**Supplementary Material for**

# Contributions of mRNA abundance, ribosome loading, and post- or peri-translational effects to temporal repression of *C. elegans* heterochronic miRNA targets

Michael Stadler , Karen Artiles, Julia Pak, and Andrew Fire

Contents:

Supplementary Methods

Supplementary Tables

Supplementary Figure Legends

Supplementary Figures

References

SUPPLEMENTARY METHODS

Methods presented in the main text are condensed for space considerations. The full methods are reported below for completeness.

**Samples.**

*C. elegans* N2 animals were cultured as described in Brenner (Brenner 1974). Animals were allowed to hatch in the absence of food for 24 hours 16°C, producing synchronized populations of L1 stage larvae. L1s were transferred to enriched agar plates with *E. coli* OP50 food source at 16°C for: 4 hours (early L1), 34 hours (L2), 45 hours (L3), or 63 hours (L4). Animals were harvested by washing and flash-freezing aliquots in liquid nitrogen.

**Ribosome footprinting and library preparation.**

Footprinting and library preparation was performed as described in Ingolia et al. (Ingolia et al. 2009), with modifications from Stadler and Fire (Stadler and Fire 2011).

**Sequence Analysis: ribosome footprinting.**

Ultra-high throughput sequencing for both mRNAseq and ribosome profiling experiments were performed on the Illumina platform. Reads were trimmed of all contiguous trailing A residues (introduced during library preparation by poly(A) polymerase) and resulting reads longer than 20 nt were used for mapping. Reference sequences for mapping RPF and mRNAseq reads consisted of the total set of predicted/verified coding sequences from the Wormbase WS220 release with 18 nt of 5' and 3' genomic flanking sequence. For transcripts with multiple isoforms, a single isoform was selected arbitrarily. Reads were mapped to these reference sets using Bowtie v.0.12.2 (Langmead et al. 2009) reporting all reads that were perfect matches or contained a single mismatch to the reference.

Library sizes for all mRNA-seq and ribosome footprint libraries were normalized using the EdgeR (Robinson and Oshlack 2010) software package within R using the TMM normalization mode. Using normalized library sizes, the reads per kilobase of exon per million mapped reads (Mortazavi et al. 2008) was determined for each gene in each library. Ribosome loading was computed separately for each ribosome footprint library as the $\log_2$ ratio of normalized tag counts in that library and the average normalized tag count from all mRNA-seq libraries derived from the same developmental stage. Figures representing alternative normalization procedures are available in supplemental materials.

Coincidence analysis was used to determine how similar two ribosome occupancy profiles were. The coincidence statistic describes the probability of choosing a random sequencing read from each of two libraries that fall on the same position. More similar profiles will show a greater probability of coincidence. Coincidence was calculated as:

$$\frac{\sum_{allpositions\ p} \text{Number of reads at } p \text{ library1} \times \text{Number of reads at } p \text{ in library 2}}{\text{Total reads mapping to transcript in library 1} \times \text{Total reads mapping to transcript in library 2}}$$

For generating ribosome occupancy profiles for individual transcripts, read coverage was assigned based on the location of the ribosomal P-site (12 nt from the 5' end of the footprint (Stadler and Fire 2011). Occupancy at each position on a transcript is the number of reads in the library whose P-site codon maps to that position. Density plots are smoothed using kernel density estimation. Smoothing is performed to allow better visualization of ribosome density profiles, especially when examining mRNAs with sparser coverage. Un-smoothed density plots are shown in Figs. S5.

Positional difference scores were calculated using a sliding window approach. A single occupancy vector was generated for each gene at each stage by combining all biological replicates from a given stage and normalizing to the total coverage for the gene within that stage. A difference score for each window was calculated as ($occupancy_{stage1}$ – $occupancy_{stage2}$) / ($occupancy_{stage1}$ + $occupancy_{stage2}$), where $occupancy_{stage\ n}$ represents the total occupancy (normalized read count) within the window in that stage. Using windows of 20 codons and steps of 10 codons, a distribution of difference scores was thus generated for each gene in a comparison between stages. The mean and median of these distributions were examined for miRNA targets and compared to transcripts of similar coverage levels (with similar coverage defined as ± 0.40 reads per codon).

