Molecular Cell, Volume *66*

Supplemental Information

Translation of CircRNAs

Nagarjuna Reddy Pamudurti, Osnat Bartok, Marvin Jens, Reut Ashwal-Fluss, Christin Stottmeister, Larissa Ruhe, Mor Hanan, Emanuel Wyler, Daniel Perez-Hernandez, Evelyn Ramberger, Shlomo Shenzis, Moshe Samson, Gunnar Dittmar, Markus Landthaler, Marina Chekulaeva, Nikolaus Rajewsky, and Sebastian Kadener

Figure S1 related to Figure 2

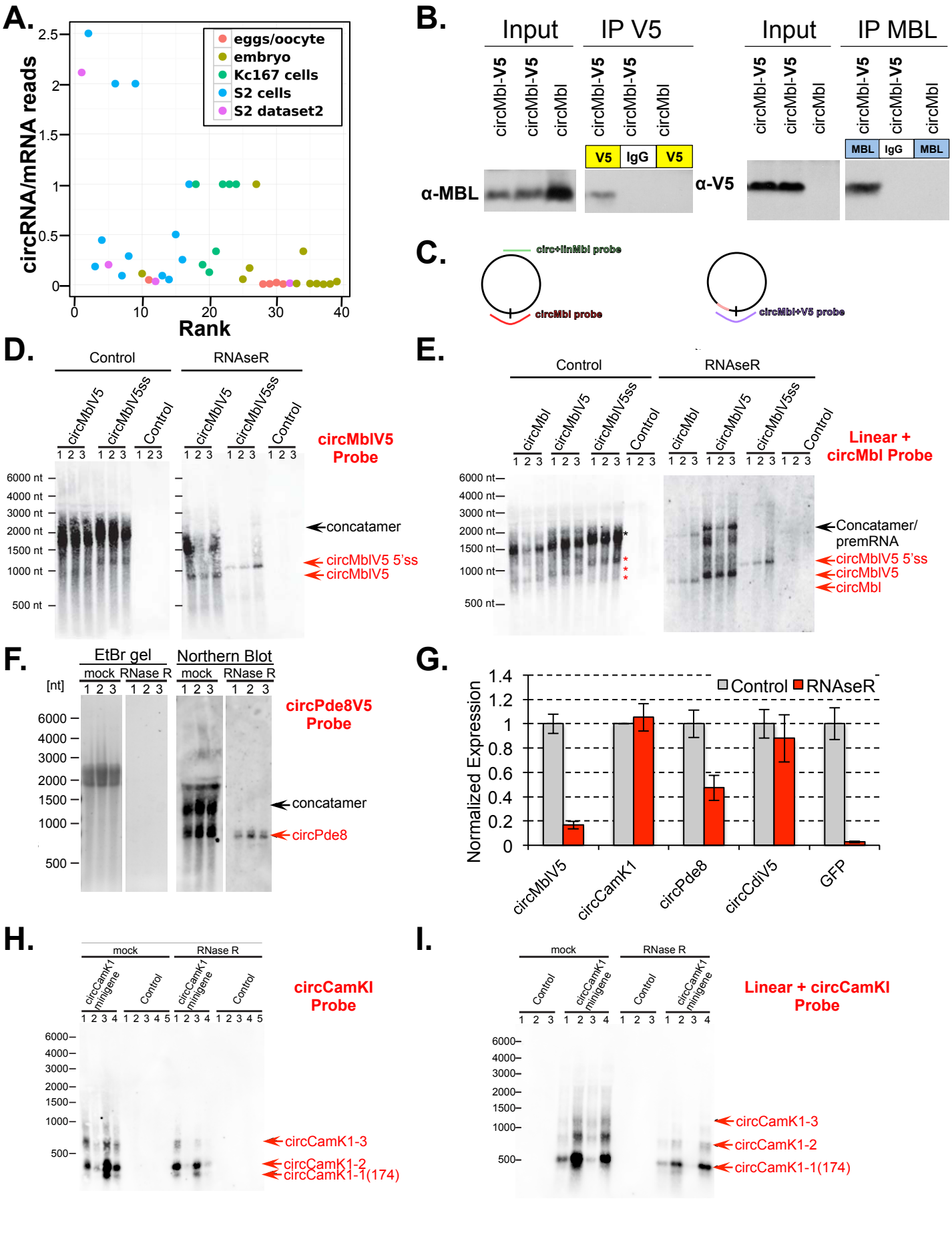


Figure S2 related to Figure 2

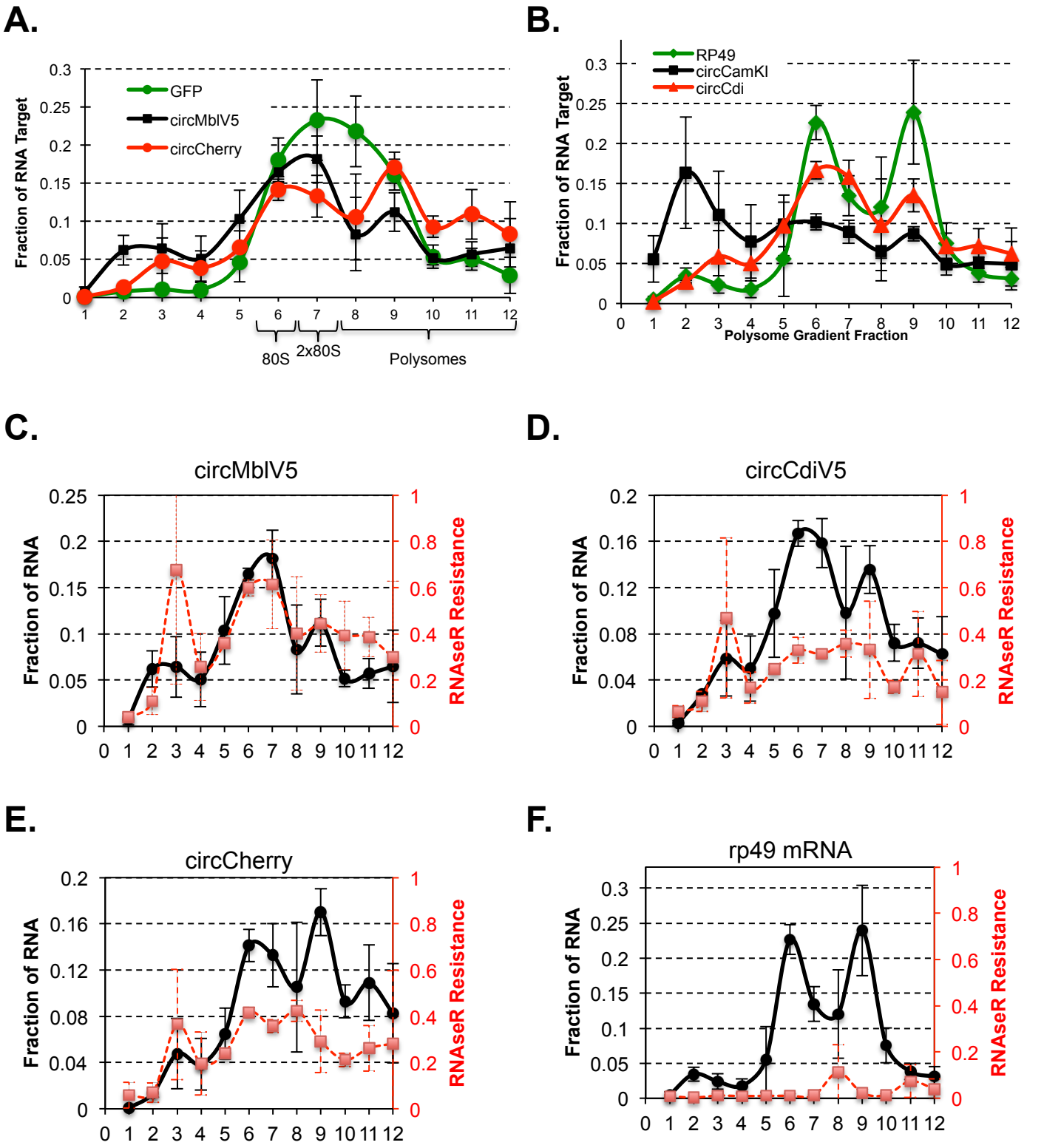


