Expanded View Figures

Figure EV1. Bioinformatics analysis reveals that GC3-content is a determinant of stability.

A Principal component analysis of the CDS codon frequencies of protein-coding genes in *S. cerevisiae*. PC1 and PC2 indicate the first and second principal components.
B PC1 factor loadings of codons from the yeast dataset ranked from the highest to the lowest. The optimal and non-optimal designation at the bottom of the figure refers to the designation according to Presnyak and colleagues [14].
C Pearson's correlation between PC1 factor loading scores and CSC for individual codons (excluding stop codons).
D Pearson's correlation between GC-content and GC3-content for 9,666 protein-coding genes.
E, F Violin plots (E) and cumulative relative frequency distributions (F) visualizing the distribution of mRNA half-lives across their respective GC3-content brackets.
G Comparison of average transcript mRNA half-lives across their respective cAI. Number of transcripts within each range is indicated above their respective points.
H Gene ontology analysis (biological processes) of the top 5% ranked genes in terms of gene GC3-content.
I Gene ontology analysis (biological processes) of the bottom 5% ranked genes in terms of gene GC3-content.

Data information: In (E), the box plots within each figure are indicative of the median and interquartile ranges. In (F), Wilcoxon signed rank tests were performed on the various distributions against the control (all transcripts) group. *P*-values are denoted. In (G), error bars represent the 95% confidence intervals.