## lin-14 gf heterozygote allele-specific qRT-PCR.

*lin-14 n536n539* (Ambros and Horvitz 1987) hermaphrodites were crossed with males harboring the integrated transgene *hlh-8::GFP* on the X chromosome (Harfe et al. 1998), and GFP was used to mark hermaphrodite cross-progeny. Animals were hatched in the presence of food, and populations of five L1 (<3 hours post hatch) or five L4 animals were frozen in GITC buffer, followed by nucleic acid preparation as described in the single worm RT-PCR protocol in (Epstein and Shakes 1995).

Reverse transcription of a *lin-14* fragment surrounding the *n539* point mutation was carried out using the primer AF-MS-51 (ATTTGCGCATTGCCTCGCGG). PCR was carried out with primers AF-MS-55 (AAGCGTGTCTTTGGACCACG) and AF-MS-56 (ATTTGTCCCAAAAGTCTTCC) and Phusion polymerase (NEB, Ipswich, MA). Primers span a 45 bp intron, allowing clear separation on an agarose gel of RNA-derived product (165 bp) from products amplified from genomic DNA (210 bp). Because amplifications were carried out using relatively small amounts of starting cDNA, we sought to control for PCR jackpot effects by first carrying out 12 independent amplifications, combining the products in equal volumes, and performing several additional cycles of amplification.

Sequencing using Sanger sequencing chemistry was carried out by Elim Biopharmaceuticals (Hayward, Ca) using AF-MS-56 as the sequencing primer. To allow sequencing by the Illumina GAIIx system, several additional rounds of PCR were carried out using primers designed to add the necessary sequences for Illumina sequencing and four-nt barcodes.

SUPPLEMENTARY TABLES

Tables are provided in a separate Excel document.

Table S1. mRNA abundance measured by mRNA-seq for heterochronic miRNA targets in four larval stages. TMM normalization is used, numbers shown are the mean of all replicates at each stage.

Table S2A. Peak heights measured from Sanger traces of *lin-14* heterozygote cDNA amplifications. Each sample is a separate biological replicate. Three controls are shown which represent mixing of DNA from WT and mutant animals in various proportions to determine consistent amplification of alleles.

Table S2B. Equivalent data to S2A except with Illumina sequencing of the same libraries from Table S2A.

Table S3. Total RPF abundance for heterochronic miRNA targets in four larval stages. TMM normalization is used, numbers shown are the mean of all replicates at each stage.

Table S4. Ribosome loading for heterochronic miRNA targets in four larval stages. TMM normalization is used, numbers shown are the mean of all replicates at each stage. Note that these numbers represent the mean of ratios (each footprinting sample provides a ratio), and not the ratio of means.

Table S5. Coincidence analysis of heterochronic miRNA targets. Individual experiments (replicates) are compared pairwise. Experiments are labeled as larval stage_replicate number. Coincidence was measured for the indicated genes between ribosome profiles for all pairs of experiments. Average similarity among L1 replicates, among L4 replicates, and between L1 (miRNA regulation not occurring) and L4 (miRNA regulation occurring for all targets) are summarized on the right.

## SUPPLEMENTARY FIGURE LEGENDS

Figure S1. **Immunoblotting of LIN-14 and LIN-28.** (A) Four *C. elegans* larval stages were probed with LIN-14 antiserum, along with the *lin-14(n536n540)* strain which contains a substantial deletion that deletes a substantial fraction of the LIN-14 ORF. The blue arrow indicates the LIN-14 band; the red arrow indicates a likely background band that may be a form of LIN-14, the green arrow represents a background band used for quantitation. (B) Blotting of four larval stages with LIN-28 antiserum. The blue arrows represent two distinct LIN-28 bands, green arrows indicate background bands used for normalization. (C) For both LIN-14 and LIN-28 blots, L1 and L4 samples were mixed together in various ratios to confirm linearity of detection. Quantitation was performed using either the indicated background bands, tubulin, or protein concentration (no additional normalization), and all strategies resulted in highly similar measurements.

Figure S2. **Alternative normalization strategies for sequencing data.** Equivalent plots to Fig. 2 and Fig. 3 are shown for miRNA target genes. 'Quantile' (microarray-like) is a

normalization strategy implemented by the EdgeR software package (Robinson and Oshlack 2010). Library size simply uses the total reads mapping to mRNA as a normalization factor, with no additional account taken of library composition.