Figure S3 related to Figure 3

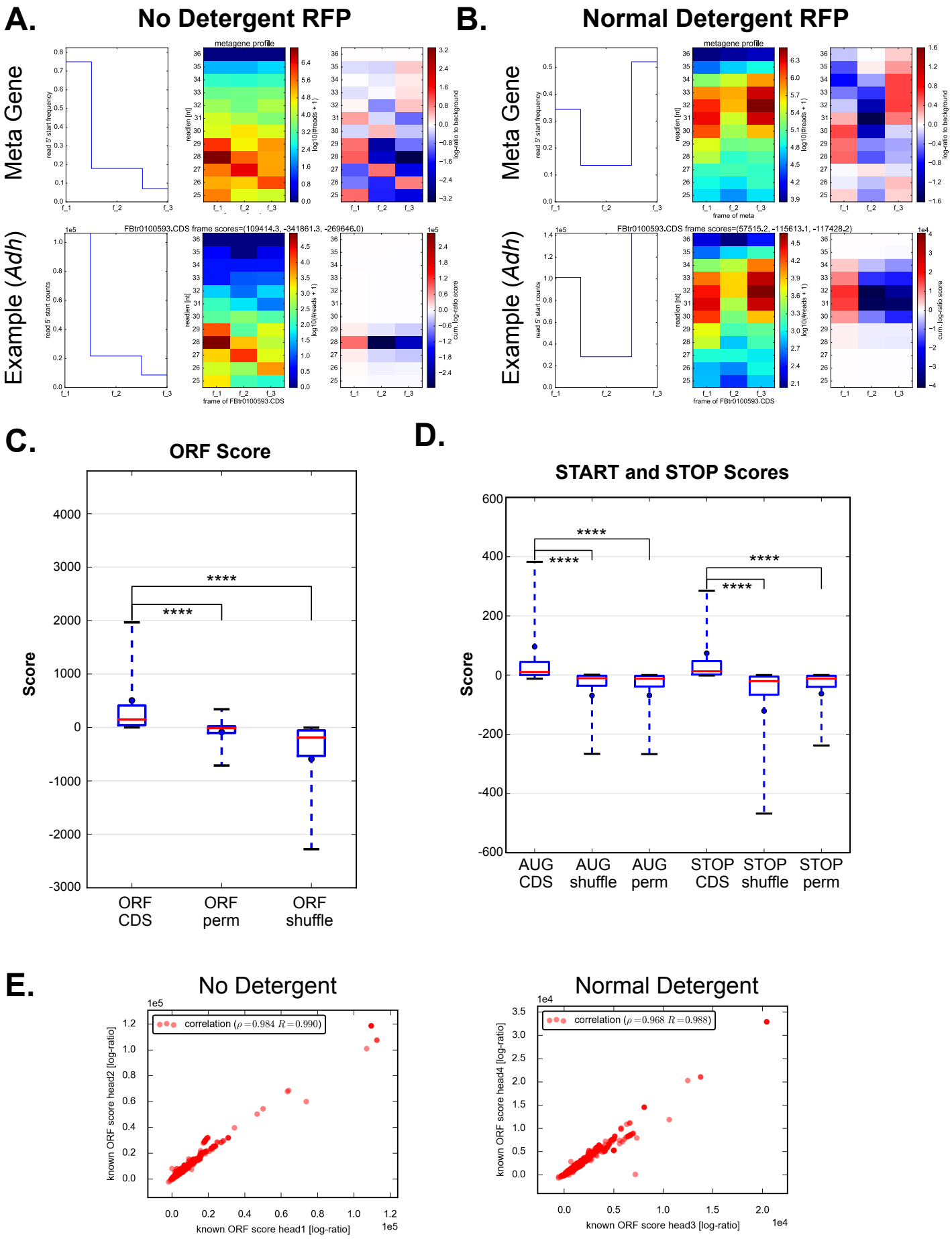


Figure S4 related to Figure 4

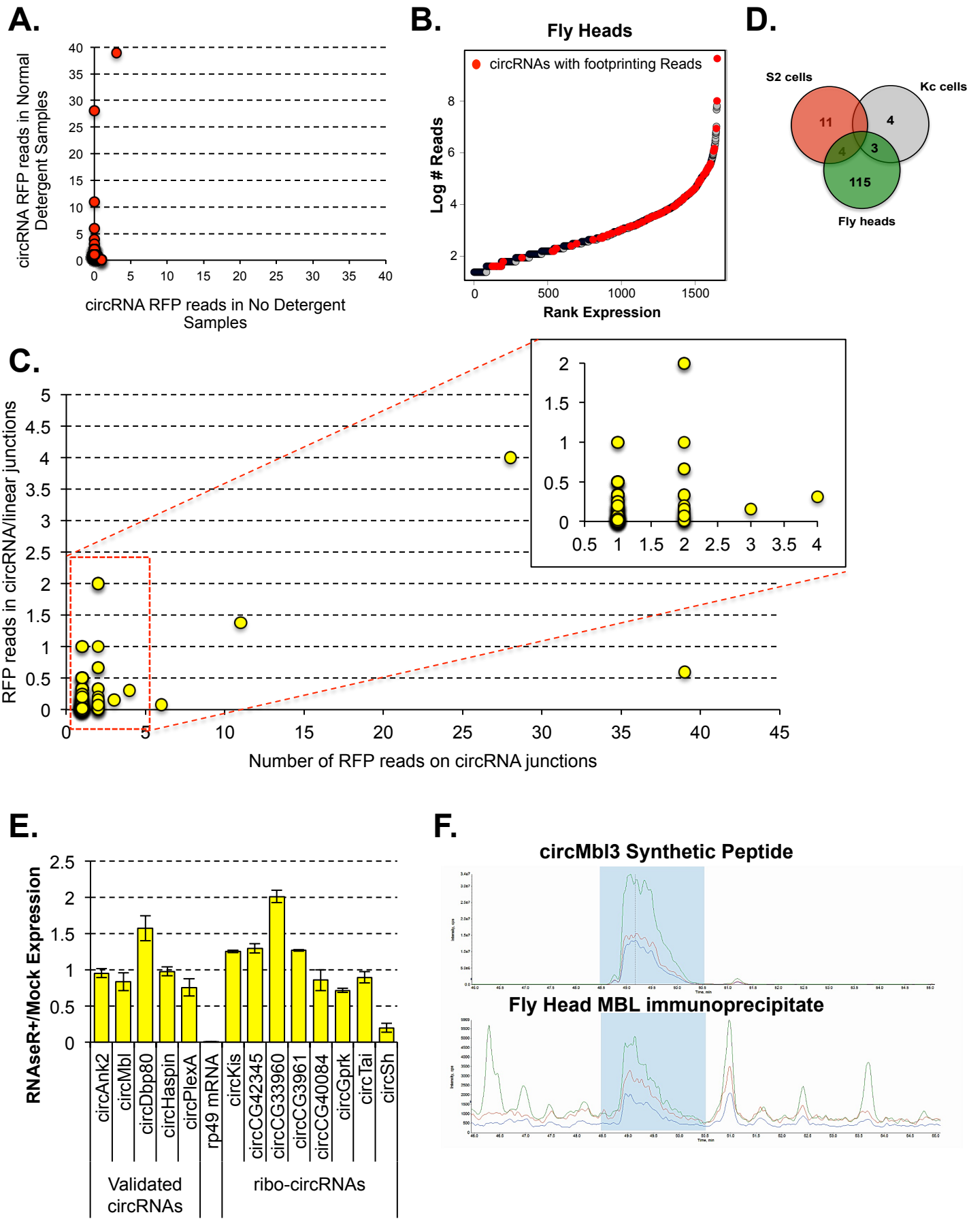


