAGTTT**ACGGTATAAG**
UTRc6	GCATAGCAATCTAATCTAAGTTT**ACTGTATAAG**
UTRc7	GCATAGCAATCTAATCTAAGTTT**ACGGTATTAG**
UTRc8	GCATAGCAATCTAATCTAAGTTT**ACTGTATTAG**

UTRc9	GCATAGCAATCTAATCTAAGTTT**AGATCGTAAA**
UTRc10	GCATAGCAATCTAATCTAAGTTT**ATATCGTAAA**
UTRc11	GCATAGCAATCTAATCTAAGTTT**AGATCGTTAA**
UTRc12	GCATAGCAATCTAATCTAAGTTT**ATATCGTTAA**
UTRc13	GCATAGCAATCTAATCTAAGTTT**AGATCGTAAT**
UTRc14	GCATAGCAATCTAATCTAAGTTT**ATATCGTAAT**
UTRc15	GCATAGCAATCTAATCTAAGTTT**AGATCGTTAT**
UTRc16	GCATAGCAATCTAATCTAAGTTT**ATATCGTTAT**
UTRt1	GCATAGCAATCTAATCTAAGTTT**CGGATTCACCA**
UTRt2	GCATAGCAATCTAATCTAAGTTT**CGGATTCACAA**
UTRt3	GCATAGCAATCTAATCTAAGTTT**CGGATTCAAAA**
UTRt4	GCATAGCAATCTAATCTAAGTTT**AAAAAAAAAAA**

Table S.3: Definitions of all 13 features used in the Partial Least Square (PLS) regression model to predict protein abundances. All 13 features were obtained from the study of Dvir *et al.* [1] and were categorized in four main groups: AUG context, short k-mer sequences, uAUG's and RNA secondary structure (RSS). The adenine of the start codon (AUG) is position +1, all preceding nucleotides are numbered relative to this adenine ending with position -1 for the nucleotide in front of the start codon. A: adenine, T: thymine, G: guanine, C: cytosine.

Feature name	Definition	Category
AG_in_min3	The presence of an A or G at position -3.	AUG context
U_in_min3	The presence of a T at position -3.	
A_in_min1	The presence of an adenine at position -1.	
AA_in_min32	The presence of an AA motif at position [-3, -2].	
CG_in_min32	The presence of a CG motif at position [-3, -2].	
AC_in_min21	The presence of an AC motif at position [-2, -1].	
GACA_kmer	The presence of a GACA motif in the 5'UTR sequence.	Short k-mer sequences
GG_kmer	The presence of a GG motif in the 5'UTR sequence.	
CACC_kmer	The presence of a CACC motif in the 5'UTR sequence.	
CA_in_min76	The presence of a CA motif at position [-7, -6].	
CC_in_min76	The presence of a CC motif at position [-7, -6].	
oof_uAUG	The number of out-of-frame uAUG's in the 5'UTR sequence.	uAUG's
dG_EFE	The ensemble free energy (EFE). The EFE is calculated using RNAfold [2] and sums the Boltzmann weighted free energies of possible secondary structures of a given RNA sequence. To calculate the EFE, the whole 5'UTR and the first 50 nucleotides of the CDS were taken into account.	RSS

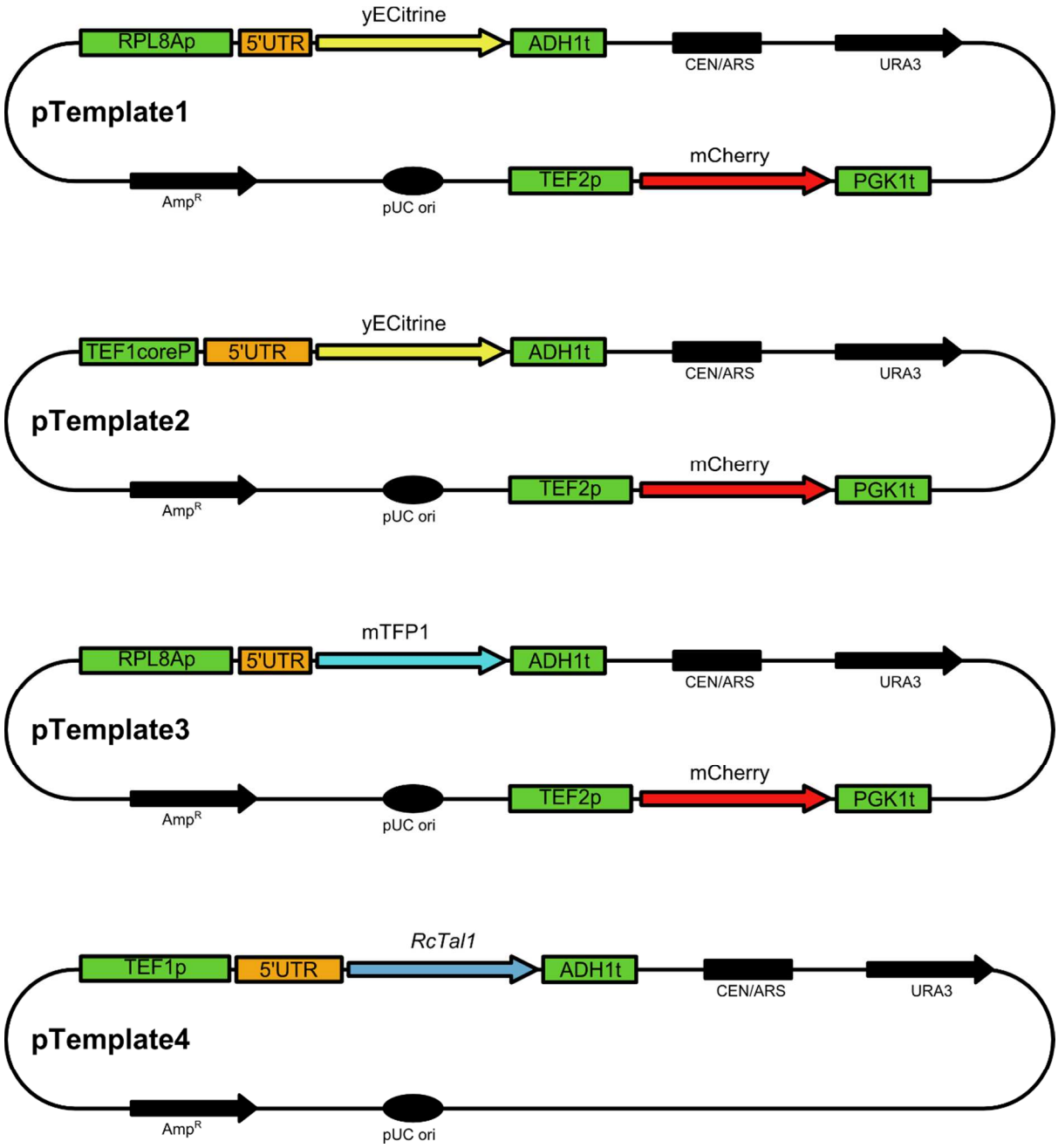


Figure S.1: Schematic overview of the pTemplate plasmids in this study. pTemplate plasmids were used for the amplification of linear DNA for the construction of the different p_yC expression vectors. pTemplate plasmids consist of a yeast low copy backbone containing a *CEN/ARS* ori and *URA3* auxotrophic marker. In addition, an ampicillin resistance gene and pUC ori is present to maintain the plasmids in *E. coli*. pTemplate1 comprises the *RPL8A* promoter with its native 5'UTR in front of the yECitrine reporter and *ADH1* terminator. pTemplate2 contains the *TEF1* core promoter with its native 5'UTR in front of the yECitrine reporter and *ADH1* terminator. pTemplate3 exists of the *RPL8A* promoter with its native 5'UTR in front of the mTFP1 reporter and *ADH1* terminator. pTemplate4 contains the *TEF1* promoter with its native 5'UTR in front of the *RcTal1* coding sequence and *ADH1* terminator. All pTemplate plasmids, except pTemplate4, contain a mCherry

transcription unit controlled by the *TEF2* promoter and *PGK1* terminator to correct for cellular background variation.

Figure S.2: Biplot of the first two components of the PLS regression model. An explanation of all features (AG_in_min3, U_in_min3, A_in_min1, AA_in_min32, CG_in_min32, AC_in_min21, GACA_kmer, GG_kmer, CACC_kmer, CA_in_min76, CC_in_min76, oof_uAUG, and dG_EFE) is given in Supplementary Table S.3.

