Cis-regulatory elements explain most of the mRNA stability variation across genes in yeast
Supplemental Figures

Jun Cheng[1,2], Kerstin Maier[3], Žiga Avsec[1,2], Petra Rus[3], and Julien Gagneur[1,2,*]

[1]Department of Informatics, Technical University of Munich, Boltzmannstraße 3, 85748 Garching, Germany
[2]Graduate School of Quantitative Biosciences (QBM), Ludwig-Maximilians-Universität München, Germany

[3]Department of Molecular Biology, Max Planck Institute for Biophysical Chemistry, Am Faßberg 11, 37077 Göttingen, Germany
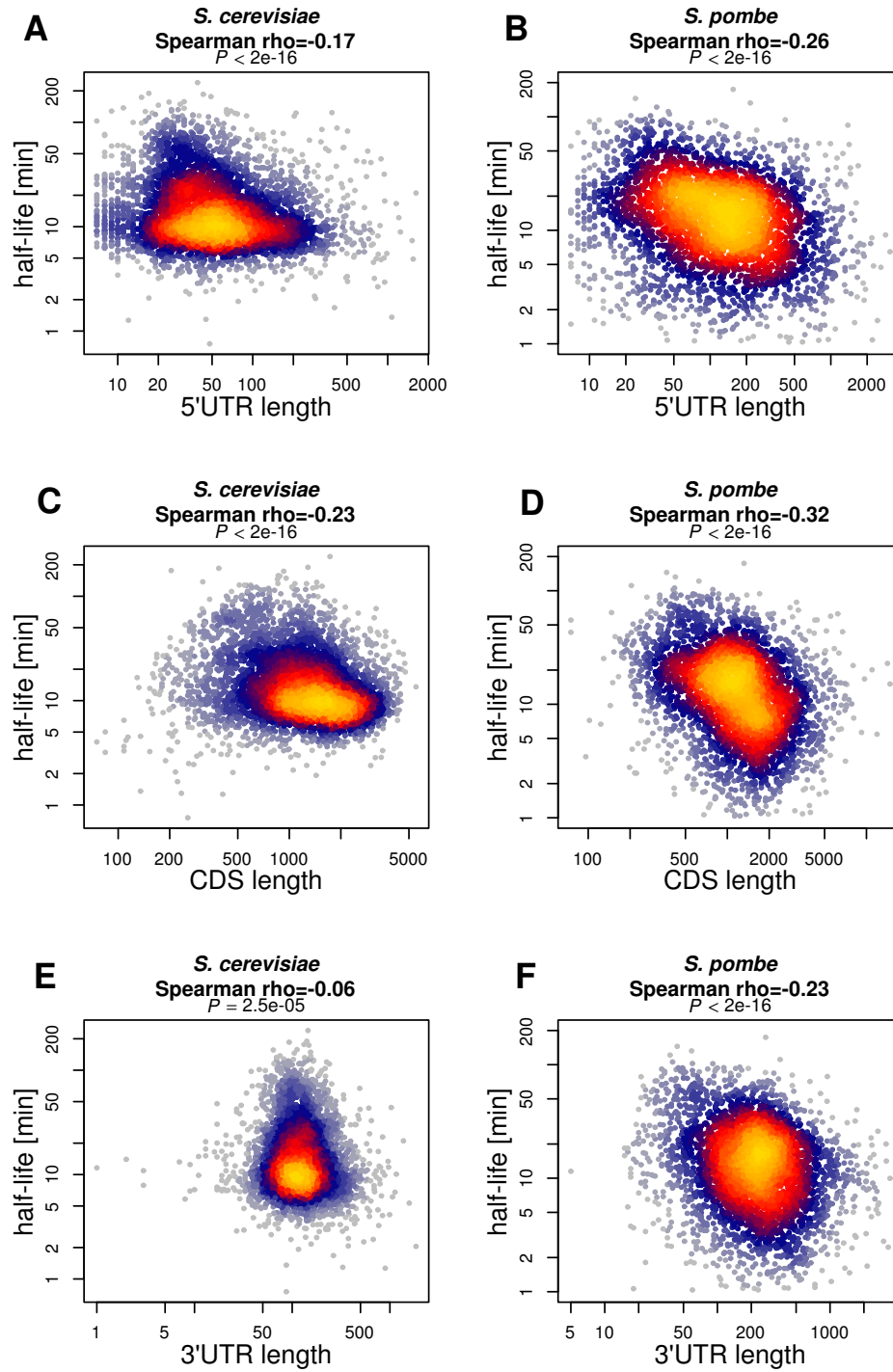[*]Correspondence *gagneur@in.tum.de*

Figure S1: Length of 5′UTR, CDS and 3′UTR correlate with mRNA half-life. (**A-B**) 5′UTR length (x-axis) versus half-life (y-axis) for *S. cerevisiae* (A) and *S. pombe* (B). (**C-D**) CDS length (x-axis) versus half-life (y-axis) for *S. cerevisiae* (C) and *S. pombe* (D). (**E-F**) 3′UTR length (x-axis) versus half-life (y-axis) for *S. cerevisiae* (E) and *S. pombe* (F).