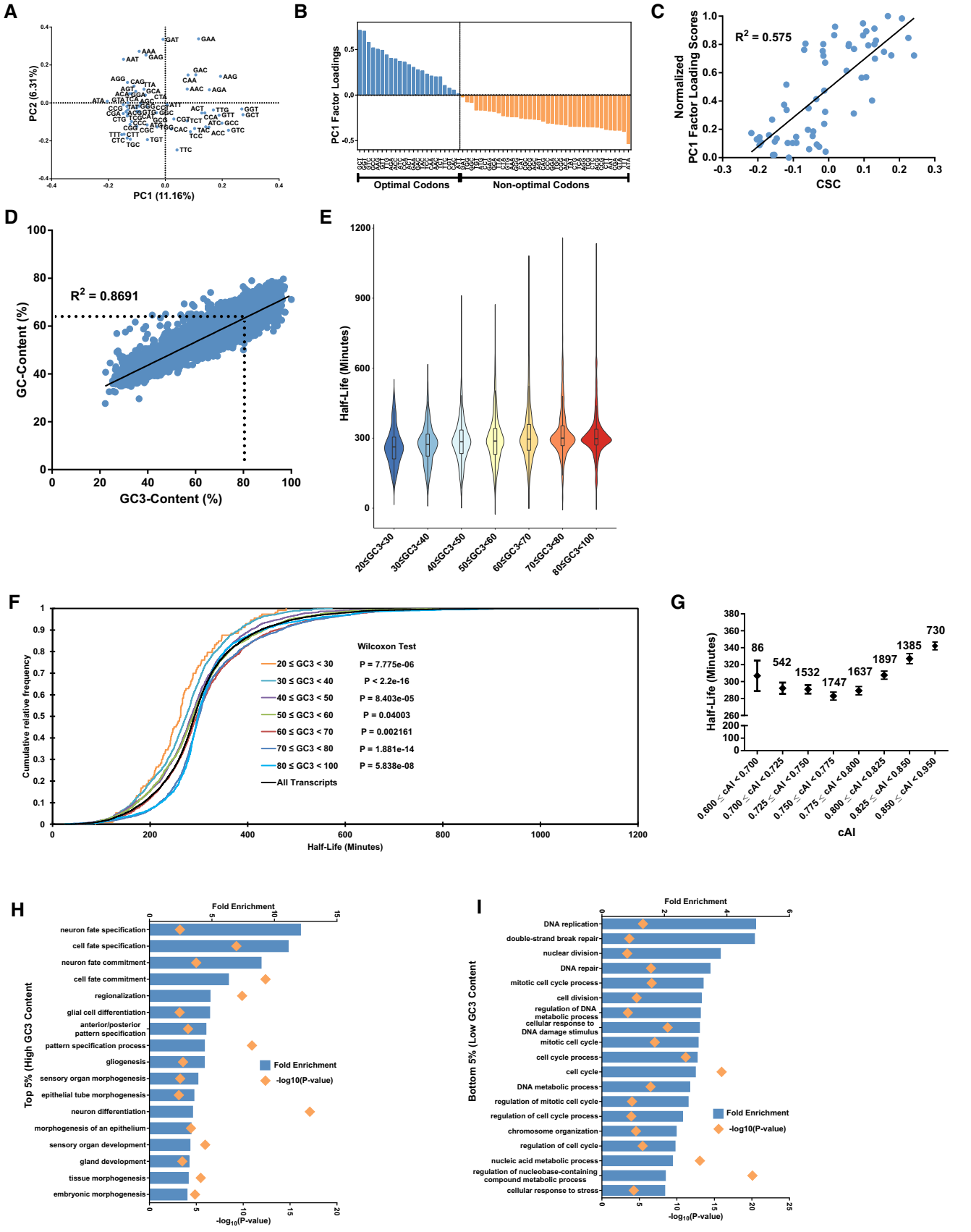
A
PC2 (6.31%)
PC1 (11.16%)
B
PC1 Factor Loadings
Optimal Codons
Non-optimal Codons
C
Normalized PC1 Factor Loading Scores
R2 = 0.575
CSC
D
GC-Content (%)
R2 = 0.8691
GC3-Content (%)
E
Half-Life (Minutes)
20≤GC3<30
30≤GC3<40
40≤GC3<50
50≤GC3<60
60≤GC3<70
70≤GC3<80
80≤GC3<100
F
Cumulative relative frequency
Half-Life (Minutes)
Wilcoxon Test
20 ≤ GC3 < 30 P = 7.775e-06
30 ≤ GC3 < 40 P < 2.2e-16
40 ≤ GC3 < 50 P = 8.403e-05
50 ≤ GC3 < 60 P = 0.04003
60 ≤ GC3 < 70 P = 0.002161
70 ≤ GC3 < 80 P = 1.881e-14
80 ≤ GC3 < 100 P = 5.838e-08
All Transcripts
G
Half-Life (Minutes)
86
542
1532
1747
1637
1897
1385
730
0.600 ≤ cAI < 0.700
0.700 ≤ cAI < 0.725
0.725 ≤ cAI < 0.750
0.750 ≤ cAI < 0.775
0.775 ≤ cAI < 0.800
0.800 ≤ cAI < 0.825
0.825 ≤ cAI < 0.850
0.850 ≤ cAI < 0.950
cAI
H
Fold Enrichment
Top 5% (High GC3 Content
neuron fate specification
cell fate specification
neuron fate commitment
cell fate commitment
regionalization
glial cell differentiation
anterior/posterior pattern specification
pattern specification process
gliogenesis
sensory organ morphogenesis
epithelial tube morphogenesis
neuron differentiation
morphogenesis of an epithelium
sensory organ development
gland development
tissue morphogenesis
embryonic morphogenesis
Fold Enrichment
-log10(P-value)
-log10(P-value)
I
Fold Enrichment
Bottom 5% (Low GC3 Content
DNA replication
double-strand break repair
nuclear division
DNA repair
mitotic cell cycle process
cell division
regulation of DNA metabolic process
cellular response to DNA damage stimulus
mitotic cell cycle
cell cycle process
cell cycle
DNA metabolic process
regulation of mitotic cell cycle
regulation of cell cycle process
chromosome organization
regulation of cell cycle
nucleic acid metabolic process
regulation of nucleobase-containing compound metabolic process
cellular response to stress
Fold Enrichment
-log10(P-value)
-log10(P-value)

Figure EV1.

Figure EV2. GC3-content can explain ribosome occupancy and translation efficiency to a certain extent.

A PC1 factor loadings of codons from the human dataset ranked from the highest to the lowest (bottom) and their corresponding normalized factor loadings after linear normalization onto a percentage scale (top).

B Pearson's correlation between the correlations derived from comparison of ribosome occupancy and codon bias-derived scores for two ribosome profiling sample replicates.

C Pearson's correlation between codon bias-derived occupancy scores and ribosome occupancy for individual codons (excluding stop codons).

D Three example transcripts (EIF2B2, DYNC1LI2, and IDH3G) that demonstrate high correlation between ribosome occupancy and codon bias-derived scores (left) as well as their corresponding Pearson's correlations over 25 bins (right).

E Comparison of average transcript translation efficiencies (TEs) across their respective GC3-content ranges after grouping by mRNA abundances. Error bars represent the 95% confidence intervals.