Figure S5 related to Figure 5

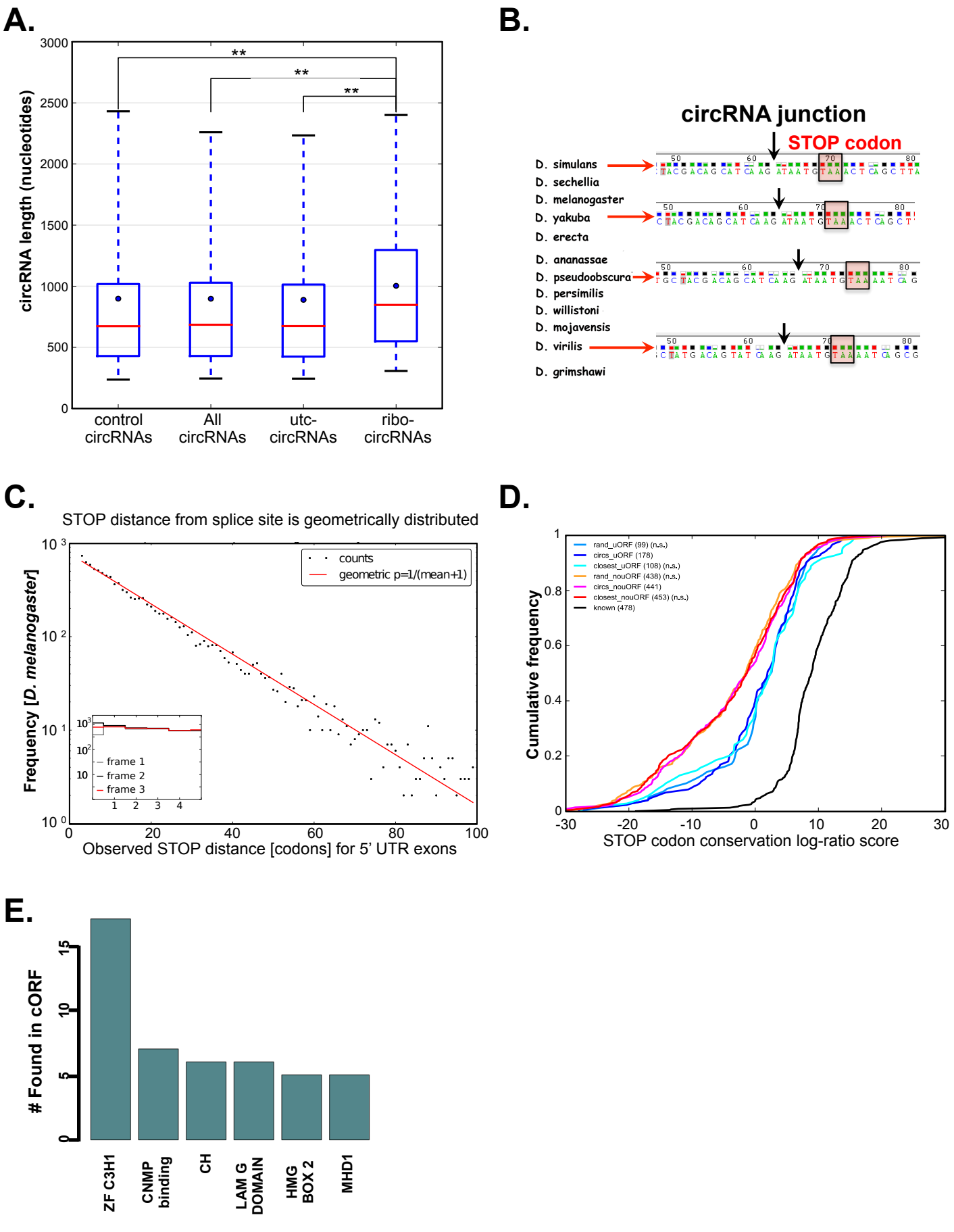


Figure S6 related to Figure 5

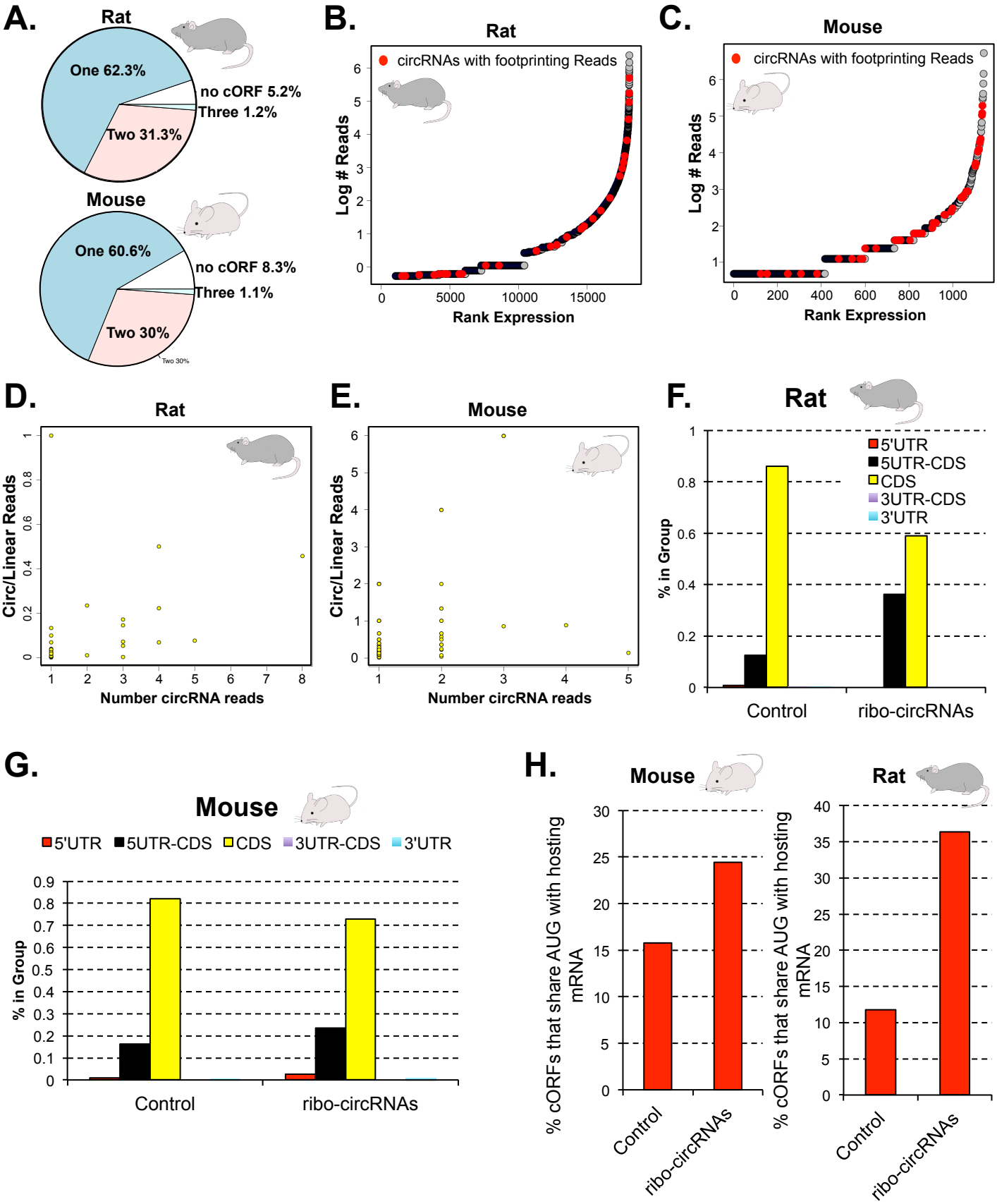
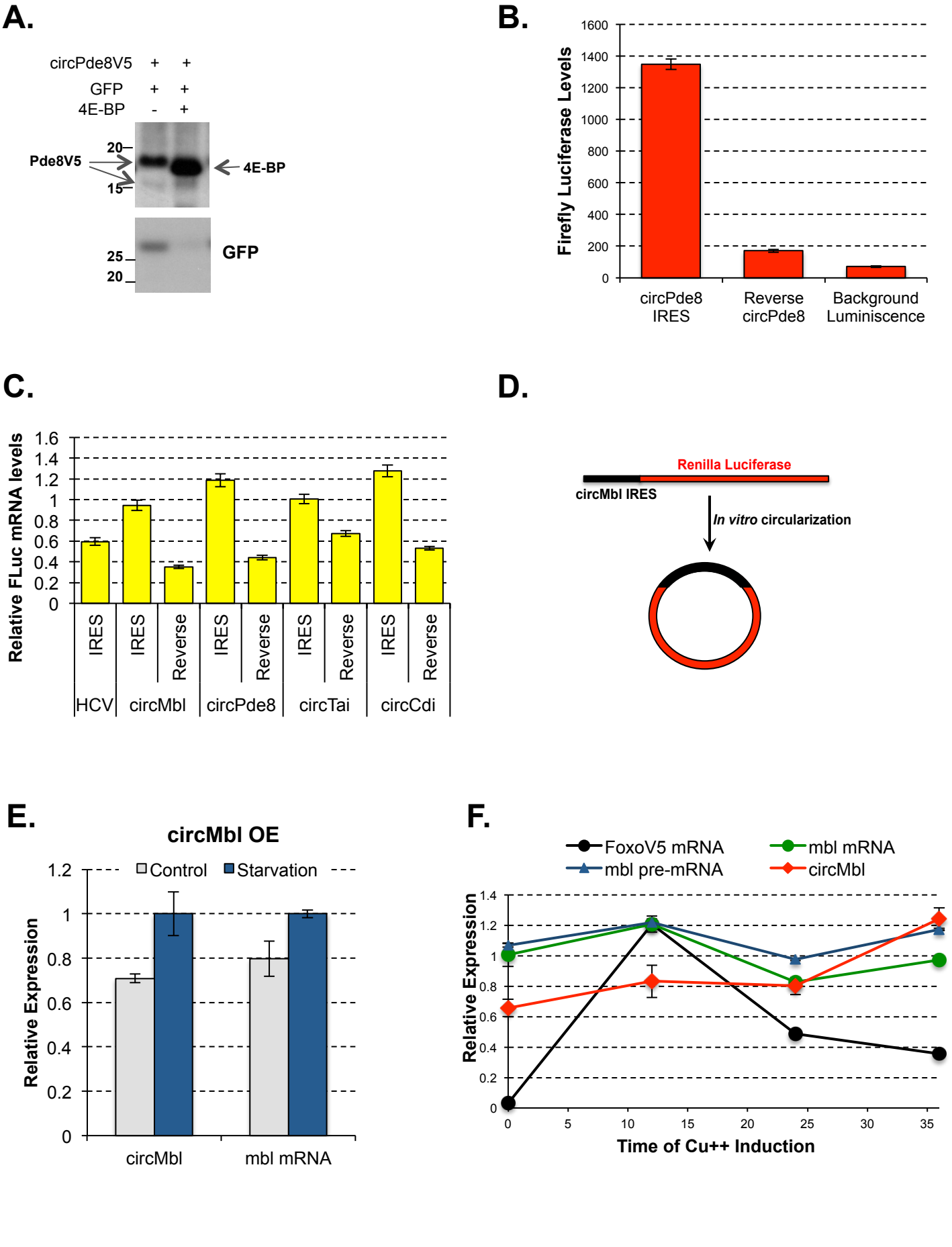