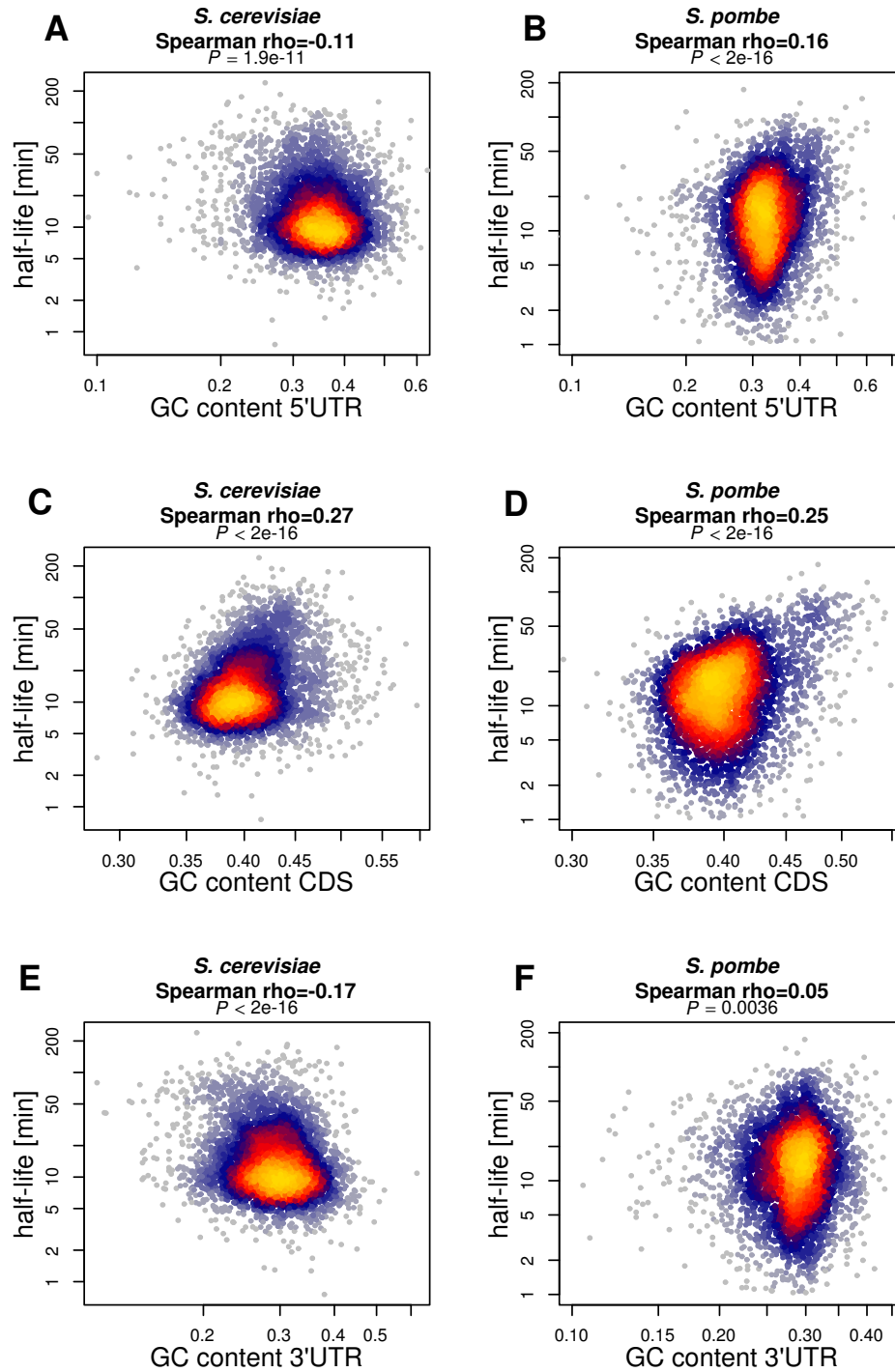
Figure S2: GC content of 5′UTR, CDS and 3′UTR correlate with mRNA half-life (**A-B**) 5′UTR GC content (x-axis) versus half-life (y-axis) for *S. cerevisiae* (A) and *S. pombe*. (**C-D**) CDS GC content (x-axis) versus half-life (y-axis) for *S. cerevisiae* (C) and *S. pombe* (D). (**E-F**) 3′UTR GC content (x-axis) versus half-life (y-axis) for *S. cerevisiae* (E) and *S. pombe* (F).