Figure S3. **Changes in mRNA levels of miRNA targets measured by Sanger sequencing.** (A) Sanger Traces for *lin-14* heterozygote RNA measurements. Top**:** Sanger sequencing traces of amplified cDNA from *lin-14* heterozygous animals at L1 (left side) and L4 (right side). The nt position of the distinguishing *n539* point mutation is highlighted, with A representing mutant transcript and G representing wild-type. Data shown for three biological replicates of five animals each. Bottom: Sanger traces for experiments in which DNA from wild-type or *lin-14(n536n539)* homozygous animals was mixed at defined ratios of 2:1, 1:1, and 1:2. Traces rendered by the 4peaks software (Mekentosj, Aalsmeer, The Netherlands). (B) Identical to Figure 2, shows equivalent data determined by measuring peak heights from Sanger sequencing traces. p-value is from two-sided T-test.

Figure S4. **Developmental changes of *lin-29* mRNA abundance and ribosome association.** mRNA abundance measured by mRNA seq, ribosome loading, and total RPF abundance is shown for *lin-29*, a transcription factor whose mRNA is targeted by LIN-41 for negative post-transcriptional regulation.

Figure S5. **Sliding window analysis of ribosome occupancy profiles.** Distributions of sliding window positional difference scores were calculated for heterochronic miRNA

genes, comparing miRNA-regulated to unregulated stages (see METHODS). Histograms are shown for the mean, median, and maximum values of the distributions for *C. elegans* mRNAs of similar coverage to the mRNA shown (defined as +- 0.40 reads/codon). The dotted orange line shows the position of the value for the mRNA of interest, and the percentile rank of the mRNA is shown below the X-axis label for each metric (a higher percentile represents a greater difference compared to the background population.

Figure S6. **Adenylation status of miRNA targets**. Abundance measured by RNA-seq of mRNAs in the poly(A)-selected pool (vertical axis) vs. abundance in total RNA (horizontal axis) for (A) L1 stage larvae, in which *lin-4, let-7*, and *miR-48/miR-84/miR-241* are at very low levels, and (B) L4 stage larvae in which each of these miRNAs is strongly expressed. Abundance is represented as reads per kilobase per million mapped reads (RPKM, Mortazavi et al. 2008). Replication-dependent histones, which lack poly(A) tails, are highlighted in yellow. (C) Changes between L1 and L4 in the adenylation state of miRNAs, measured as the ratio of tag counts in poly(A)+ RNA vs. total RNA. Dotted line in A and B reflects unity (equivalent representation in total and polyA-selected libraries).

Figure S7. **Unsmoothed ribosome occupancy profiles for heterochronic miRNA targets.** Unsmoothed ribosome occupancy profiles are generated by assigning a single count for each sequenced read to the codon corresponding to the ribosomal P-site, measured from the 5' end of the read. Data shown is un-normalized; the Y-axis represents raw read counts. Data is shown for two biological replicates for each

heterochronic miRNA target at a miRNA-unregulated stage (top, green) and a miRNA-regulated stage (bottom, red).

Figure S8. **Correlation of replicate experiments.** Correlation between RPKM values (using TMM normalization from EdgeR) for mRNA-seq and ribosome footprinting experiments from two biological replicates at each developmental time point measured. Data also shown for ribosome loading, a derived from both footpring and mRNA-seq data. Spearman's rank coefficient is shown for each plot.

Figure S9. **Examples of ribosome loading changes during development for assorted gene sets.** Ribosome loading is shown across four larval stages for gene sets drawn from the glycolysis pathway, ribosomal proteins, and RNAi-related genes.

## REFERENCES


Ambros V, and Horvitz H. 1987. The lin-14 locus of Caenorhabditis elegans controls the time of expression of specific postembryonic developmental events. *Genes Dev* **1**: 398–414.

Brenner S. 1974. The genetics of Caenorhabditis elegans. *Genetics* **77**: 71–94.

Epstein HF, and Shakes DC. 1995. *Caenorhabditis elegans: modern biological analysis of an organism*. Academic Press, Salt Lake City, UT.

Harfe B, Vaz Gomes A, Kenyon C, Liu J, Krause M, and Fire A. 1998. Analysis of a Caenorhabditis elegans Twist homolog identifies conserved and divergent aspects of mesodermal patterning. *Genes Dev* **12**: 2623–35.