Figure 7

Figure S7 related to Figure 6



SUPPLEMENTAL FIGURE AND TABLE LEGENDS:

Legends for Figure S1-S7

Figure S1 related to Figure 2:

A. Non-ambiguous circRNAs foot-printing reads are in the same scale as the ones in junctions of the same exons. **B.** Both anti-V5 and anti-MBL antibodies recognize and precipitate (IP) the protein produced from circMblV5. **C.** Scheme of the probe design for the detection of circRNA and linear molecules by Northern Blot. **D-F.** Northern blot assay for detecting circMbl, circMbl 5'ss and circPde8 in transfected *Drosophila* S2 cells. RNA was treated with RNAseR or mock treated before running the gels. Blotting was done using probes for the relevant circRNA junctions or exons (which detect both the linear and circRNA molecules). Concatamer products migrate higher due to their size and are sensitive to RNAseR treatment. **G.** RT-PCR analysis for RNA extracted from *Drosophila* S2 cells, transfected with different minigenes, -/+ RnaseR treatment. Gene expression was normalized to an endogenous circRNA. Data is presented as mean ±SD (n=3). **H,I.** Northern blot assay for detecting circCamKI in transfected *Drosophila* S2 cells. In **H.** and **I.** three circRNA products originate from the minigene due to alternative splicing of the internal introns.