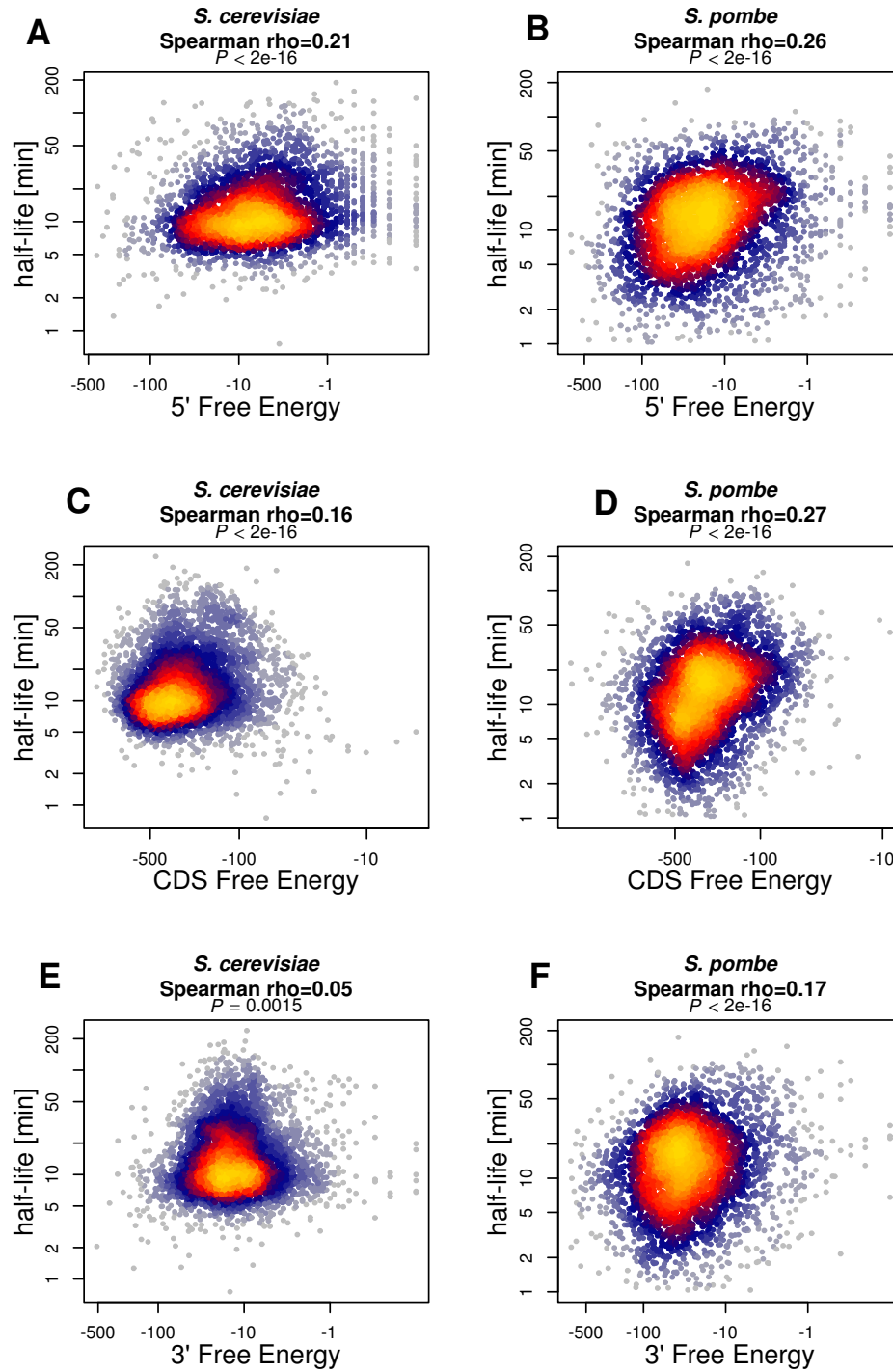
Figure S3: Folding energy of 5′UTR, CDS and 3′UTR correlate with mRNA half-life. (**A-B**) 5' free energy (x-axis) versus half-life (y-axis) for *S. cerevisiae* (A) and *S. pombe* (B). (**C-D**) CDS free energy (x-axis) versus half-life (y-axis) for *S. cerevisiae* (C) and *S. pombe* (D). (**E-F**) 3' free energy (x-axis) versus half-life (y-axis) for *S. cerevisiae* (E) and *S. pombe* (F).