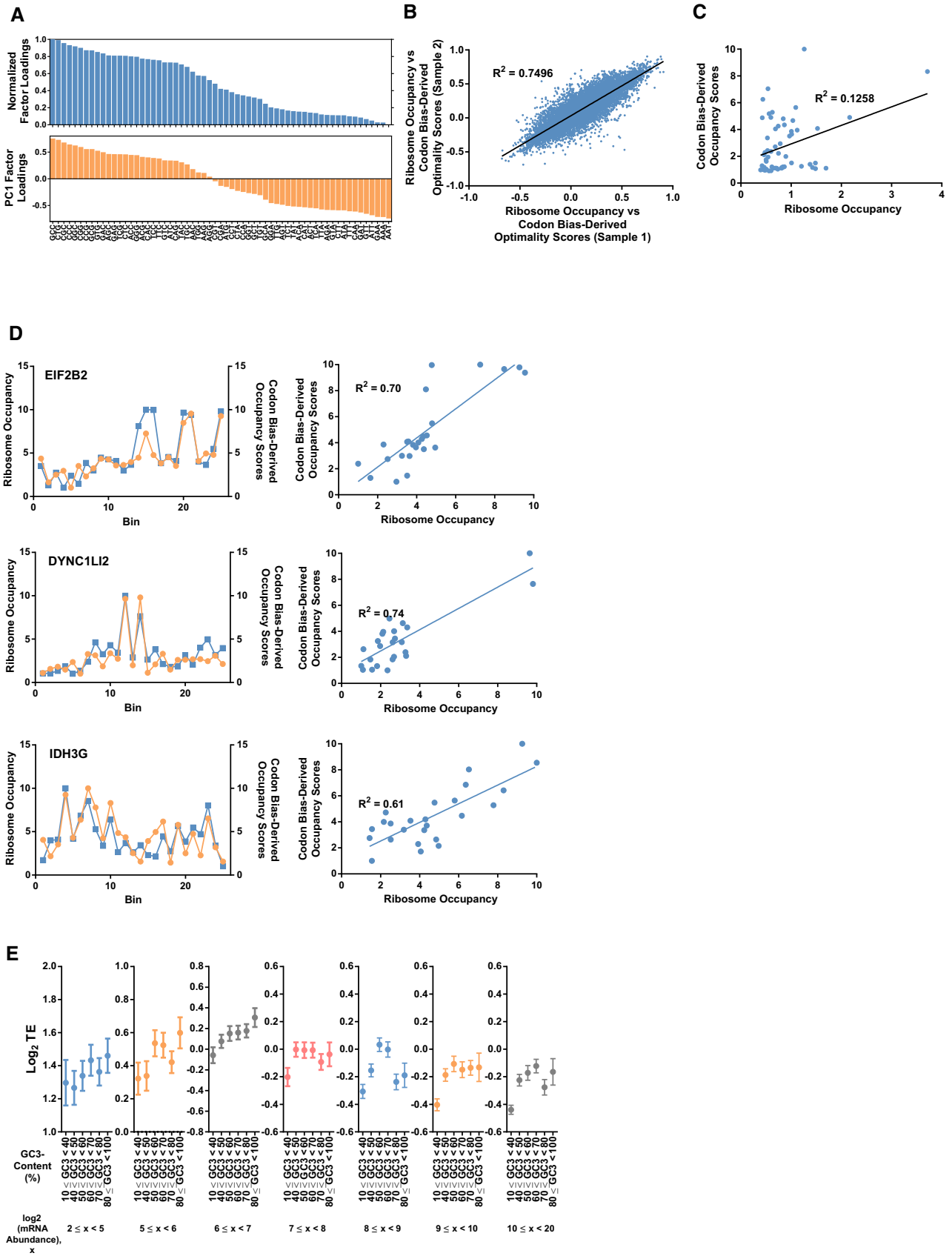
A
Normalized Factor Loadings
PC1 Factor Loadings
B
Ribosome Occupancy vs Codon Bias-Derived Optimality Scores (Sample 2)
Ribosome Occupancy vs Codon Bias-Derived Optimality Scores (Sample 1)
R2 = 0.7496
C
Codon Bias-Derived Occupancy Scores
Ribosome Occupancy
R2 = 0.1258
D
EIF2B2
DYNC1LI2
IDH3G
Ribosome Occupancy
Codon Bias-Derived Occupancy Scores
Bin
R2 = 0.70
R2 = 0.74
R2 = 0.61
E
Log2 TE
GC3-Content (%)
log2 (mRNA Abundance), x
2 ≤ x < 5
5 ≤ x < 6
6 ≤ x < 7
7 ≤ x < 8
8 ≤ x < 9
9 ≤ x < 10
10 ≤ x < 20

Figure EV2.