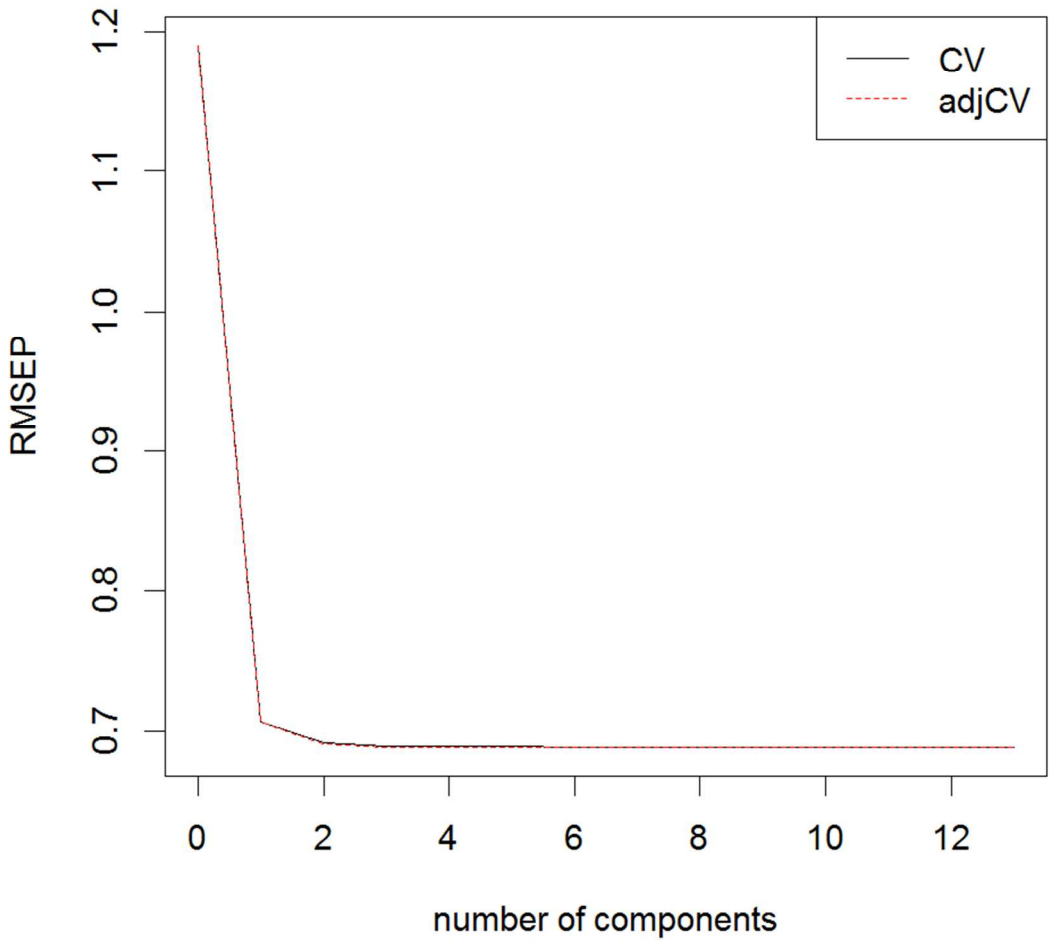


Figure S.3: Cross-validated root mean squared error of prediction (RMSEP) curves. CV is the ordinary cross-validation estimate, adjCV is a bias-corrected cross-validation estimate. From 4 components (*i.e.* latent variables), no further decrease in RMSEP was observed.

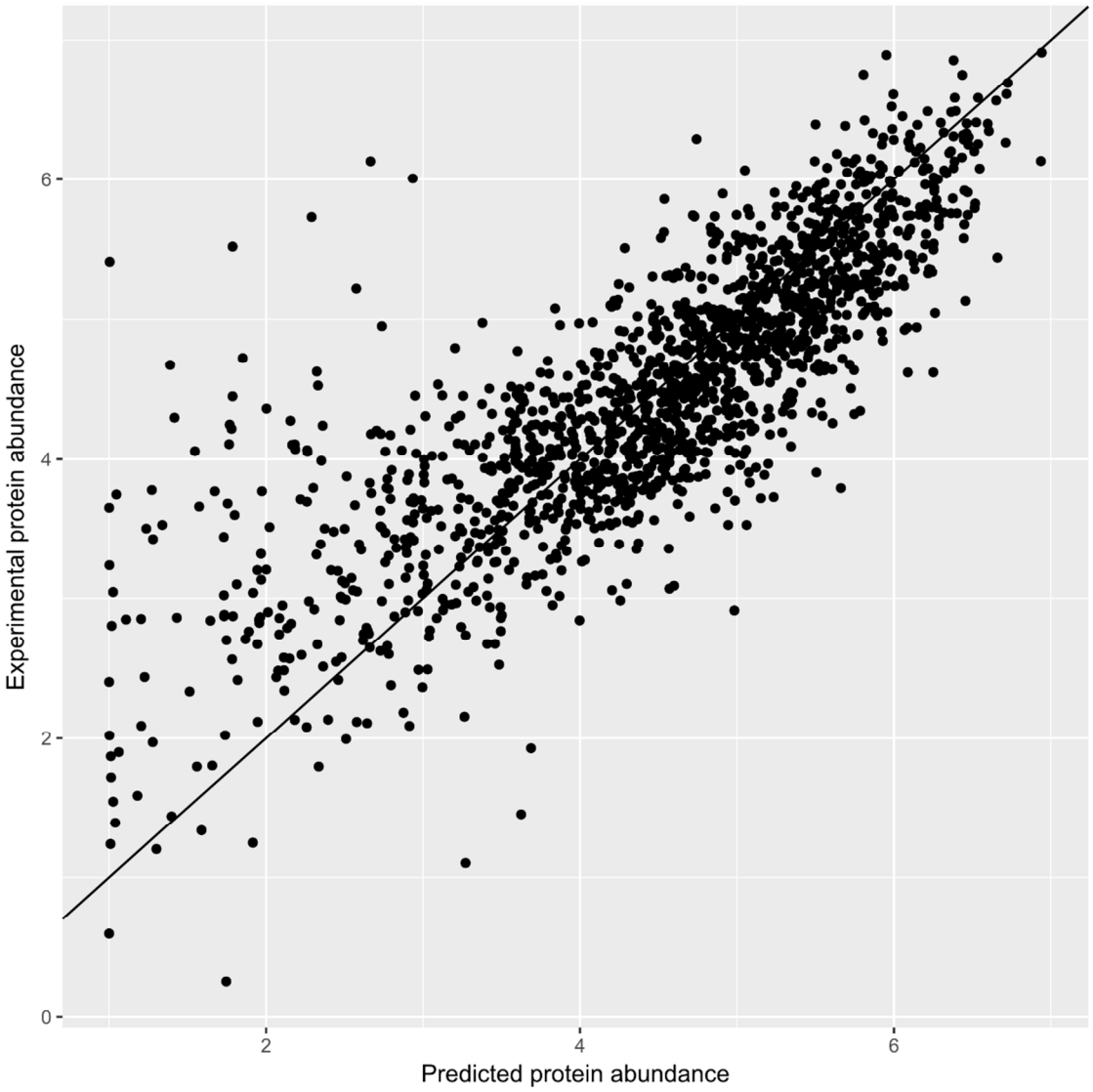


Figure S.4: Validation of the PLS regression model on the training set used for model calibration. The model uses 13 features of the 5'UTR of *Saccharomyces cerevisiae* (Supplementary Table S.3) to predict protein abundance. This plot represents the experimental [1] versus the predicted protein abundance, calculated via the PLS model, for the training set of 1633 5'UTRs. A coefficient of determination (R^2) of 0.67 was obtained.

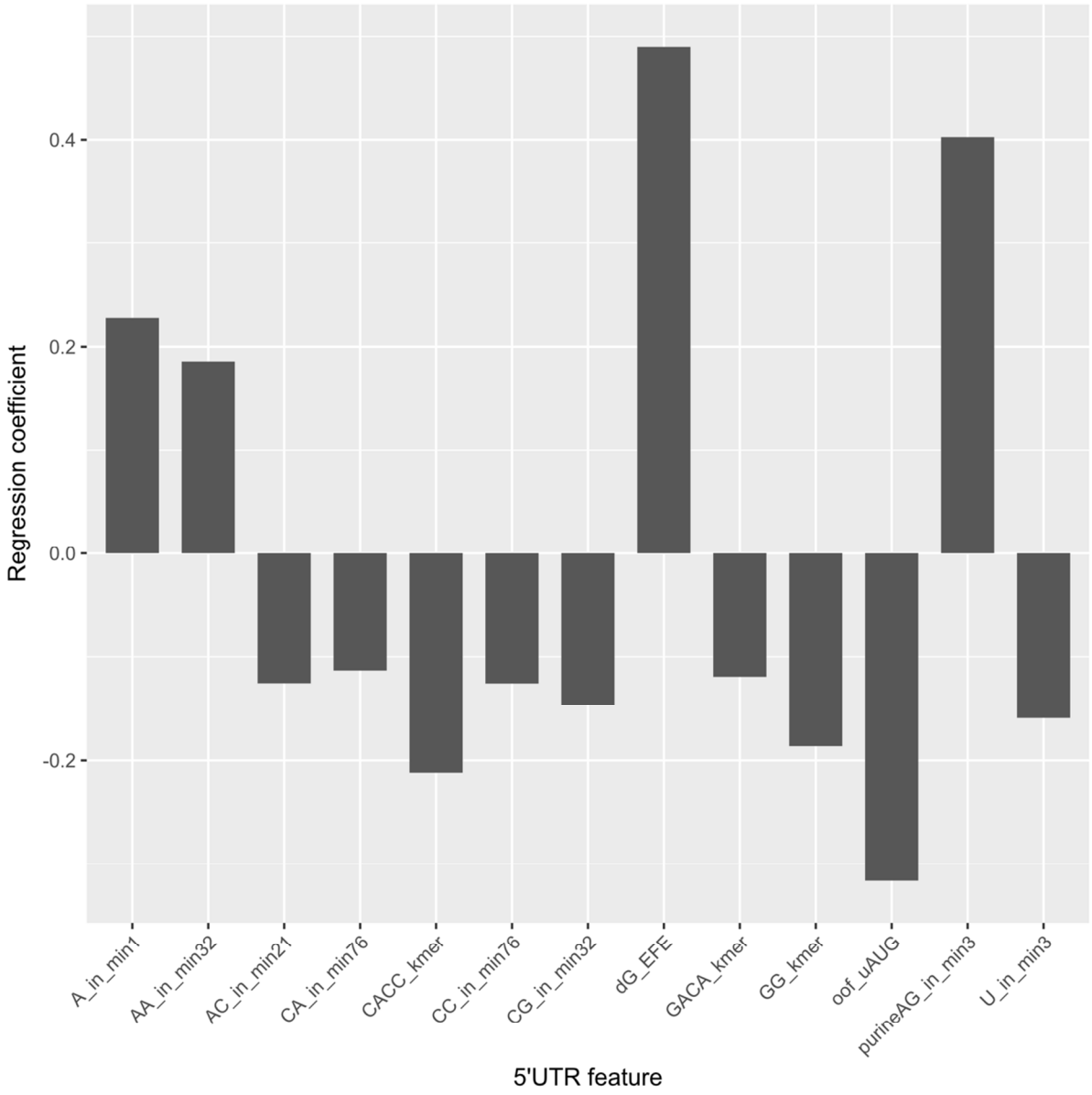


Figure S.5: The estimated regression coefficients of all 5'UTR features. An explanation of all features (AG_in_min3, U_in_min3, A_in_min1, AA_in_min32, CG_in_min32, AC_in_min21, GACA_kmer, GG_kmer, CACC_kmer, CA_in_min76, CC_in_min76, oof_uAUG, and dG_EFE) is available in Supplementary Table S.3.