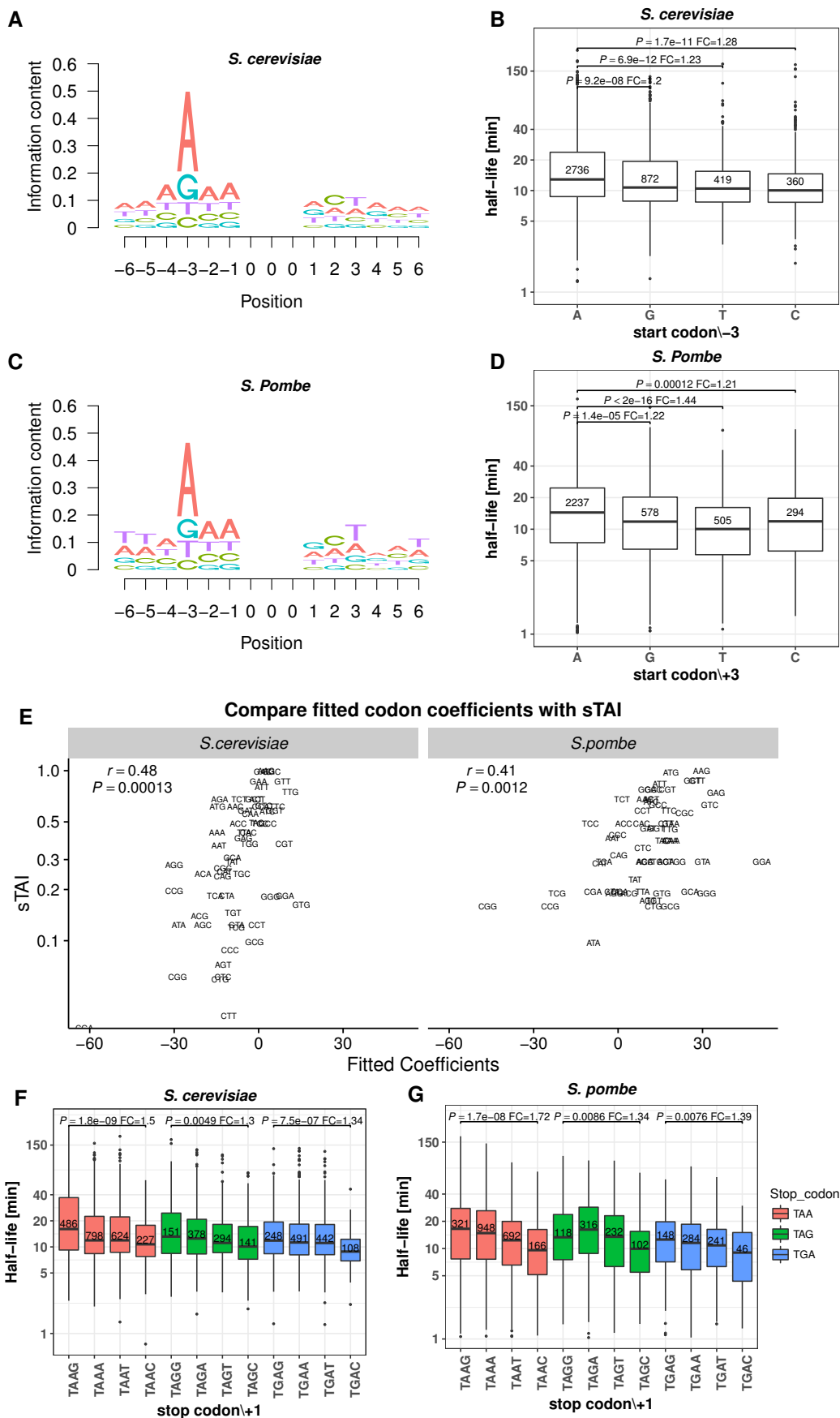
Figure S4: Translation initiation, elongation and termination features associate with mRNA half-life. (**A**) Start codon context (Kozak sequence) generated from 4388 *S. cerevisiae* genes. (**B**) Distribution of half-life for mRNAs grouped by the third nucleotide before the start codon for *S. cerevisiae*. Group sizes (numbers in boxes) show that nucleotide frequency at this position positively associates with half-life. (**C**) Same as (A) for 3713 *S. pombe* genes. (**D**) Same as (B) for *S. pombe*. (**E**) Codon coefficients of linear regression model with codon frequencies as covariates (x-axis) versus species-specific tRNA adaptation index (sTAI) (y-axis, log scale). (**F-G**) Distribution of half-life for mRNAs grouped by the stop codon and the following nucleotide for *S. cerevisiae* (F) and *S. pombe* (G). Colors represent three different stop codons (TAA, TAG and TGA), within each stop codon group, boxes are shown in G, A, T, C order of their following base. Only the P-values for the most drastic pairwise comparisons (A versus C within each stop codon group) are shown. All p-values in boxplot were calculated with Wilcoxon rank-sum test. Boxplots computed as in Figure 3.