Ingolia N, Ghaemmaghami S, Newman J, and Weissman J. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**: 218–23.

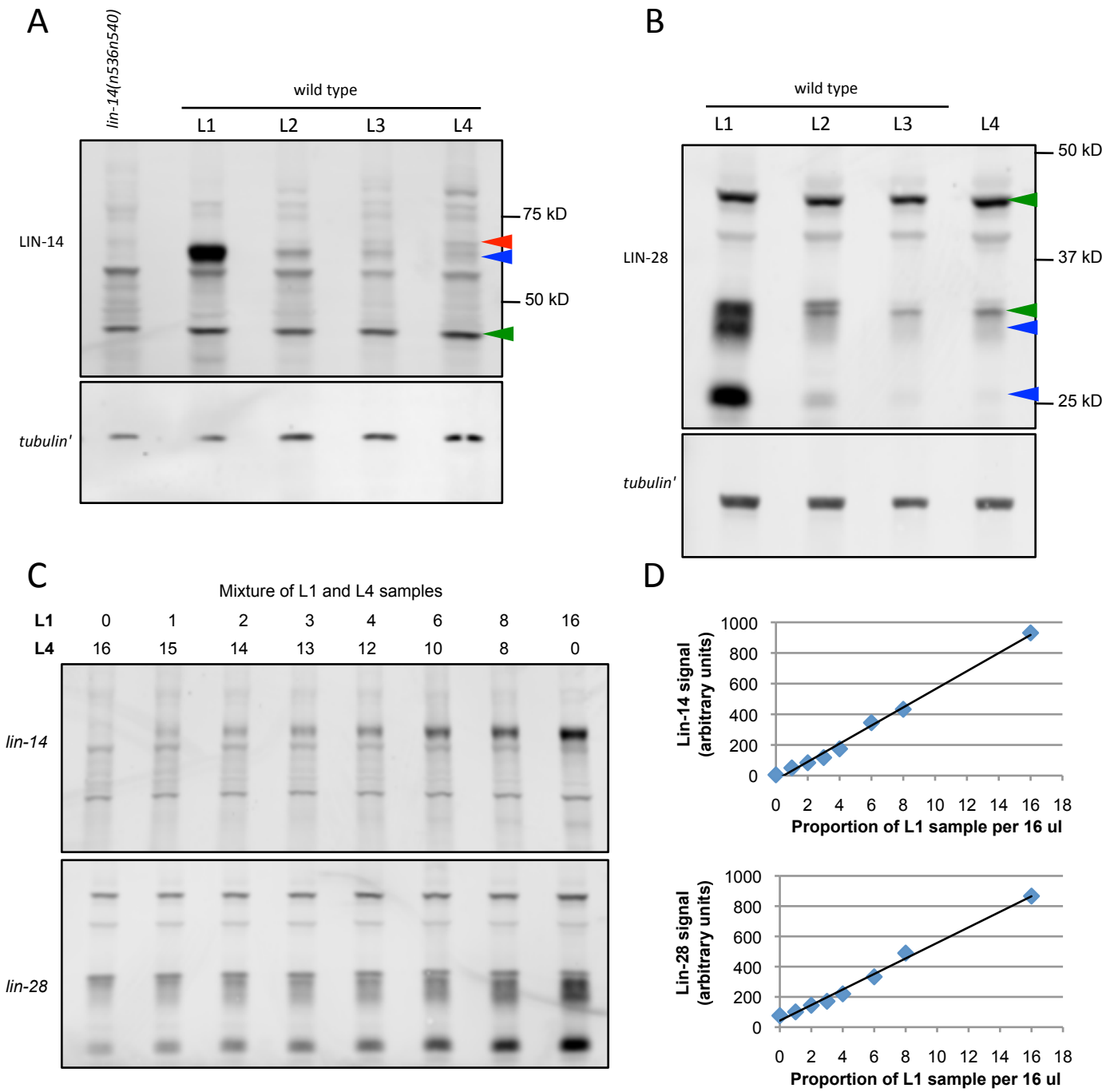
Langmead B, Trapnell C, Pop M, and Salzberg S. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Mortazavi A, Williams B, McCue K, Schaeffer L, and Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**: 621–8.

Robinson M, and Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**: R25.