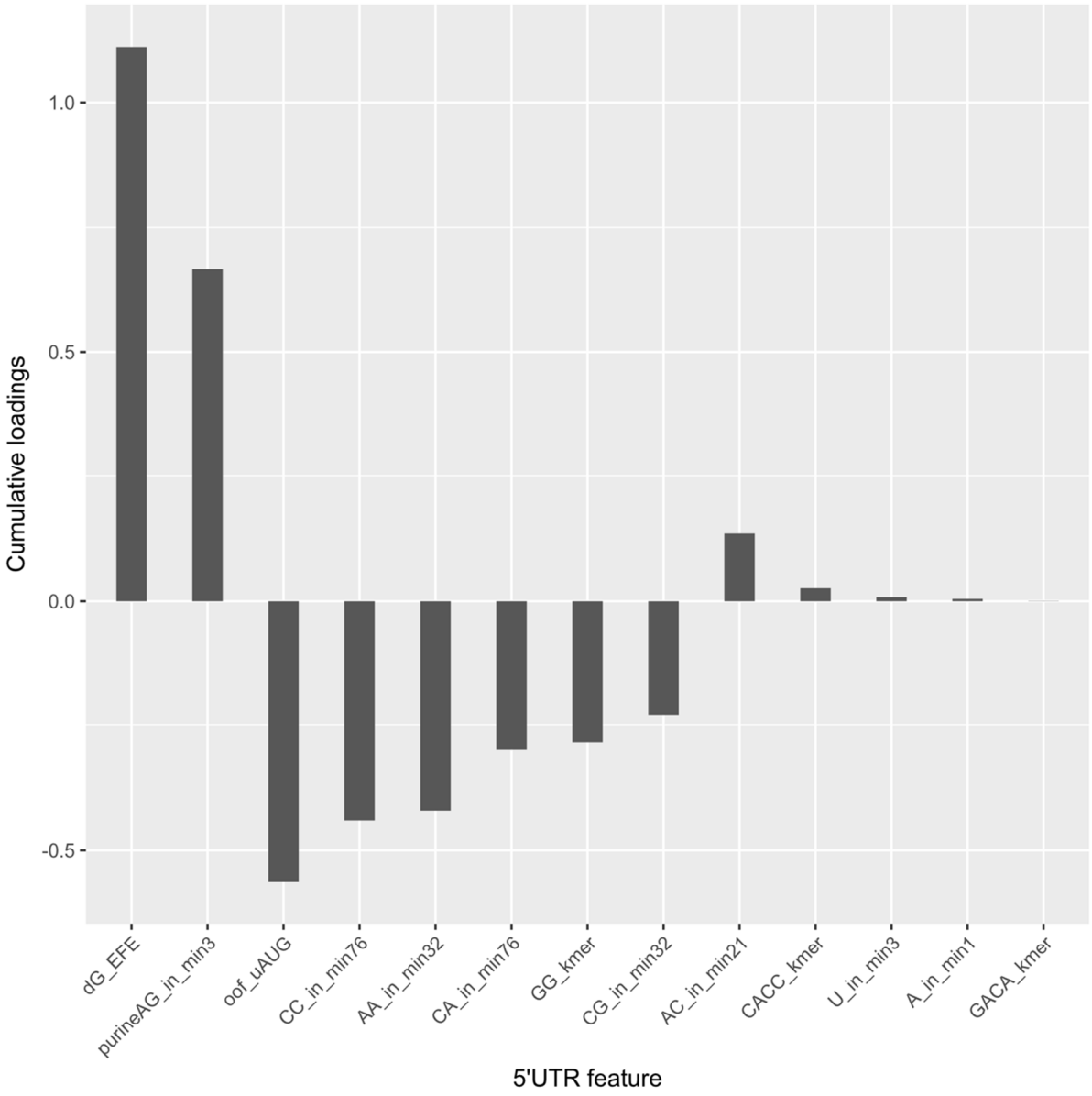
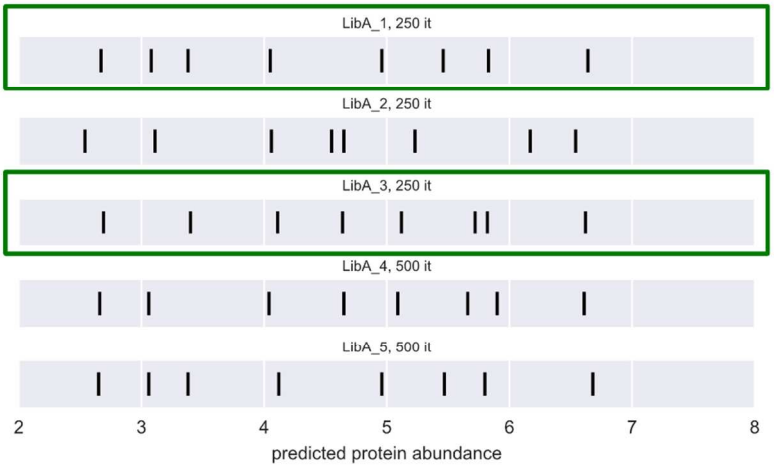
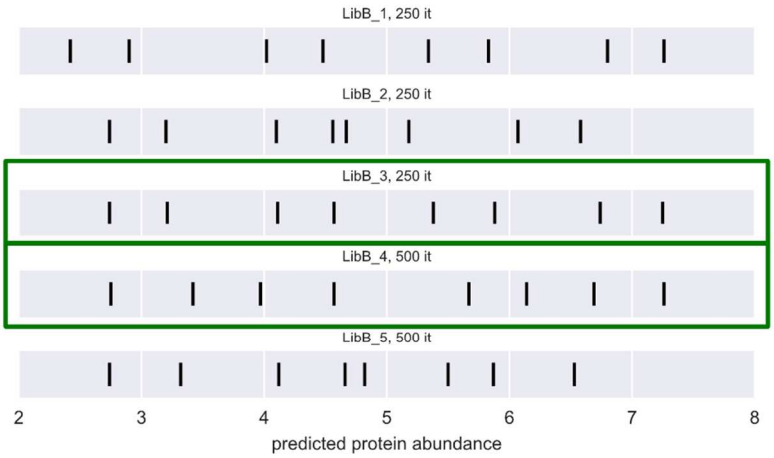


Figure S.6: Cumulative loadings of the four components used in the PLS model. An explanation of all features (AG_in_min3, U_in_min3, A_in_min1, AA_in_min32, CG_in_min32, AC_in_min21, GACA_kmer, GG_kmer, CACC_kmer, CA_in_min76, CC_in_min76, oof_uAUG, and dG_EFE) is available in Supplementary Table S.3.