Figure S2 related to Figure 2:

Polysome profile analysis of *Drosophila* S2 cells transfected with plasmids expressing the circMblV5, circCdiV5 and circCamKIV5 minigenes, a plasmid driving expression of GFP and a plasmid driving expression of a circRNA in which a split Cherry protein is under the control of the putative circMbl IRES (see Figure 1F). **A.** RT-PCR analysis of the different fractions of a polysome gradient. The results are the average of three biological replicates and are expressed as fraction of the RNA target over the gradient. *Gadph* mRNA amplified from a mouse RNA spike-in was used for normalizing between samples. Error bars indicate standard error of the mean (SEM). **B.** RT-PCR analysis of the different fractions for the indicated targets. The results are the average of three biological replicates and are expressed as fraction of the RNA target over the whole gradient. We utilized mouse *gadph* mRNA amplified from a mouse RNA spike-in for normalizing between samples across the gradient. Error bars indicate standard error of the mean (SEM). **C-F.** RT-PCR results showing the RNaseR sensitivity/resistance of the indicated RNA targets across the polysome gradients. The RNAseR sensitivity was obtained by calculating the RNAseR/Mock ratio for each target in each fraction. The data was normalized to mouse *gadph* mRNA; we added a mouse RNA spike to each fraction. To correct for the different efficiency of the RNAseR treatment across samples and fractions we used the GFP mRNA. Outliers were excluded for normalization and all the values are the average of at least 2 biological replicates.

Figure S3 related to Figure 3:

A. Top: Metagene profile of RFP read phasing analysis for no detergent samples. First (left) the overall frequency with which a read 5' end falls into a particular frame, relative to the annotated start codons. Next (middle), these frequencies subdivided by read length, represented by color in a heatmap plot. Last (right), the log-ratio scores of the position and length-dependent frequencies to a background model with uniform distribution across frames. Red indicates overrepresented, blue underrepresented frame/length combinations. Bottom: The highly expressed *Adh* gene as an example. Last panel contains the aggregate support each length-group of reads gives to the three possible frames relative to the annotated start. The correct frame1 receives very strong support. **B.** Same as A but for detergent containing lysis buffer conditions. **C.** Boxplots showing the distribution of aggregate log-ratio scores for translation (ORF-score) of known ORFs (ORF CDS). Negative controls are scores of the same ORFs when randomly permuting the scoring matrix for each ORF (ORF perm), or randomly shuffling the position of each read (ORF shuffle). Red line indicates the median and blue dot the mean. Box shows interquartile range. Whiskers show 5^{th} to 95^{th} percentile range. Outliers are omitted. Significantly different medians are indicated by stars *: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$; **** $P < 0.0001$; two-sided Mann-Whitney-U test. E. Box plots indicating the distribution of aggregate log-ratio score for start or stop codons. **D.** Meta-analysis of the RFP reads in the proximity of the start (top) and stop (bottom) codons of all annotated genes. The no detergent samples were used for these graphs. From left to right: the total frequency of RFP read 5' end positions plotted against the relative distance (in nucleotides) to the start/stop codon (Left). The same information, but subdivided into reads of different length, with frequency represented by color (Middle). Relative enrichment of RFP reads around real start/stop codons over the background frequency for reads of such length in the vicinity of start/stop codons, expressed as log-ratio of these frequencies (Right). Positive scores (red) indicate consistency with the signature of start/stop codons, negative scores (blue) with the uniform background distribution. **E.** Scatter plot showing the reproducibility of ORF scoring between biological replicates.

The ORF log-ratio scores for the coding sequence of each annotated gene are highly correlated between replicates (rho=Spearman rank correlation, R=Pearson R).

Figure S4 related to Figure 4:
A subset of circRNAs is associated with translating ribosomes. **A.** Comparison of RFP backsplicing reads between the FRP libraries prepared from with and without detergent samples. **B.** Backsplice RFP reads originate from circRNAs expressed at different levels of expression. Data from fly heads. **C.** Graph comparing the number of RFP reads on circRNA junctions with the number of observed reads in the linear junctions of the hosting mRNA. In the y-axis we plotted for each ribo-circRNA the ratio between the number of backspliced RFP reads and the highest number of flanking junction reads of the linear mRNA. **D.** Overlap of ribo-circRNAs between the RFP fly head and cell lines datasets. **E.** RT-PCR results showing RNAseR sensitivity (expressed as the ratio between the expression values in the RNAseR and mock-treated samples) for the candidate ribo-circRNAs. Values were normalized to mouse *gadph* mRNA. Equal amounts of mouse spike-in RNA were added to the sample before the RT reaction. Previously validated circRNAs (left) were used as negative (resistant) controls and rp49 mRNA as a positive (sensitive) control. **F.** Detection of the RAADTTDMFPLIM peptide by selected reaction monitoring. The uppertrace shows the elution profile for the three monitored transitions on a Q-TRAP 6500 mass spectrometer (b8-red, b9-green and b11-blue). Bottom: chromatographic profile of the native RAADTTDMFPLIM peptide from the MBL IP experiment. Concentration of the peptide is 80 pmol.