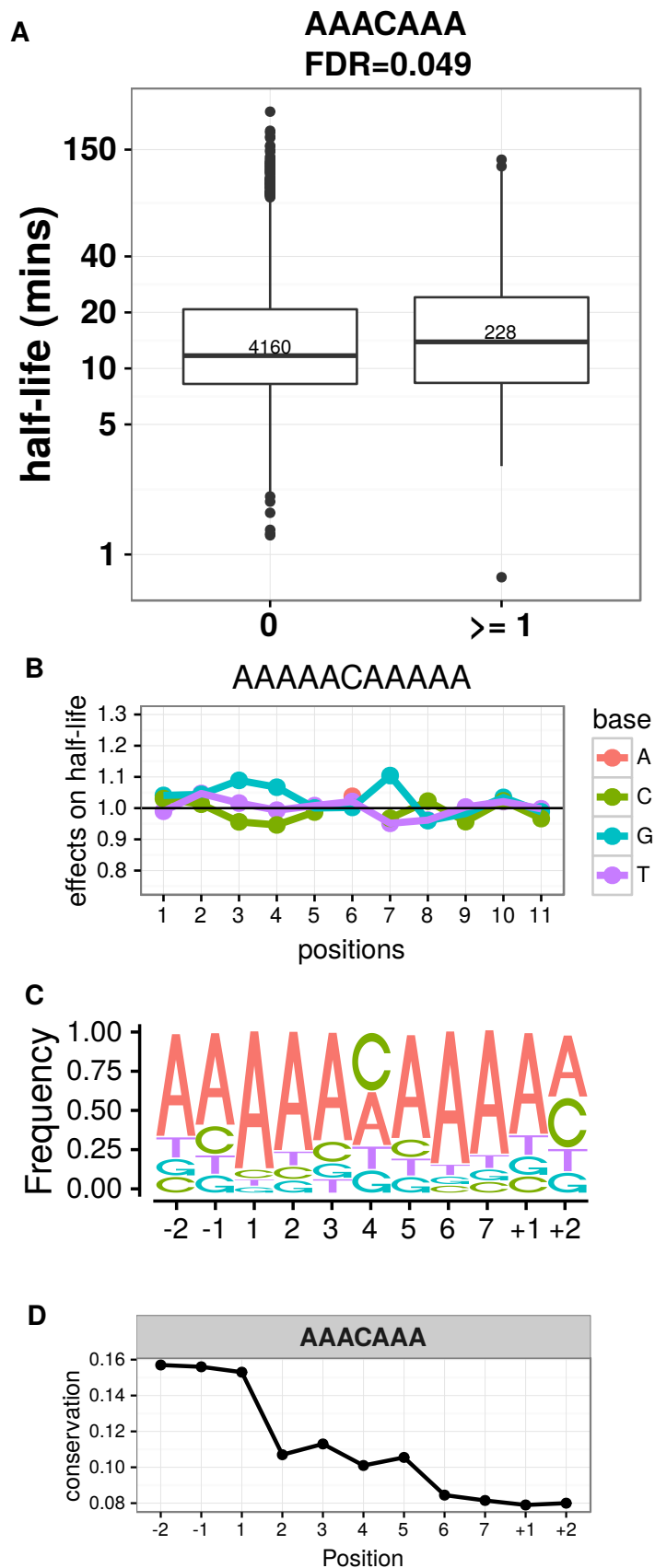
Figure S5: *S. cerevisiae* 5′UTR mRNA half-life associated motif. (**A**) Distribution of half-lives for mRNAs grouped by the number of occurrence(s) of the motif AAACAAA in their 5′UTR sequence. Numbers in the boxes represent the number of members in each box. FDR were reported from the linear mixed effect model (Materials and Methods). (**B**) Prediction of the relative effect on half-life (y-axis) for single-nucleotide substitution in the motif with respect to the consensus motif (y=1, horizontal line). The motifs were extended 2 bases at each flanking site (positions +1, +2, -1, -2). (**C**) Nucleotide frequency within motif instances, when allowing for one mismatch compared to the consensus motif. (**D**) Mean conservation score (phastCons, Materials and Methods) of each base in the consensus motif with 2 flanking nucleotides (y-axis).