Figure EV3. GC3-content of transcripts affects translation efficiency and stability.

A Example of how transcript GC3 optimization and GC3 deoptimization were performed to generate GC3-optimized and GC3-deoptimized versions of *REL* and *IL6* transcripts.

B GC3- and GC-content of *REL-OPT/WT*, *REL-OPT* (+1 Frameshift), and *IL6-OPT/WT/DE* and *IL6-OPT* (+1 Frameshift) transcripts.

C Protein abundance of immunoblot of FLAG-tagged REL-OPT and REL-WT in HEK293T cells (normalized by respective mRNA levels) transfected with either empty plasmids, or plasmids bearing REL-OPT or REL-WT (corresponding to Fig 3B). The data are representative of threew independent experiments. The respective steady-state mRNA levels (transcript copy numbers) are shown on the right.

D IL6 protein abundance as determined by ELISA of IL6-OPT, IL6-WT, and IL6-DE in HEK293T cells (normalized by respective mRNA levels) transfected with either empty plasmids, or plasmids bearing IL6-OPT, IL6-WT, or IL6-DE (corresponding to Fig 3C). The ELISA quantification is representative of three independent experiments. The respective steady-state mRNA levels (transcript copy numbers) are shown on the right.

E Representative immunoblot of FLAG-tagged REL-OPT and REL-WT expressed in HeLa cells (left). The immunoblot is representative of three independent experiments.

F mRNA stability experiments showing the degradation of *REL-OPT* and *REL-WT* transcripts in HeLa cells, post-actinomycin-D addition.

Data information: In (C), the densitometry data are representative of three independent experiments. Unpaired *t*-tests were performed within the REL-OPT and REL-WT samples, $P < 0.05$ (*). In (D), a one-way ANOVA with Tukey's multiple comparisons was performed between samples where $P < 0.01$ (**) and $P < 0.001$ (***). In (F), data are representative of three independent experiments each with three replicates. The data represent the mean $\pm$ SD for three replicates. A two-way ANOVA with the Holm–Sidak multiple comparisons was performed. *P*-values are denoted as follows: $P < 0.001$ (***).