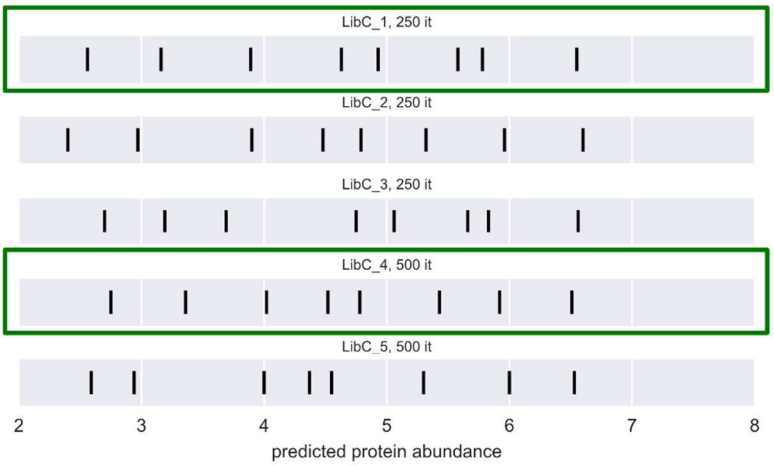


Figure S.7: Event plot representing the distribution of the predicted protein abundances for the calculated 5'UTR libraries (UTRa, UTRb and UTRc). Respectively libA_1 & libA_3, libB_3 & libB_4 and libC_1 & libC_4 were selected (indicated by a green box) and form library UTRa, UTRb and UTRc. All three libraries are available in Supplementary Table S.2.

```
LOCUS       linearDNA-p_yUTR        1278 bp    DNA     linear   SYN 05-APR-2017
DEFINITION  p_yUTRA cut 7094 to 957
ACCESSION   linearDNA-p_yUTR
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown
            Unclassified.
REFERENCE   1  (bases 1 to 1278)
  AUTHORS   Self
  JOURNAL   Unpublished.
COMMENT     SECID/File created by Clone Manager, Scientific & Educational Software
FEATURES             Location/Qualifiers
     misc_feature    1..334
                     /gene="pRPL8A"
                     /product="RPL8A promoter"
                     /SECDrawAs="Region"
                     /SECStyleId=1
     misc_feature    335..351
                     /gene="RPL8A 5'UTR"
                     /product="17 bp 5'UTR of the RPL8A gene"
                     /SECDrawAs="Region"
                     /SECStyleId=1
     misc_signal     342..351
                     /label=UTRa
                     /SECDrawAs="Label"
     CDS             352..1068
                     /gene="'yECitrine"
                     /product="yeast enhanced yellow fluorescent protein"
                     /codon_start=2
                     /translation="CLKVKNYSLVLSQFWLN"
                     /SECDrawAs="Gene"
                     /SECStyleId=1
                     /SECName="yECitrine"
                     /SECDescr="yeast enhanced yellow fluorescent protein"
     misc_feature    1076..1278
                     /gene="ADH1t"
                     /product="ADH1 terminator"
                     /SECDrawAs="Region"
                     /SECStyleId=1
ORIGIN
        1 acataaataa tttctattaa caatgtaatt tccataattt tatattcctc tccaccttct
       61 attgcatcat gtactattca aatgactgta acactagtat tatgaagaaa acacccaaac
      121 atatctaggc catcagattt ttttttttc atttttcatt tttttctcat tttcttattt
      181 atttttattg aaaaataata accgacgcaa acaaattgga aaaaccaacg caaaaaaaaa
      241 aagacgctaa attgtttata aaggcgagga atttgtatct atcaattact attccagttg
      301 tcagtttaca ttgcttaccc tctattatca catcaaaaca annnnnnnnn natgtctaaa
      361 ggtgaagaat tattcactgg tgttgtccca attttggttg aattagatgg tgatgttaat
      421 ggtcacaaat tttctgtctc cggtgaaggt gaaggtgatg ctacttacgg taaattgacc
      481 ttaaaattta tttgtactac tggtaaattg ccagttccat ggccaacctt agtcactact
      541 ttaggttatg gtttgatgtg ttttgctaga tacccagatc atatgaaaca acatgacttt
      601 ttcaagtctg ccatgccaga aggttatgtt caagaaagaa ctatttttt caaagatgac
      661 ggtaactaca agaccagagc tgaagtcaag tttgaaggtg ataccttagt taatagaatc
      721 gaattaaaag gtattgattt taaagaagat ggtaacattt taggtcacaa attggaatac
      781 aactataact ctcacaatgt ttacatcatg gctgacaaac aaaagaatgg tatcaaagtt
      841 aacttcaaaa ttagacacaa cattgaagat ggttctgttc aattagctga ccattatcaa
      901 caaaatactc caattggtga tggtccagtc ttgttaccag acaaccatta cttatcctat
      961 caatctgcct tatccaaaga tccaaacgaa aagagagacc acatggtctt gttagaattt
     1021 gttactgctg ctggtattac ccatggtatg gatgaattgt acaaataagg cgcgccactt
     1081 ctaaataagc gaatttctta tgatttatga tttttattat taaataagtt ataaaaaaaa
     1141 taagtgtata caaattttaa agtgactctt aggttttaaa acgaaaattc ttattcttga
     1201 gtaactcttt cctgtaggtc aggttgcttt ctcaggtata gtatgaggtc gctcttattg
     1261 accacacctc taccggca
//
```

Figure S.8: Annotated Genbank file of the *RPL8Ap-UTRai-yECitrine-ADH1t* transcription unit in expression vector p_yC^{I}-i (i varies from 1 to 16). The 5'UTR is underlined and indicated in bold. The respective sequences of library UTRa are represented in Supplementary Table S.2.

```
LOCUS       linearDNA-p_yUTR        1120 bp    DNA     linear   SYN 05-APR-2017
DEFINITION  p2a_TEF1coreP-RPL8A_UTRlib1_G1-yECit cut 1 to 1120
ACCESSION   linearDNA-p_yUTR
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown
            Unclassified.
REFERENCE   1  (bases 1 to 1120)
  AUTHORS   Self
  JOURNAL   Unpublished.
COMMENT     SECID/File created by Clone Manager, Scientific & Educational Software
FEATURES             Location/Qualifiers
     misc_feature    1..176
                     /gene="cpTEF1"
                     /product="TEF1 core promoter"
                     /SECDrawAs="Region"
                     /SECStyleId=1
     misc_feature    177..193
                     /gene="RPL8A 5'UTR"
                     /product="17 bp 5'UTR of the RPL8A gene"
                     /SECDrawAs="Region"
                     /SECStyleId=1
     misc_signal     184..193
                     /label=UTRa
                     /SECDrawAs="Label"
     CDS             194..910
                     /gene="'yECitrine"
                     /product="yeast enhanced yellow fluorescent protein"
                     /codon_start=3
                     /SECDrawAs="Gene"
                     /SECStyleId=1
                     /SECName="yECitrine"
                     /SECDescr="yeast enhanced yellow fluorescent protein"
     misc_feature    918..1120
                     /gene="ADH1t"
                     /product="ADH1 terminator"
                     /SECDrawAs="Region"
                     /SECStyleId=1
ORIGIN
        1 aataaaaatt tttatcacgt ttctttttct tgaaaatttt tttttttgat ttttttctct
       61 ttcgatgacc tcccattgat atttaagtta ataaacggtc ttcaatttct caagtttcag
      121 tttcattttt cttgttctat tacaactttt tttacttctt gctcattaga aagaaaaaaa
      181 caannnnnnn nnnatgtcta aaggtgaaga attattcact ggtgttgtcc caattttggt
      241 tgaattagat ggtgatgtta atggtcacaa attttctgtc tccggtgaag gtgaaggtga
      301 tgctacttac ggtaaattga ccttaaaatt tatttgtact actggtaaat tgccagttcc
      361 atggccaacc ttagtcacta ctttaggtta tggtttgatg tgttttgcta gatacccaga
      421 tcatatgaaa caacatgact ttttcaagtc tgccatgcca gaaggttatg ttcaagaaag
      481 aactattttt ttcaaagatg acggtaacta caagaccaga gctgaagtca agtttgaagg
      541 tgatacctta gttaatagaa tcgaattaaa aggtattgat tttaaagaag atggtaacat
      601 tttaggtcac aaattggaat acaactataa ctctcacaat gtttacatca tggctgacaa
      661 acaaaagaat ggtatcaaag ttaacttcaa aattagacac aacattgaag atggttctgt
      721 tcaattagct gaccattatc aacaaaatac tccaattggt gatggtccag tcttgttacc
      781 agacaaccat tacttatcct atcaatctgc cttatccaaa gatccaaacg aaaagagaga
      841 ccacatggtc ttgttagaat ttgttactgc tgctggtatt acccatggta tggatgaatt
      901 gtacaaataa ggcgcgccac ttctaaataa gcgaatttct tatgatttat gatttttatt
      961 attaaataag ttataaaaaa aataagtgta tacaaatttt aaagtgactc ttaggtttta
     1021 aaacgaaaat tcttattctt gagtaactct ttcctgtagg tcaggttgct ttctcaggta
     1081 tagtatgagg tcgctcttat tgaccacacc tctaccggca
//
```

Figure S.9: Annotated Genbank file of the *TEF1coreP-UTRai-yECitrine-ADH1t* transcription unit in expression vector p_yC^{II}-i (i varies from 1 to 16). The 5'UTR is underlined and indicated in bold. The respective sequences of library UTRa are represented in Supplementary Table S.2.

```
LOCUS       linearDNA-p_yUTR        1272 bp    DNA     linear   SYN 05-APR-2017
DEFINITION  p2a_RPL8Ap-RPL8A-libB-mTFP1_1 cut 7068 to 931
ACCESSION   linearDNA-p_yUTR
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown
            Unclassified.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Self
  JOURNAL   Unpublished.
COMMENT     SECID/File created by Clone Manager, Scientific & Educational Software
FEATURES             Location/Qualifiers
     misc_feature    1..334
                     /gene="pRPL8A"
                     /product="RPL8A promoter"
                     /SECDrawAs="Region"
                     /SECStyleId=1
     misc_feature    335..351
                     /gene="RPL8A 5'UTR"
                     /product="17 bp 5'UTR of the RPL8A gene"
                     /SECDrawAs="Region"
                     /SECStyleId=1
     misc_signal     342..351
                     /label=UTRb
                     /SECDrawAs="Label"
     CDS             352..1062
                     /gene="mTFP1_co_Sc"
                     /product="yeast codon optimized mTFP1 fluorescent protein"
                     /SECDrawAs="Gene"
                     /SECStyleId=1
                     /SECName="mTFP1_co_Sc"
                     /SECDescr="yeast codon optimized mTFP1 fluorescent protein"
     misc_feature    1070..1272
                     /gene="ADH1t"
                     /product="ADH1 terminator"
                     /SECDrawAs="Region"
                     /SECStyleId=1
ORIGIN
        1 acataaataa tttctattaa caatgtaatt tccataattt tatattcctc tccaccttct
       61 attgcatcat gtactattca aatgactgta acactagtat tatgaagaaa acacccaaac
      121 atatctaggc catcagattt ttttttttc atttttcatt tttttctcat tttcttattt
      181 atttttattg aaaaataata accgacgcaa acaaattgga aaaaccaacg caaaaaaaaa
      241 aagacgctaa attgtttata aaggcgagga atttgtatct atcaattact attccagttg
      301 tcagtttaca ttgcttaccc tctattatca catc aaaaca annnnnnnnn natggtgagt
      361 aaggggggaag aaactactat gggagtaatc aagcccgaca tgaagattaa gttaaagatg
      421 gaagggaacg tgaacggtca cgcattcgtt atcgagggag aaggagaagg caagccctat
      481 gatggcacaa atacgataaa tctggaagtg aaagaaggag cgcctctgcc tttttcctac
      541 gatatactga caacggcgtt tgcctacgga aacagggcgt tcaccaagta ccctgacgat
      601 atcccgaatt acttcaagca atcattccct gaaggatata gttgggagcg tacgatgacg
      661 tttgaggata agggaatagt caaggttaag tcagatatat ctatggaaga agattccttt
      721 atatatgaga tacatttgaa aggtgagaac ttccccccga atggtcctgt aatgcaaaaa
      781 aagaccactg ggtgggacgc gtctaccgag cgtatgtacg tcagagatgg ggtactaaaa
      841 ggagatgtga aacataagtt attattggag ggcggcggcc atcaccgtgt ggacttcaaa
      901 actatttata gagcgaaaaa agccgtgaag ctaccagatt atcattttgt agaccacaga
      961 atcgagattc tgaaccatga taaagactat aataaggtta ctgtgtatga gagcgccgtt
     1021 gcgaggaact ctactgacgg aatggatgaa ttatataaat aaggcgcgcc acttctaaat
     1081 aagcgaattt cttatgattt atgattttta ttattaaata agttataaaa aaaataagtg
     1141 tatacaaatt ttaaagtgac tcttaggttt taaaacgaaa attcttattc ttgagtaact
     1201 ctttcctgta ggtcaggttg ctttctcagg tatagtatga ggtcgctctt attgaccaca
     1261 cctctaccgg ca
//
```