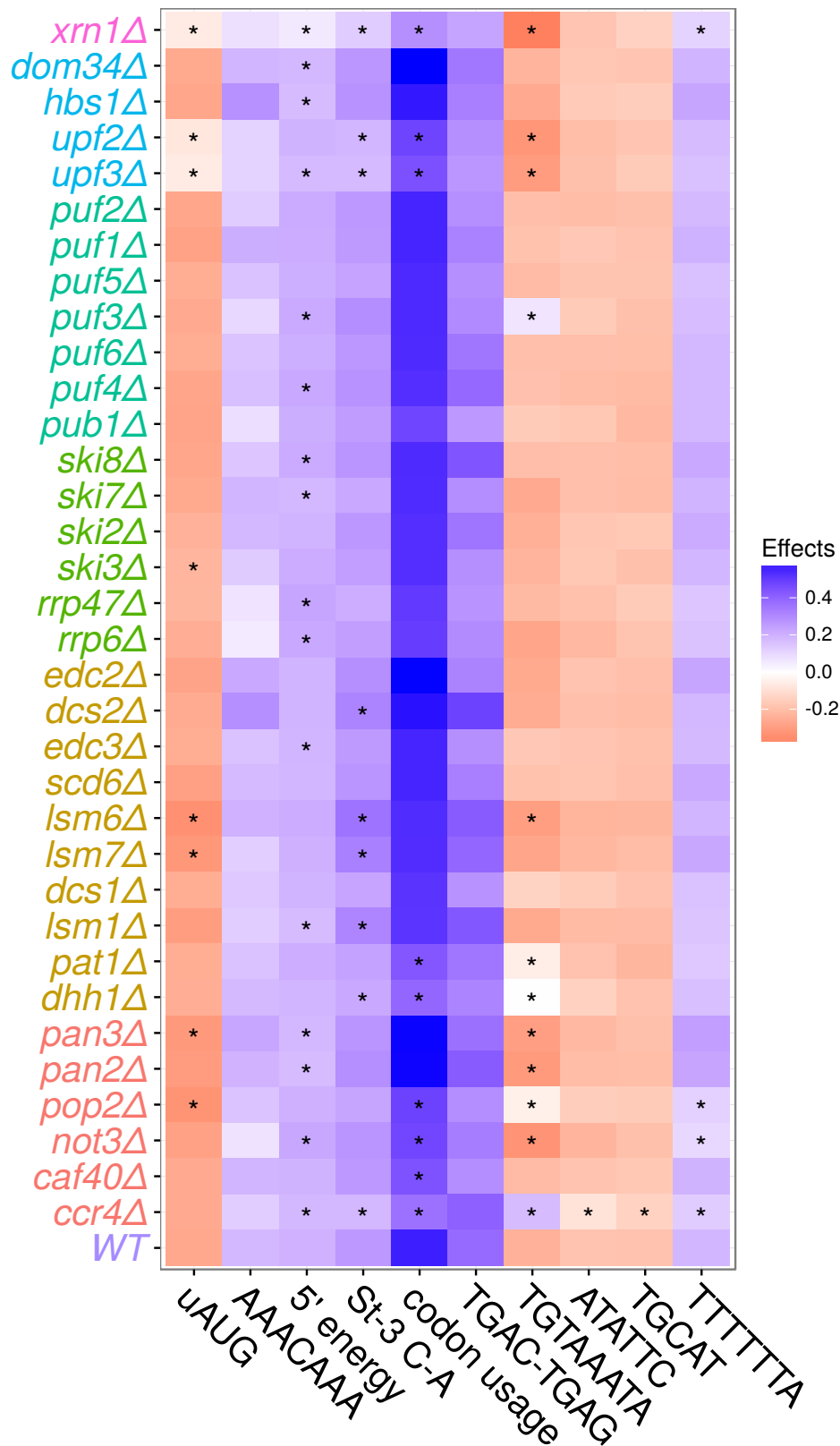
Figure S6: Summary of CREs effect changes across all 34 knockouts comparing with WT. Colour represent the relative effect size (motifs, St-3 C-A, TGAG-TGAC, uAUG), correlation (5' folding energy) or explained variance (codon usage) upon knockout of different genes (y-axis) (Materials and Methods for detailed description). Wild-type label is shown in the bottom (WT) P-values calculated with Wilcoxon rank-sum test by comparing each mutant to wild-type level, multiple testing p-values corrected with Benjamini & Hochberg (FDR). Stars indicating significance of statistical testing (FDR <0.1). 5' energy: correlation of 5′UTR folding energy with mRNA half-lives; St-3 C-A: relative median half-life difference between genes with cytosine and adenine at start codon -3 position; TGAC-TGAG: relative