Figure S5 related to Figure 5:
Computational analysis of ribo-circRNAs. **A.** ribo-circRNAs are longer than utc-circRNAs as well as control and all circRNAs. Boxplots showing the distribution circRNA spliced lengths, assuming all introns are spliced out. Negative controls ("ctrl") are consecutive non-circRNA exons that are selected to match the spliced-length distribution of all circRNAs ("all"). "ribo"/"utc" are circRNAs with a cORF and with/without ribo-seq back-spliced read support. Red line indicates the median and blue dot the mean. Box shows interquartile range. Whiskers show 5^{th} to 95^{th} percentile range. Outliers are omitted. Significantly different medians are indicated by stars: * : $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$; **** $P < 0.0001$; two-sided Mann-Whitney-U test. **B.** Validation of the circMbl expression in *D. simulams, D. Yaacuba, D. pseudoscura, D. virilis* using Sanger sequencing of the head to tail junction. Black arrow indicates the junction. The stop codon is highlighted in a pink box. **C.** The distribution of distances between 5' splice site and the closest stop codon is well approximated by a geometric distribution. Black dots: observed stop distances for all internal (not first) *Drosophila melanogaster* 5'UTR exons. Red line: geometric distribution with same mean. Insert: zoom into the first 5 nucleotides shows frame-specific deviation from the geometric distribution due to overlap with the exonic splice site motif. **D.** Distribution of stop codon conservation scores for different sets of cORF stop codons and controls shown as cumulative relative frequency plots. prefix "rand" means randomly selected from the 5'UTR sequences also contained in the corresponding cORFs. prefix "closest" means, control stop codons, selected for minimum distance to cORF stop, ignoring frame. Suffix "uORF" indicates that stop codon also terminates a possible uORF. "Known" are stop codons from the annotated mRNA ORFs corresponding to the circRNAs. "n.s" for no significant difference of the medians by double-sided Mann-Whitney U test at P=5% cutoff. **E.** Proteins domains enriched in cORFs and present in ribo-circRNAs. We used ScanProsite to predict the domains from cORF predicted proteins. The same analysis was applied on exons selected randomly from the same hosting genes and was used as a control for filtering out randomly found domains. The most frequented domains found in cORF predicted proteins after filtering is presented. The y-axis indicates the number of those domains found in the cORF group.

Figure S6 related to Figure 5:
Analysis of cORFs in mammals. **A.** Top: Number of predicted cORFs *per* circRNA in rat circRNAs. Bottom: Number of predicted cORFs *per* circRNA in mouse circRNAs. **B, C.** Backsplice RFP reads originate from circRNAs expressed at different levels of expression in rat brain (A) and C2C12 cells (B)**. D.** Non-ambiguous circRNAs foot-printing reads from rat brain are in the same scale than the ones in junctions of the same exons. **E.** Non-ambiguous circRNAs foot-printing reads in C2C12 cells are in the same scale than the ones in junctions of the same exons. **F,G.** Rat and Mouse ribo-circRNAs are strongly enriched for 5' UTR and CDS overlap compared to control exons. **H.** Rat and Mouse ribo-circRNAs are more likely to start their cORF in the AUG of the hosting/linear mRNA than control.

Figure S7 related to Figure 6:

Ribo-circRNAs have IRES sequences. **A.** Western blot analysis for *Drosophila* S2 cells, transfected with pAcGFP and pMT-circPde8V5 with or without the 4E-BPV5 expressing plasmid. The V5 blot recognizes both the protein originated from the circPde8V5 minigene, which shows two bands (expected sizes: 14.5 and 17KD) and the 4E-BP (17 KD) that masks the higher Pde8 band. **B.** Firefly luciferase levels of the bicistronic carrying the putative circPde8 IRES in the straight or inverted orientation. The levels of S2 cells are presented as background luminescence. **C.** RT-PCR for firefly luciferase in cells transfected with the different bicistronic constructs. N=3. Error represents standard error of the mean (SEM). Data was normalized to the endogenous rp49 mRNA. **D.** Schematic representation of circMbl reporter used in cell-free translation. The UTR region of the circMBL was fused with the Renilla coding sequence and circularized. **E.** RT-PCR analysis examining circular and linear mbl transcript levels in response to starvation in circMbl OE flies. Data is presented as mean ±SD (N=2). Data were normalized to *rp49* and 28S rRNA. **F.** Levels of *foxo* mRNA, *mbl* pre and mRNA and circMbl from *Drosophila* S2 cells at different times following *foxo* expression. The measurements were performed by RT-PCR at the indicated times after *foxo* induction by addition of copper (MT-FOXO stably transfected cell line). N=3. Error represents standard error of the mean (SEM).

Legends for Tables S1-S7

Table S1 related to Figure 1. List of cORFs in *Drosophila melanogaster*.

Table S2 related to Figure 3. List of *Drosophila* ribo-circRNAs.

Table S3 related to Figure 5. Evolutionary Conservation scores for *Drosophila* cORFs and ribo-circRNAs.

Table S4 related to Figure 5. Proteins Domains found on ribo-circRNAs.

Table S5 related to Figure 5. Summary of back-spliced reads detected in C2C12 ribosomal foot printing data (PRJEB7207).

Table S6 related to Figure 5. Summary of back-spliced reads detected in rat samples ribosomal foot printing data (GSE66715).

Table S7 related to Figures 1-7. List of oligonucleotides utilized in this study.