Figure S.10: Annotated Genbank file of the *RPL8Ap-UTRbi-mTFP1-ADH1t* transcription unit in expression vector p_yC^{III}-i (i varies from 1 to 16). The 5'UTR is underlined and indicated in bold. The respective sequences of library UTRb are represented in Supplementary Table S.2.

```
LOCUS       linearDNA-p_yUTR        1136 bp    DNA     linear   SYN 05-APR-2017
DEFINITION  p2a_cTEF1p_UTR-TEF1-libC-yECit_1 cut 7074 to 937
ACCESSION   linearDNA-p_yUTR
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown
            Unclassified.
REFERENCE   1  (bases 1 to 1136)
  AUTHORS   Self
  JOURNAL   Unpublished.
COMMENT     SECID/File created by Clone Manager, Scientific & Educational Software
FEATURES             Location/Qualifiers
     misc_feature    1..176
                     /gene="cpTEF1"
                     /product="TEF1 core promoter"
                     /SECDrawAs="Region"
                     /SECStyleId=1
     misc_feature    177..209
                     /gene="TEF1 5'UTR"
                     /product="33 bp 5'UTR of the TEF1 gene"
                     /SECDrawAs="Region"
                     /SECStyleId=1
     misc_signal     200..209
                     /label=UTRc
                     /SECDrawAs="Label"
     CDS             210..926
                     /gene="'yECitrine"
                     /product="yeast enhanced yellow fluorescent protein"
                     /codon_start=2
                     /SECDrawAs="Gene"
                     /SECStyleId=1
                     /SECName="yECitrine"
                     /SECDescr="yeast enhanced yellow fluorescent protein"
     misc_feature    934..1136
                     /gene="ADH1t"
                     /product="ADH1 terminator"
                     /SECDrawAs="Region"
                     /SECStyleId=1
ORIGIN
        1 aataaaaatt tttatcacgt ttctttttct tgaaaatttt tttttttgat ttttttctct
       61 ttcgatgacc tcccattgat atttaagtta ataaacggtc ttcaatttct caagtttcag
      121 tttcattttt cttgttctat tacaactttt tttacttctt gctcattaga aagaaagcat
      181 agcaatctaa tctaagtttn nnnnnnnnna tgtctaaagg tgaagaatta ttcactggtg
      241 ttgtcccaat tttggttgaa ttagatggtg atgttaatgg tcacaaattt tctgtctccg
      301 gtgaaggtga aggtgatgct acttacggta aattgacctt aaaatttatt tgtactactg
      361 gtaaattgcc agttccatgg ccaaccttag tcactacttt aggttatggt ttgatgtgtt
      421 ttgctagata cccagatcat atgaaacaac atgacttttt caagtctgcc atgccagaag
      481 gttatgttca agaaagaact atttttttca aagatgacgg taactacaag accagagctg
      541 aagtcaagtt tgaaggtgat accttagtta atagaatcga attaaaaggt attgatttta
      601 aagaagatgg taacatttta ggtcacaaat tggaatacaa ctataactct cacaatgttt
      661 acatcatggc tgacaaacaa aagaatggta tcaaagttaa cttcaaaatt agacacaaca
      721 ttgaagatgg ttctgttcaa ttagctgacc attatcaaca aaatactcca attggtgatg
      781 gtccagtctt gttaccagac aaccattact tatcctatca atctgcctta tccaaagatc
      841 caaacgaaaa gagagaccac atggtcttgt tagaatttgt tactgctgct ggtattaccc
      901 atggtatgga tgaattgtac aaataaggcg cgccacttct aaataagcga atttcttatg
      961 atttatgatt tttattatta aataagttat aaaaaaaata agtgtataca aattttaaag
     1021 tgactcttag gttttaaaac gaaaattctt attcttgagt aactctttcc tgtaggtcag
     1081 gttgctttct caggtatagt atgaggtcgc tcttattgac cacacctcta ccggca
//
```

Figure S.11: Annotated Genbank file of the *TEF1coreP-UTRci-yECitrine-ADH1t* transcription unit in expression vector p_yC^{IV}-i (i varies from 1 to 16). The 5'UTR is underlined and indicated in bold. The respective sequences of library UTRc are represented in Supplementary Table S.2.

```
LOCUS       linearDNA-p_yUTR        1272 bp    DNA     linear   SYN 05-APR-2017
DEFINITION  p2a_RPL8Ap-RPL8AUTR-libA_G1 cut 4125 to 5397
ACCESSION   linearDNA-p_yUTR
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown
            Unclassified.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Self
  JOURNAL   Unpublished.
COMMENT     SECID/File created by Clone Manager, Scientific & Educational Software
FEATURES             Location/Qualifiers
     misc_feature    1..341
                     /gene="pRPL8A"
                     /product="RPL8A promoter"
                     /SECDrawAs="Region"
                     /SECStyleId=1
     misc_feature    335..351
                     /gene="RPL8A 5'UTR"
                     /product="17 bp 5'UTR of the RPL8A gene"
                     /SECDrawAs="Region"
                     /SECStyleId=1
     misc_signal     342..351
                     /label=UTRa
                     /SECDrawAs="Label"
     CDS             352..1062
                     /gene="mTFP1_co_Sc"
                     /product="yeast codon optimized mTFP1 fluorescent protein"
                     /SECDrawAs="Gene"
                     /SECStyleId=1
                     /SECName="mTFP1_co_Sc"
                     /SECDescr="yeast codon optimized mTFP1 fluorescent protein"
     misc_feature    1070..1272
                     /gene="ADH1t"
                     /product="ADH1 terminator"
                     /SECDrawAs="Region"
                     /SECStyleId=1
ORIGIN
        1 acataaataa tttctattaa caatgtaatt tccataattt tatattcctc tccaccttct
       61 attgcatcat gtactattca aatgactgta acactagtat tatgaagaaa acacccaaac
      121 atatctaggc catcagattt ttttttttc atttttcatt tttttctcat tttcttattt
      181 atttttattg aaaaataata accgacgcaa acaaattgga aaaaccaacg caaaaaaaaa
      241 aagacgctaa attgtttata aaggcgagga atttgtatct atcaattact attccagttg
      301 tcagtttaca ttgcttaccc tctattatca catcaaaaca annnnnnnnn natggtgagt
      361 aaggggggaag aaactactat gggagtaatc aagcccgaca tgaagattaa gttaaagatg
      421 gaagggaacg tgaacggtca cgcattcgtt atcgagggag aaggagaagg caagccctat
      481 gatggcacaa atacgataaa tctggaagtg aaagaaggag cgcctctgcc tttttcctac
      541 gatatactga caacggcgtt tgcctacgga aacagggcgt tcaccaagta ccctgacgat
      601 atcccgaatt acttcaagca atcattccct gaaggatata gttgggagcg tacgatgacg
      661 tttgaggata agggaatagt caaggttaag tcagatatat ctatggaaga agattccttt
      721 atatatgaga tacatttgaa aggtgagaac ttccccccga atggtcctgt aatgcaaaaa
      781 aagaccactg ggtgggacgc gtctaccgag cgtatgtacg tcagagatgg ggtactaaaa
      841 ggagatgtga aacataagtt attattggag ggcggcggcc atcaccgtgt ggacttcaaa
      901 actatttata gagcgaaaaa agccgtgaag ctaccagatt atcattttgt agaccacaga
      961 atcgagattc tgaaccatga taaagactat aataaggtta ctgtgtatga gagcgccgtt
     1021 gcgaggaact ctactgacgg aatggatgaa ttatataaat aaggcgcgcc acttctaaat
     1081 aagcgaattt cttatgattt atgattttta ttattaaata agttataaaa aaaataagtg
     1141 tatacaaatt ttaaagtgac tcttaggttt taaaacgaaa attcttattc ttgagtaact
     1201 ctttcctgta ggtcaggttg ctttctcagg tatagtatga ggtcgctctt attgaccaca
     1261 cctctaccgg ca
//
```