median half-life difference between genes with stop codon +1 TGAC and TGAG. codon usage: codon usage explained mRNA half-lives variance. uAUG: relative median half-life difference between genes without and with upstream AUG in the 5′UTR (Materials and Methods)
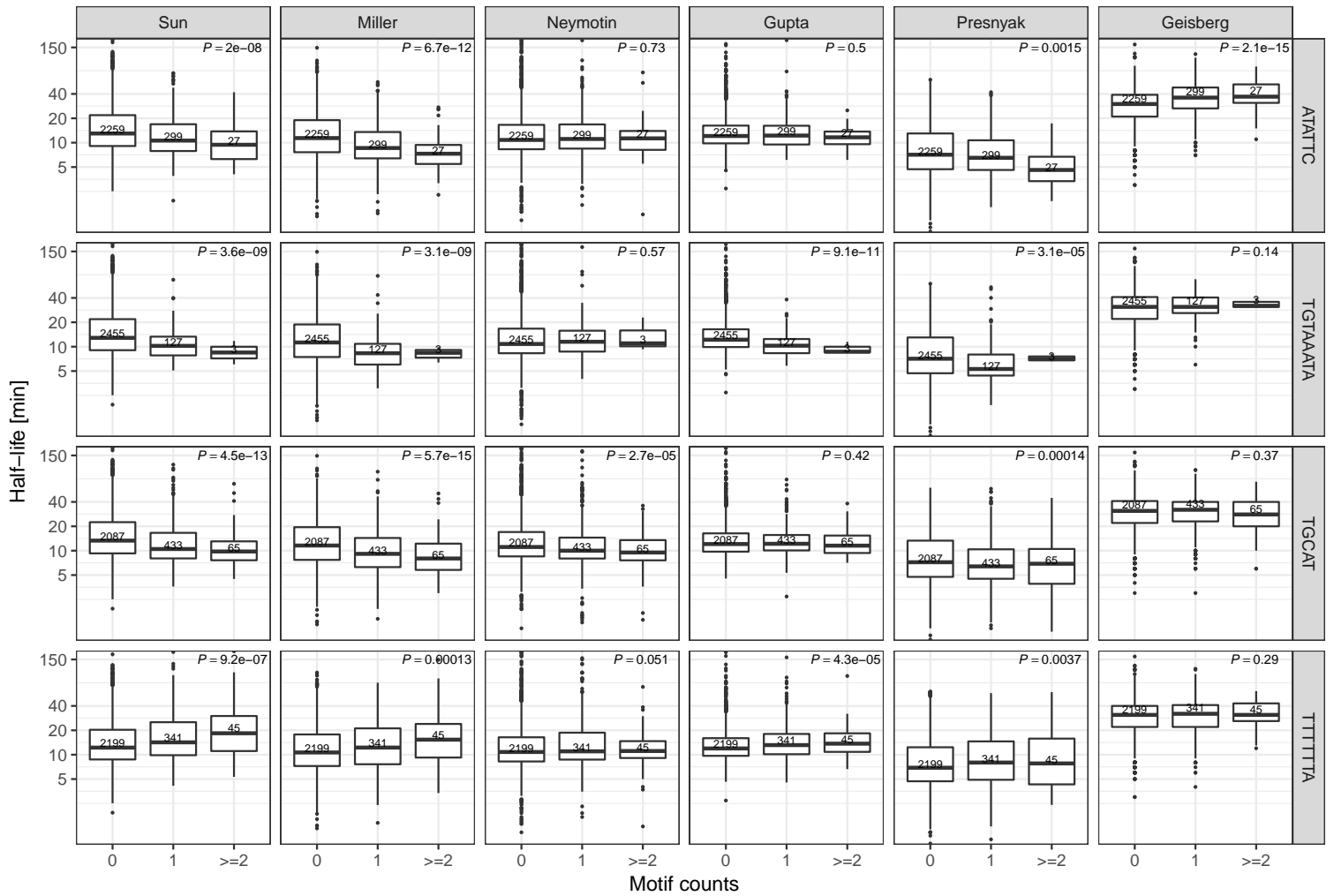
Figure S7: Association of half-life determinant motifs to mRNA half-life in 6 different studies. P-values were calculated with Wilcoxon rank-sum test by comparing half-life of genes without the corresponding motif with genes with the corresponding motif.
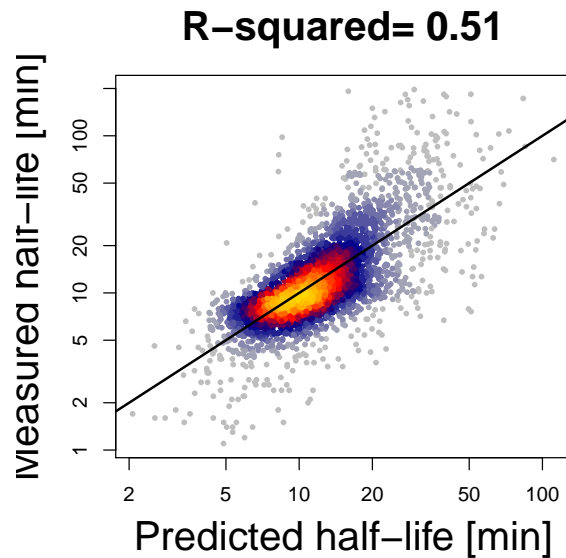

Figure S8: Genome-wide prediction of mRNA half-lives from sequence features with RATE-seq data. mRNA half-lives predicted (x-axis) versus measured (y-axis) with RATE-seq data for 3,539 genes that have complete profiles of all features.