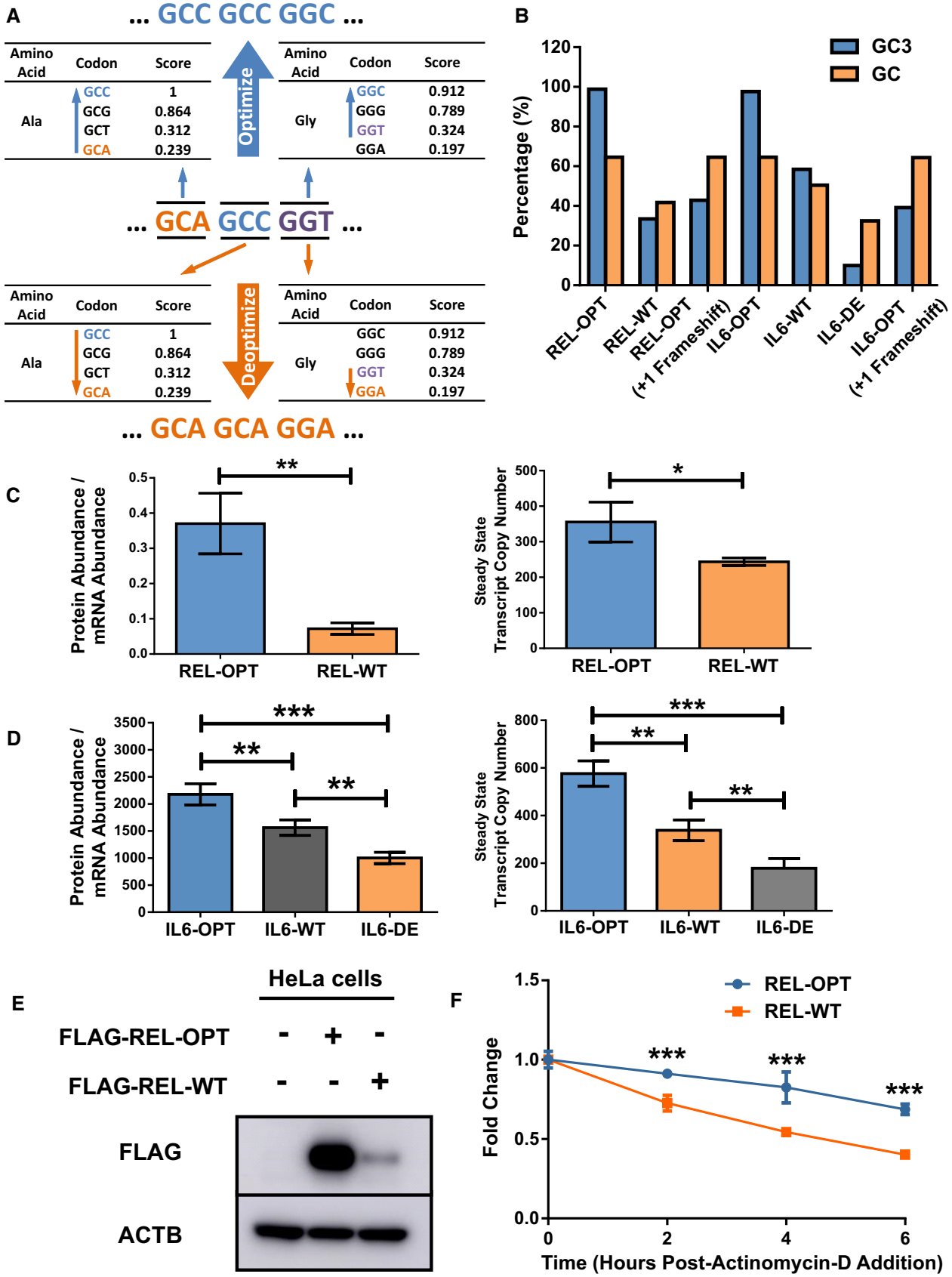
A
... GCC GCC GGC ...
Amino Acid
Codon
Score
Ala
GCC 1
GCG 0.864
GCT 0.312
GCA 0.239
Gly
GGC 0.912
GGG 0.789
GGT 0.324
GGA 0.197
Optimize
... GCA GCC GGT ...
Deoptimize
... GCA GCA GGA ...
B
GC3
GC
Percentage (%)
REL-OPT
REL-WT
REL-OPT (+1 Frameshift)
IL6-OPT
IL6-WT
IL6-DE
IL6-OPT (+1 Frameshift)
C
Protein Abundance / mRNA Abundance
Steady State Transcript Copy Number
D
IL6-OPT
IL6-WT
IL6-DE
E
HeLa cells
FLAG-REL-OPT
FLAG-REL-WT
FLAG
ACTB
F
Fold Change
Time (Hours Post-Actinomycin-D Addition)

Figure EV3.