Figure S.12: Annotated Genbank file of the *RPL8Ap-UTRai-mTFP1-ADH1t* transcription unit in expression vector p_yC^{V}-i (i varies from 1 to 16). The 5'UTR is underlined and indicated in bold. The respective sequences of library UTRa are represented in Supplementary Table S.2.

```
LOCUS       TEF1p-RcTal1-ADH        2219 bp    DNA     linear   SYN 13-DEC-2017
DEFINITION  p2a33_5UTRnative_TAL1 cut 2590 to 4808
ACCESSION   TEF1p-RcTal1-ADH
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown
            Unclassified.
REFERENCE   1  (bases 1 to 2219)
  AUTHORS   Self
  JOURNAL   Unpublished.
COMMENT     SECID/File created by Clone Manager, Scientific & Educational Software
FEATURES             Location/Qualifiers
     misc_feature    1..203
                     /gene="UAS_TEF1"
                     /product="upstream activating sequence TEF1 gene"
                     /SECDrawAs="Region"
                     /SECStyleId=1
     misc_feature    204..413
                     /gene="TEF1 core promoter"
                     /SECDrawAs="Region"
                     /SECStyleId=1
     misc_feature    380..413
                     /gene="TEF1 5'UTR"
                     /SECDrawAs="Region"
                     /SECStyleId=1
     misc_signal     403..413
                     /label=RcTal1-UTRlib
                     /SECDrawAs="Label"
     CDS             414..2009
                     /gene="co_Rc_TAL1'"
                     /SECDrawAs="Gene"
                     /SECStyleId=1
                     /SECName="Rc_coSc_TAL1"
     misc_feature    2017..2219
                     /gene="ADH1t"
                     /product="ADH terminator"
                     /SECDrawAs="Region"
                     /SECStyleId=1
ORIGIN
        1 atagcttcaa aatgtttcta ctccttttt actcttccag attttctcgg actccgcgca
       61 tcgccgtacc acttcaaaac acccaagcac agcatactaa atttcccctc tttcttcctc
      121 tagggtgtcg ttaattaccc gtactaaagg tttggaaaag aaaaaagaga ccgcctcgtt
      181 tctttttctt cgtcgaaaaa ggcaataaaa atttttatca cgtttctttt tcttgaaaat
      241 ttttttttt gatttttttc tctttcgatg acctcccatt gatatttaag ttaataaacg
      301 gtcttcaatt tctcaagttt cagtttcatt tttcttgttc tattacaact ttttttactt
      361 cttgctcatt agaaagaaag g catagcaatc taatctaagt ttnnnnnnnn nnnatgacct
      421 tacaatccca aactgccaaa gactgcttag ccttagacgg tgccttgacc ttggttcaat
      481 gtgaagcaat tgccacacat agatccagaa taagtgtcac cccagctttg agagaaagat
      541 gcgctagagc acatgccaga ttagaacacg ctattgcaga acaaagacac atctatggta
      601 taactacagg ttttggtcct ttggctaata gattaatagg tgccgatcaa ggtgctgaat
      661 tgcaacaaaa cttaatctac catttggcta ctggtgttgg tccaaaattg tcttgggccg
      721 aagctagagc attgatgttg gcaagattga actcaatctt gcaaggtgca tctggtgcct
      781 cacctgaaac aatcgacaga attgttgctg tcttaaacgc tggtttcgca ccagaagtcc
      841 ctgcccaagg tactgtaggt gcttccggtg acttgacacc attggcacat atggttttgg
      901 ccttacaagg tagaggtaga atgattgatc ctagtggtag agttcaagaa gccggtgctg
      961 tcatggacag attatgtggt ggtccattga ctttagctgc aagagatggt ttggctttag
     1021 ttaatggtac ttctgccatg acagctatcg ccgctttgac aggtgttgaa gcagccagag
     1081 ctattgatgc tgcattaaga cattccgcag tattaatgga agttttgagt ggtcatgcag
     1141 aagcctggca cccagctttt gcagaattaa gaccacaccc tggtcaatta agagctaccg
     1201 aaagattagc ccaagctttg gatggtgcag gtagagtttg cagaaccttg actgccgcta
     1261 gaagattgac agcagccgac ttaagaccag aagatcatcc tgcacaagac gcctattctt
     1321 tgagagttgt cccacaatta gttggtgctg tctgggatac tttggactgg cacgatagag
     1381 tagttacctg tgaattgaac tcagtcactg ataacccaat atttcctgaa ggttgcgctg
     1441 tacctgcatt acatggtggt aatttcatgg gtgtacacgt tgcattggcc tccgacgctt
     1501 taaacgctgc attagtaaca ttggctggtt tagttgaaag acaaatcgca agattgaccg
     1561 atgaaaagtt gaataagggt ttgccagcat ttttgcatgg tggtcaagca ggtttacaat
     1621 caggtttcat gggtgctcaa gttacagcta ccgcattgtt agcagaaatg agagccaacg
     1681 ctacccctgt ctctgtacaa tctttgtcaa ctaatggtgc taaccaagat gtcgtatcaa
     1741 tgggtactat cgccgctaga agagcaagag cccaattgtt gccattgtct caaatccaag
     1801 caatcttggc tttagcattg gcccaagcta tggacttgtt agatgaccct gaaggtcaag
     1861 caggttggtc cttgacagcc agagacttaa gagatagaat tagagctgtt agtccaggtt
```

```
     1921 tgagagctga tagaccttta gcaggtcata tagaagcagt cgcacaaggt ttgagacatc
     1981 catccgccgc agcagaccct ccagcctaag gcgcgccact tctaaataag cgaatttctt
     2041 atgatttatg atttttatta ttaaataagt tataaaaaaa ataagtgtat acaaatttta
     2101 aagtgactct taggttttaa aacgaaaatt cttattcttg agtaactctt tcctgtaggt
     2161 caggttgctt tctcaggtat agtatgaggt cgctcttatt gaccacacct ctaccggca
//
```

Figure S.13: Annotated Genbank file of the *TEF1p-UTRti-RcTal1-ADH1t* transcription unit in expression vector p_yC^{VI}-i (i varies from 1 to 4). The 5'UTR is underlined and indicated in bold. The respective sequences of library UTRt are represented in Supplementary Table S.2.

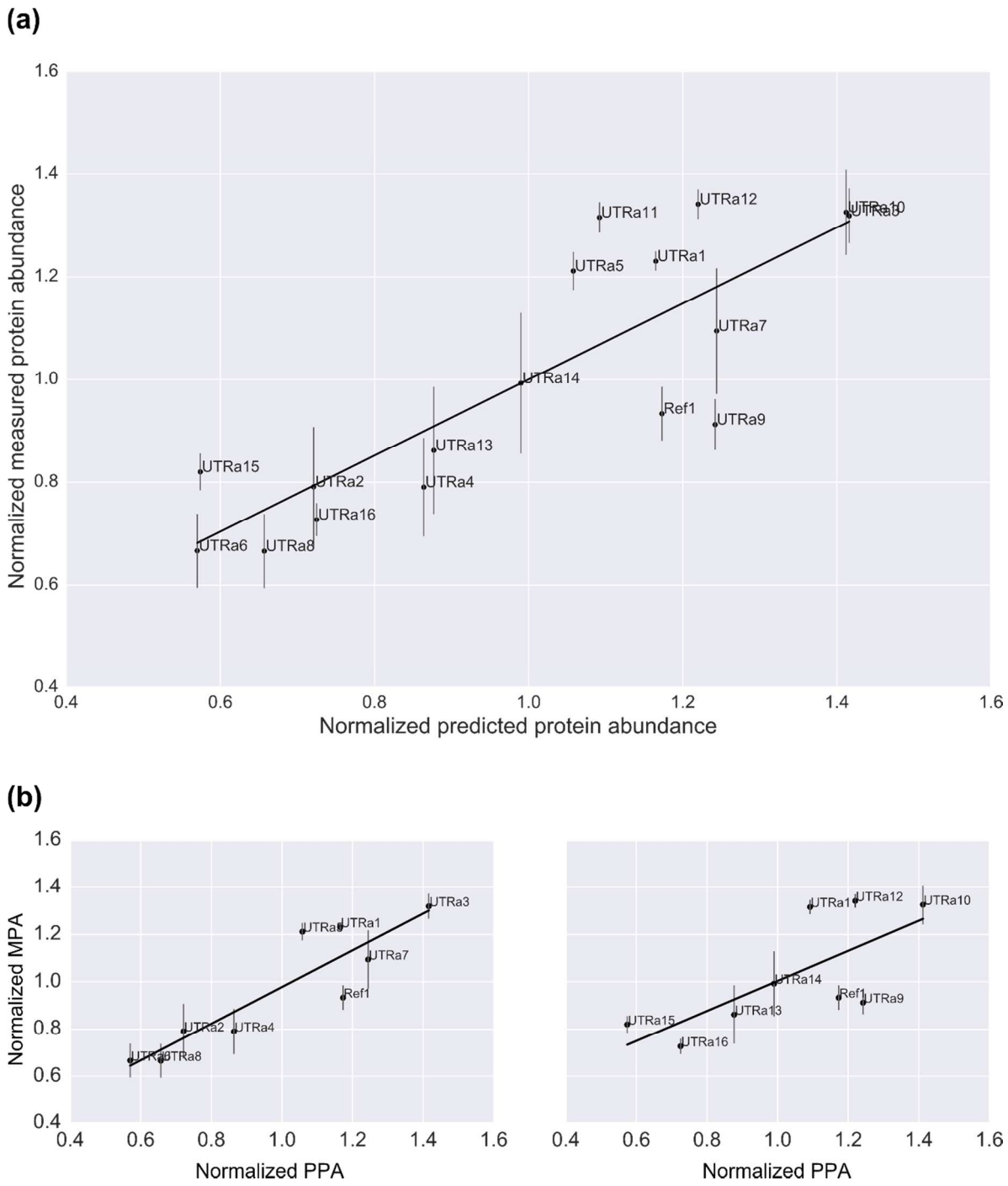


Figure S.14: OLS regression plots comparing the normalized predicted protein abundance (PPA) calculated by forward engineering with our model and the normalized measured protein abundance (MPA), determined by measuring yECitrine-to-mCherry ratios. (a) Regression plot of both calculated 8-containing 5'UTR libraries representing strains s_yC^{I}-1 to s_yC^{I}-16, additionally, reference strain sTemplate1 was included (R^2 = 0.70). (b) Left: Regression plot of the first part of library UTRa consisting of eight 5'UTR candidates representing strains s_yC^{I}-1 to s_yC^{I}-8 including reference strain sTemplate1 (R^2 = 0.81). Right: Regression plot of the second part of library UTRa consisting of eight 5'UTR candidates representing strains s_yC^{I}-9 to s_yC^{I}-16 including reference strain sTemplate1 (R^2 = 0.51). Error bars represent standard deviations of four biological replicates.