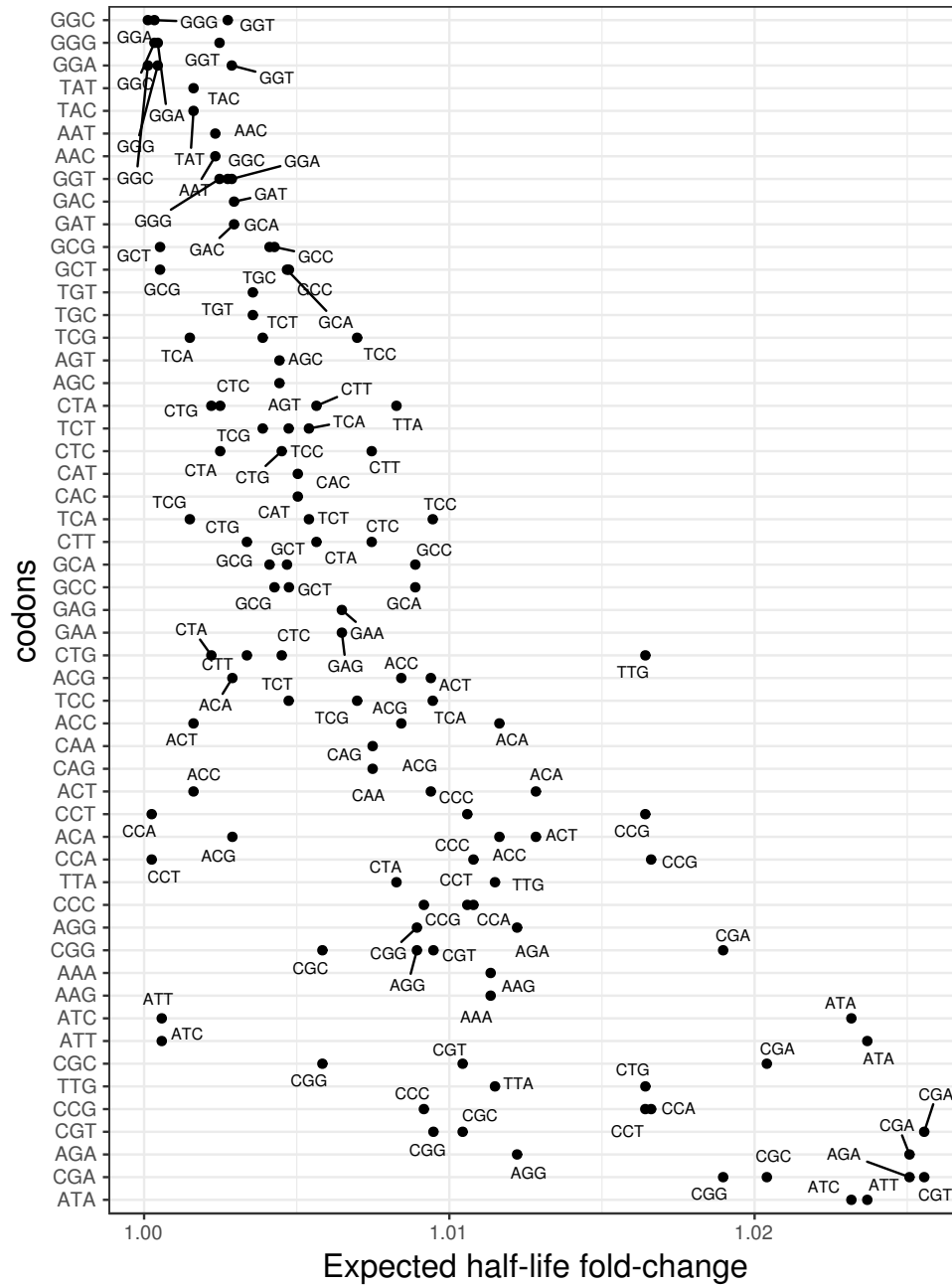
Figure S9: Predicted effects of synonymous codon transitions on half-life. Expected half-life fold-change (x-axis) at each synonymous codon transitions. Each row represent transition from one codon (y-axis) to its synonymous partners. Only synonymous codons that differ by one base were considered.