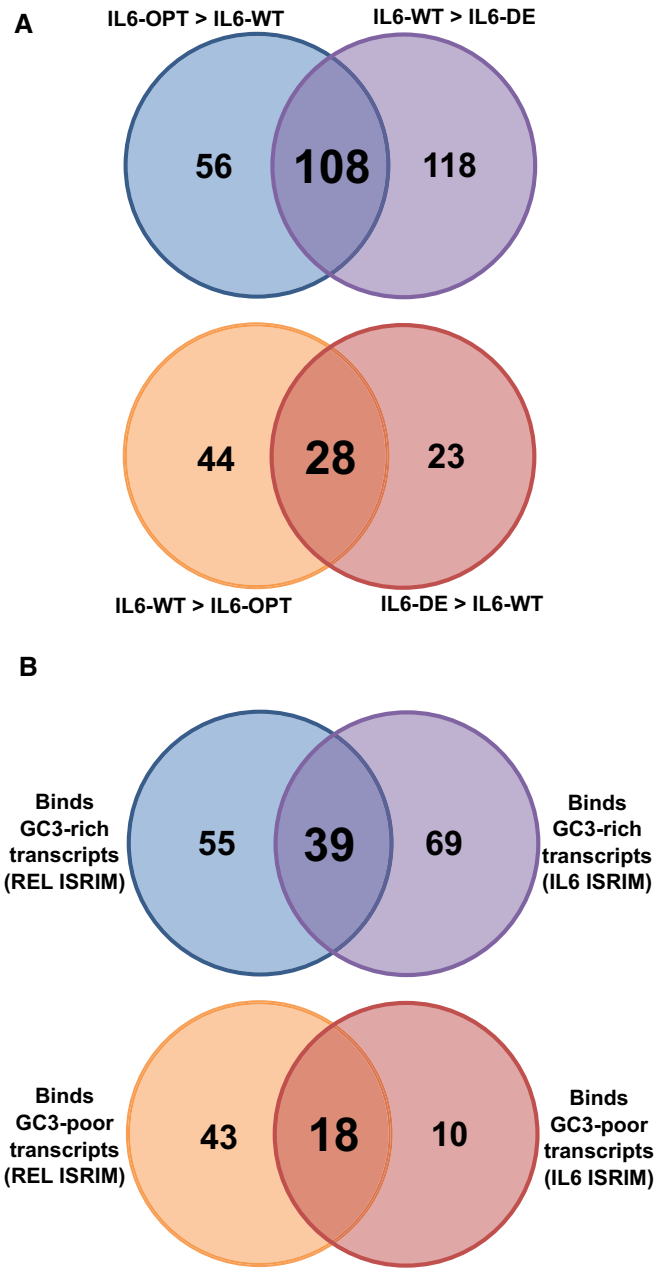

Figure EV4. RNA-binding proteins can be identified from ISRIM experiments.

A Venn diagram indicating the number of RBPs identified from the IL6 ISRIM experiments.

B Venn diagram indicating the number of RBPs identified from the REL and IL6 ISRIM experiments.

Figure EV5. ILF2 is an RNA-binding protein that can bind differentially to transcripts with different levels of GC3-content.

A Cumulative distribution plots showing the GC3-content distribution of transcripts bound to by ILF2 in H929 (top) and JJN3 cells (bottom). Wilcoxon signed rank tests were performed on the ILF2 RIP group against the control group. *P*-values are denoted in the figure.

B Scatterplot of the RPKM values of mRNA transcripts in K562 cells subject to ILF2 CRISPR interference and its corresponding WT control. mRNA transcripts are colored according to their respective GC3-content.

C Fold changes of example mRNA representing low, average, and high GC3-content transcripts from the RPKM values of mRNA transcripts in K562 cells subject to ILF2 CRISPR interference.

D Densitometric analysis of immunoblot of FLAG-tagged REL-OPT and REL-WT expressed in HEK293T cells under ILF2 and ILF3 siRNA treatment (corresponding to Fig 6D).

E Densitometric analysis of immunoblot of FLAG-tagged REL-OPT and REL-WT expressed in HEK293T cells co-expressed with two different isoforms of ILF2 (corresponding to Fig 6E).

F, G Top three RNA motifs enriched in upregulated transcripts (> 5 fold) in ILF2 RIP-seq data (A) derived from both H929 and JJN3 datasets (left) and their corresponding annotations in transcripts (right) (F), followed by the ILF3 motif and its distribution identified by Dotu *et al* [47] from ILF3 RNA Bind-n-seq experiments (G).

Data information: In (D, E), densitometry data are representative of three independent experiments. A one-way ANOVA with Tukey's multiple comparisons was performed within the REL-OPT and REL-WT samples. *P*-values are denoted as follows: $P < 0.05$ (*), $P < 0.01$ (**), and $P < 0.001$ (***).