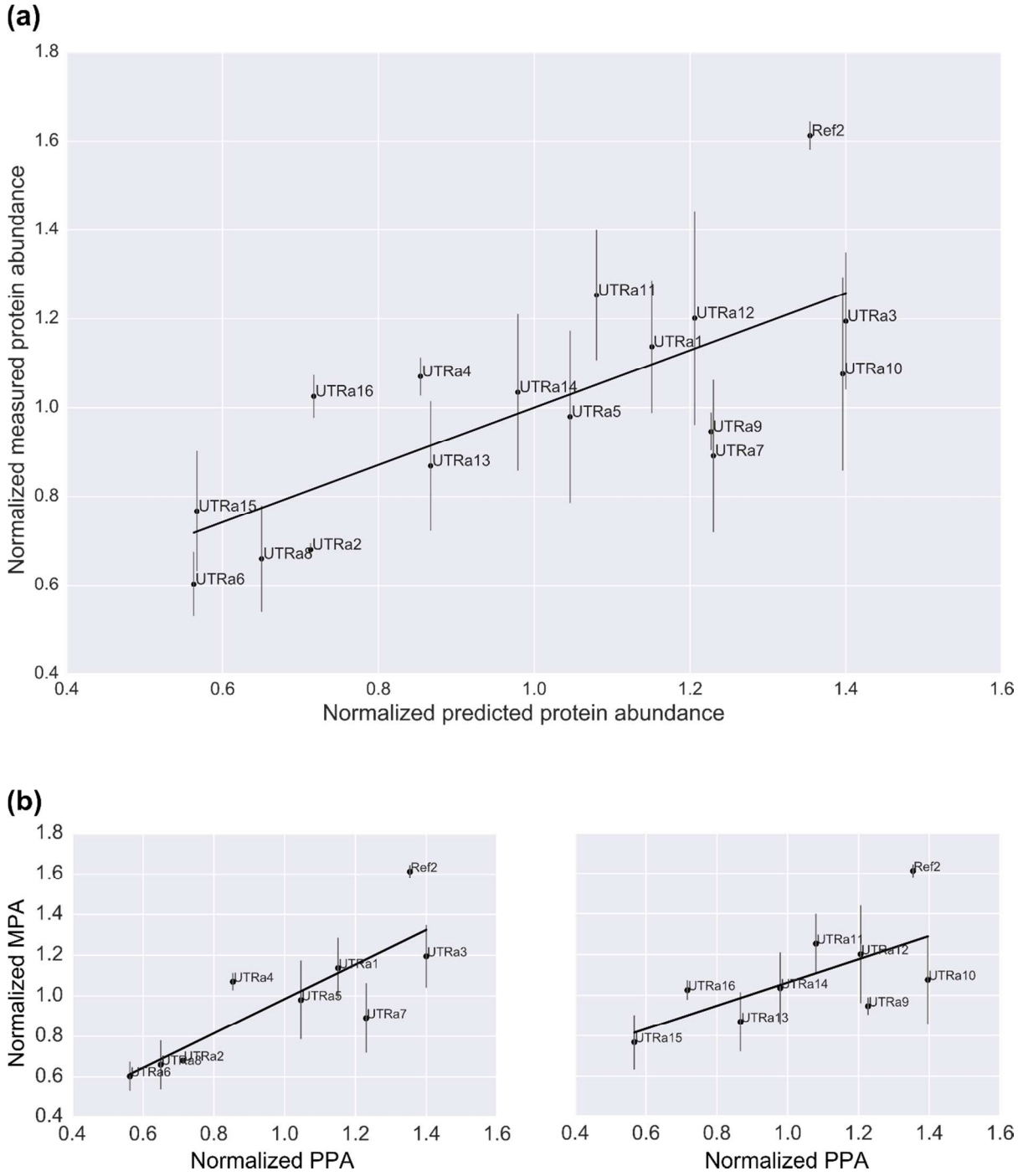


Figure S.15: OLS regression plots comparing the normalized predicted protein abundance (PPA) calculated by forward engineering with our model and the normalized measured protein abundance (MPA), determined by measuring yECitrine-to-mCherry ratios. (a) Regression plot of both calculated 8-containing 5'UTR libraries representing strains s_yC^{II}-1 to s_yC^{II}-16, additionally, reference strain sTemplate2 was included (R^2 = 0.54). (b) Left: Regression plot of the first part of library UTRa consisting of eight 5'UTR candidates representing strains s_yC^{II}-1 to s_yC^{II}-8 including reference strain sTemplate2 (R^2 = 0.69). Right: Regression plot of the second part of library UTRa consisting of eight 5'UTR candidates representing strains s_yC^{II}-9 to s_yC^{II}-16 including reference strain sTemplate2 (R^2 = 0.43). Error bars represent standard deviations of four biological replicates.

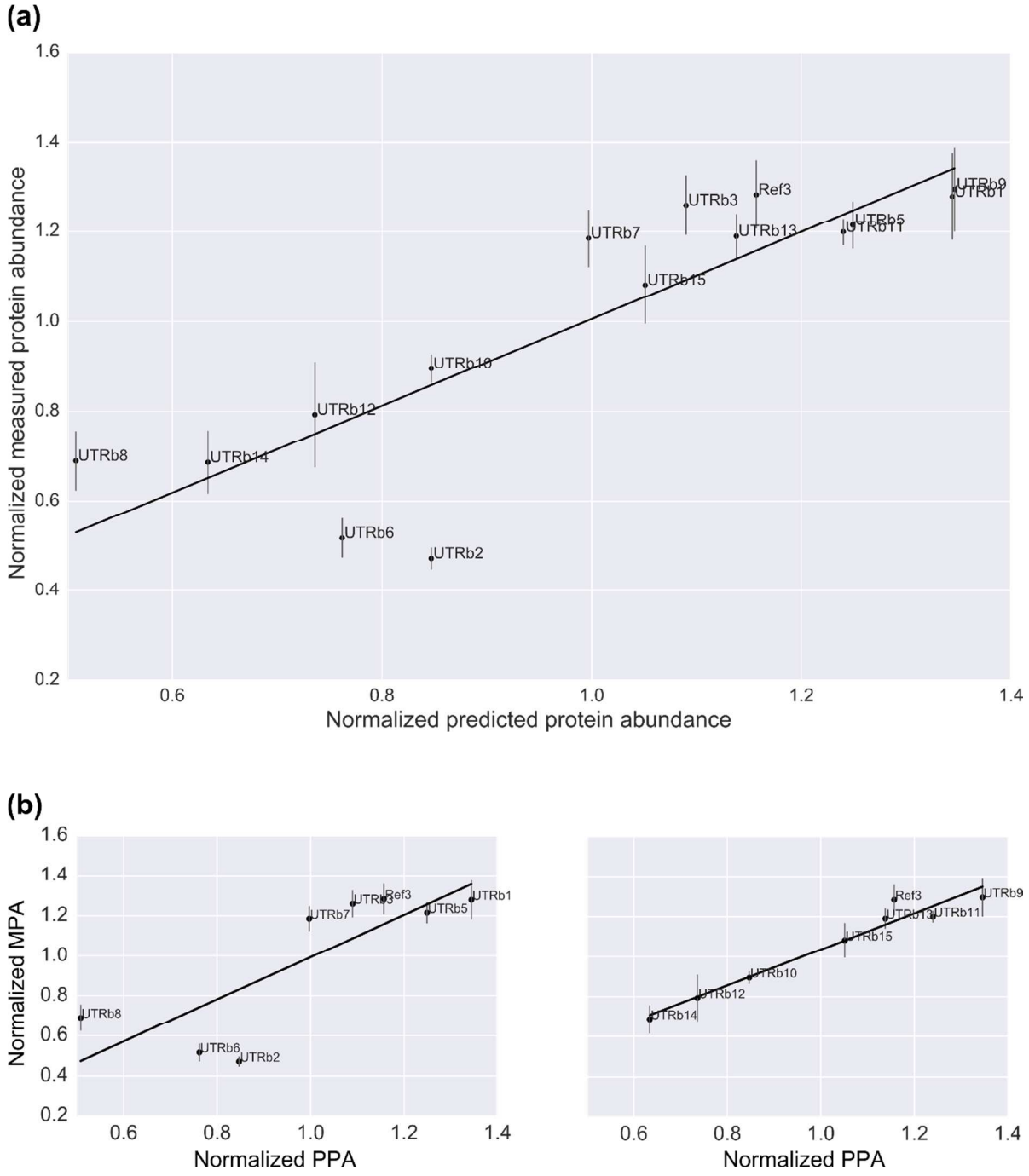


Figure S.16: OLS regression plots comparing the normalized predicted protein abundance (PPA) calculated by forward engineering with our model and the normalized measured protein abundance (MPA), determined by measuring mTFP1-to-mCherry ratios. (a) Regression plot of both calculated 8-containing 5'UTR libraries representing strains s_yC^{III}-1 to s_yC^{III}-15, additionally, reference strain sTemplate3 was included (R^2 = 0.73). (b) Left: Regression plot of the first part of library UTRb consisting of eight 5'UTR candidates representing strains s_yC^{III}-1 to s_yC^{III}-8 including reference strain sTemplate3 (R^2 = 0.65). Right: Regression plot of the second part of library UTRb consisting of eight 5'UTR candidates representing strains s_yC^{III}-9 to s_yC^{III}-15 including reference strain sTemplate3 (R^2 = 0.95). Error bars represent standard deviations of four biological replicates. Due to cloning issues, strains s_yC^{III}-4 and s_yC^{III}-16 are not included.

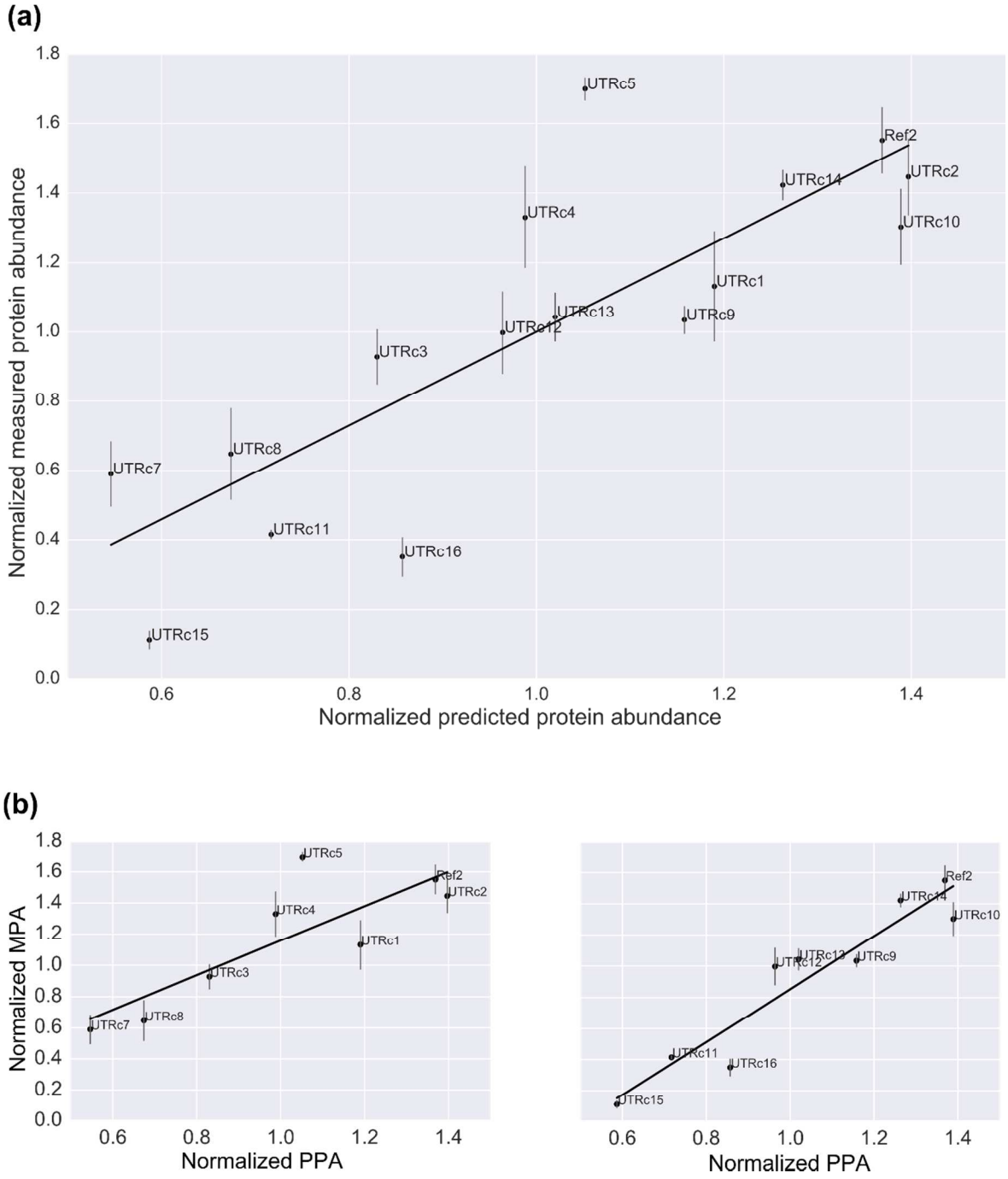


Figure S.17: OLS regression plots comparing the normalized predicted protein abundance (PPA) calculated by forward engineering with our model and the normalized measured protein abundance (MPA), determined by measuring yECitrine-to-mCherry ratios. (a) Regression plot of both calculated 8-containing 5'UTR libraries representing strains s_yC^{IV}-1 to s_yC^{IV}-16, additionally, reference strain sTemplate2 was included (R^2 = 0.67). (b) Left: Regression plot of the first part of library UTRc consisting of eight 5'UTR candidates representing strains s_yC^{IV}-1 to s_yC^{IV}-8 including reference strain sTemplate2 (R^2 = 0.69). Right: Regression plot of the second part of library UTRc consisting of eight 5'UTR candidates representing strains s_yC^{IV}-9 to s_yC^{IV}-16 including reference strain sTemplate2 (R^2 = 0.90). Error bars represent standard deviations of four biological replicates. Due to cloning issues, strain s_yC^{IV}-6 is not included.

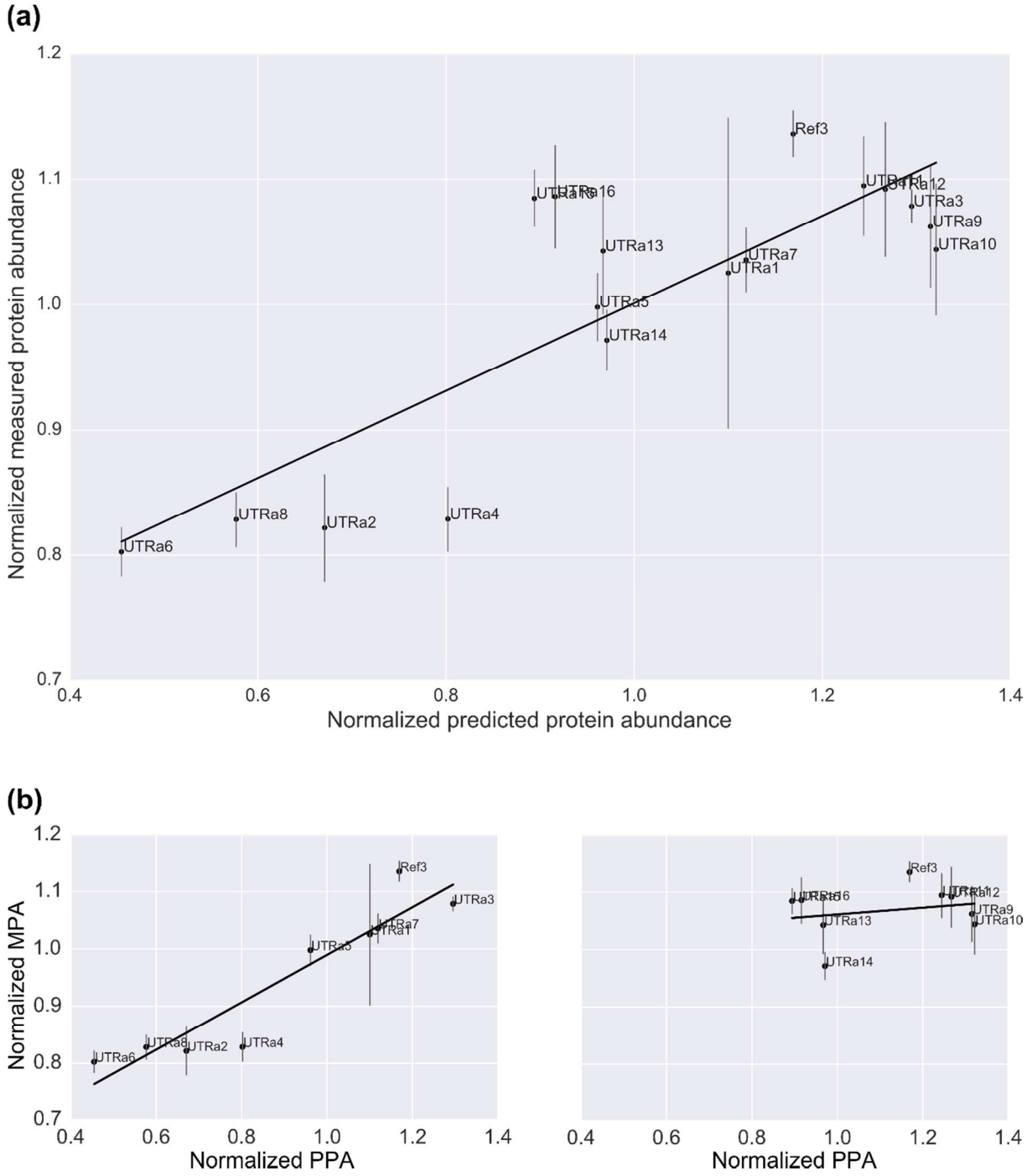


Figure S.18: OLS regression plots comparing the normalized predicted protein abundance (PPA) recalculated by reverse engineering with our model and the normalized measured protein abundance (MPA), determined by measuring mTFP1-to-mCherry ratios. (a) Regression plot of both calculated 8-containing 5'UTR libraries representing strains s_yC^{v}-1 to s_yC^{v}-16, additionally, reference strain sTemplate3 was included (R^2 = 0.69). (b) Left: Regression plot of the first part of library UTRa consisting of eight 5'UTR candidates representing strains s_yC^{v}-1 to s_yC^{v}-8 including reference strain sTemplate3 (R^2 = 0.88). Right: Regression plot of the second part of library UTRa consisting of eight 5'UTR candidates representing strains s_yC^{v}-9 to s_yC^{v}-16 including reference strain sTemplate3 (R^2 = 0.05). Error bars represent standard deviations of four biological replicates.

References

(1) Dvir, S., Velten, L., Sharon, E., Zeevi, D., Carey, L. B., Weinberger, A., and Segal, E. (2013) Deciphering the rules by which 5′-UTR sequences affect protein expression in yeast. *PNAS 110*, e2792-801.

(2) Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., and Schuster, P. (1994) Fast Folding and Comparison of RNA Secondary Structure. *Monatshefte für Chemie 125*, 167–188.