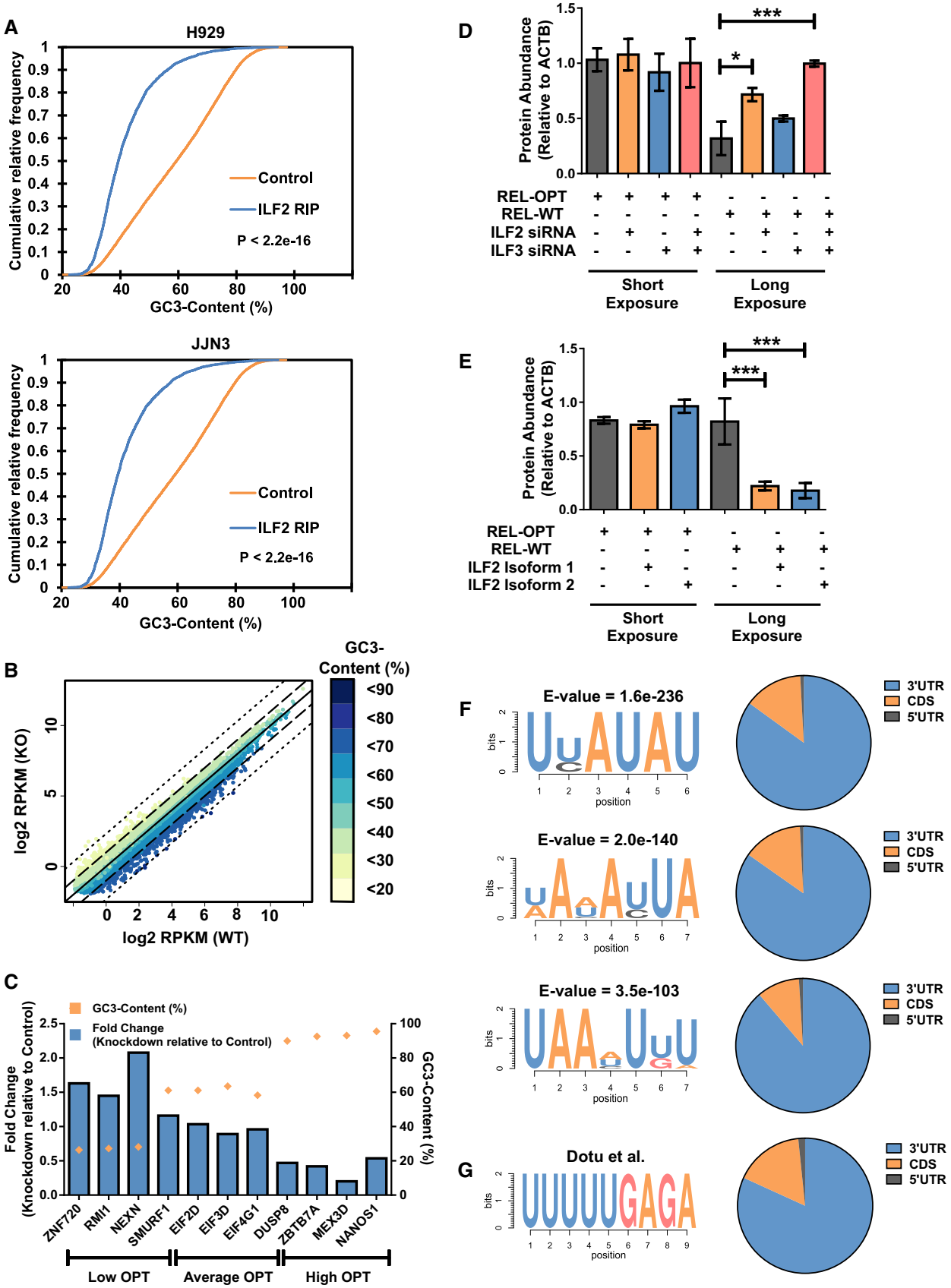
A
H929
Cumulative relative frequency
GC3-Content (%)
Control
ILF2 RIP
P < 2.2e-16
JJN3
Cumulative relative frequency
GC3-Content (%)
Control
ILF2 RIP
P < 2.2e-16
B
log2 RPKM (KO)
log2 RPKM (WT)
GC3-Content (%)
<90
<80
<70
<60
<50
<40
<30
<20
C
GC3-Content (%)
Fold Change (Knockdown relative to Control)
Fold Change (Knockdown relative to Control)
GC3-Content (%)
ZNF720
RMI1
NEXN
SMURF1
EIF2D
EIF3D
EIF4G1
DUSP8
ZBTB7A
MEX3D
NANOS1
Low OPT
Average OPT
High OPT
D
Protein Abundance (Relative to ACTB)
***
*
REL-OPT
REL-WT
ILF2 siRNA
ILF3 siRNA
Short Exposure
Long Exposure
E
Protein Abundance (Relative to ACTB)
***
***
REL-OPT
REL-WT
ILF2 Isoform 1
ILF2 Isoform 2
Short Exposure
Long Exposure
F
E-value = 1.6e-236
E-value = 2.0e-140
E-value = 3.5e-103
bits
position
3'UTR
CDS
5'UTR
G
Dotu et al.
bits
position
3'UTR
CDS
5'UTR

Figure EV5.