Stadler M, and Fire A. 2011. Wobble base-pairing slows in vivo translation elongation in metazoans. *RNA* **17**: 2063–73.

# Figure S1


A
lin-14(n536n540)
wild type
L1
L2
L3
L4
LIN-14
75 kD
50 kD
tubulin'
B
wild type
L1
L2
L3
L4
LIN-28
50 kD
37 kD
25 kD
tubulin'
C
Mixture of L1 and L4 samples
L1 0 1 2 3 4 6 8 16
L4 16 15 14 13 12 10 8 0
lin-14
lin-28
D
Lin-14 signal (arbitrary units)
1000
800
600
400
200
0
0 2 4 6 8 10 12 14 16 18
Proportion of L1 sample per 16 ul
Lin-28 signal (arbitrary units)
1000
800
600
400
200
0
0 2 4 6 8 10 12 14 16 18
Proportion of L1 sample per 16 ul

lin-4 targets

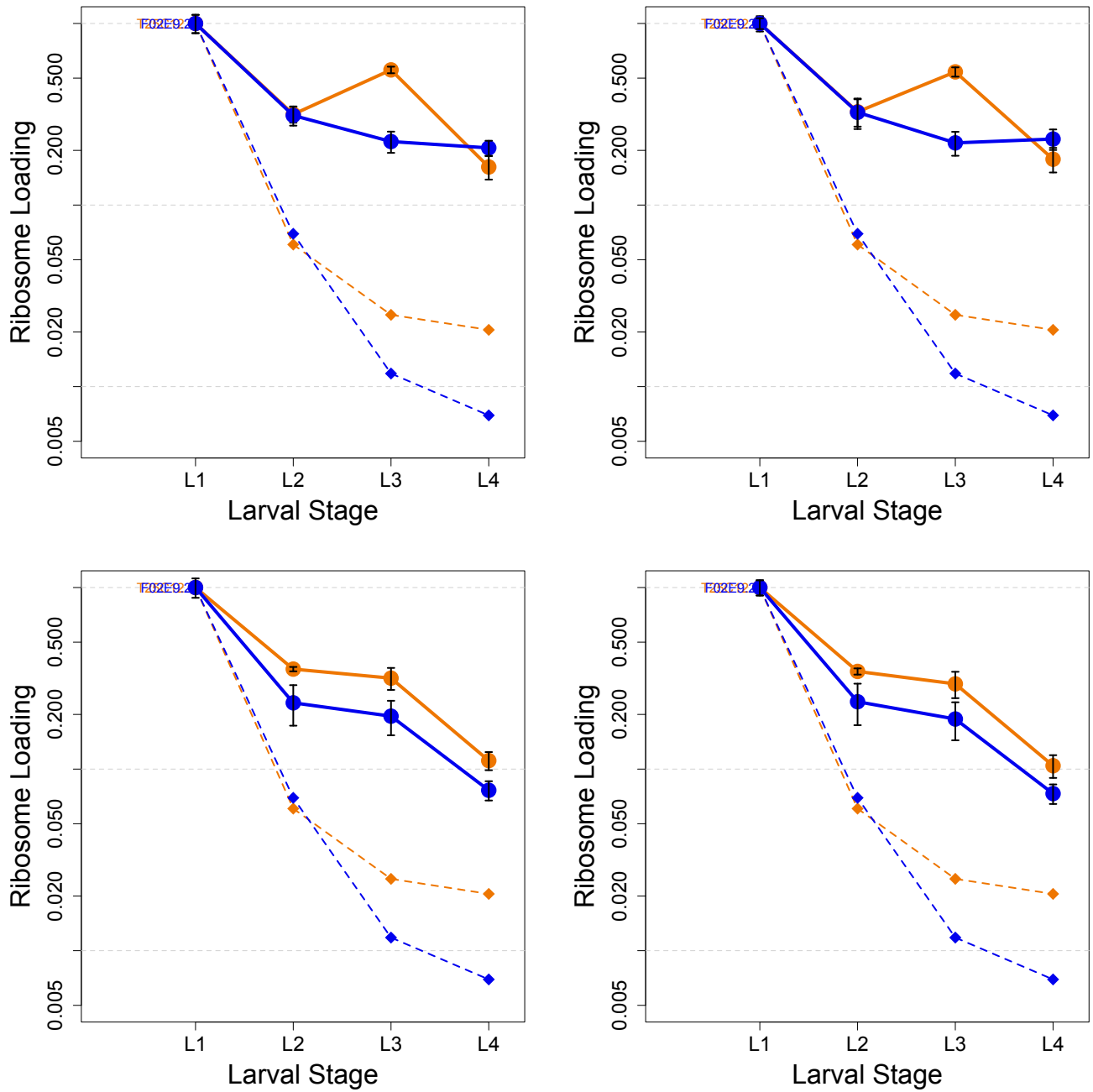
Figure S2

quantile

library size

mRNA-seq

RPFs



let-7 targets

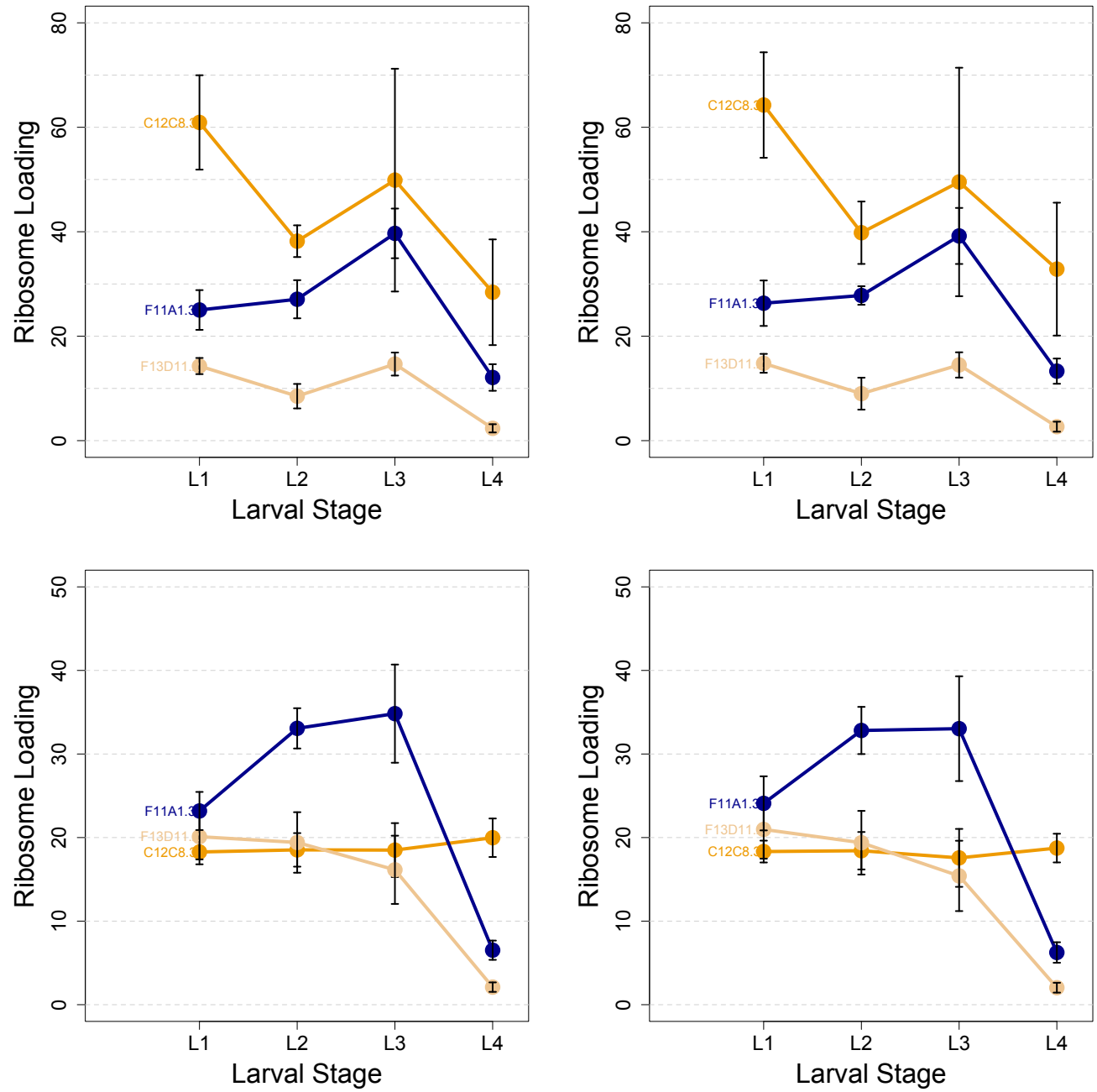
Figure S2

quantile

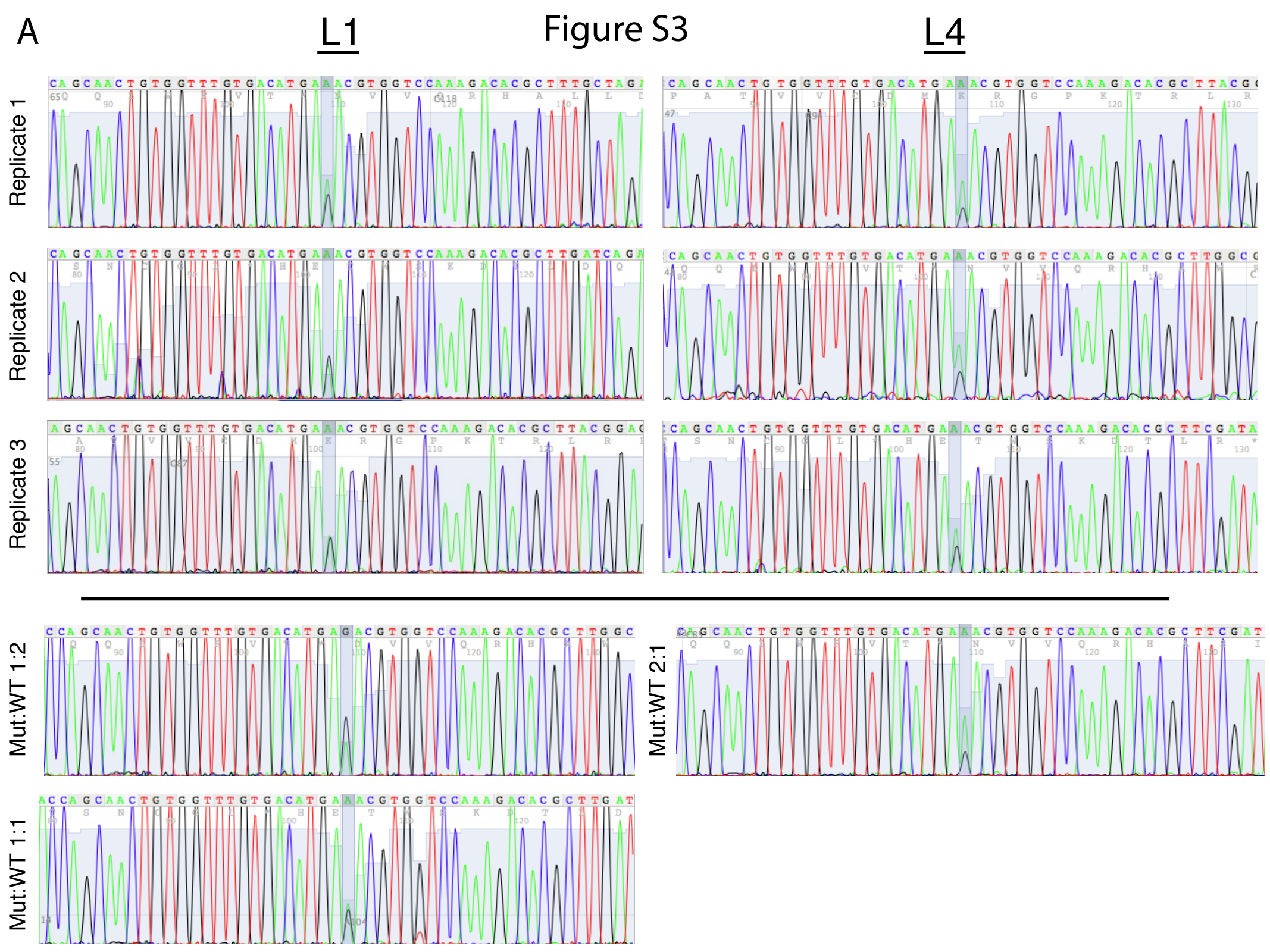
library size

mRNA-seq

RPFs

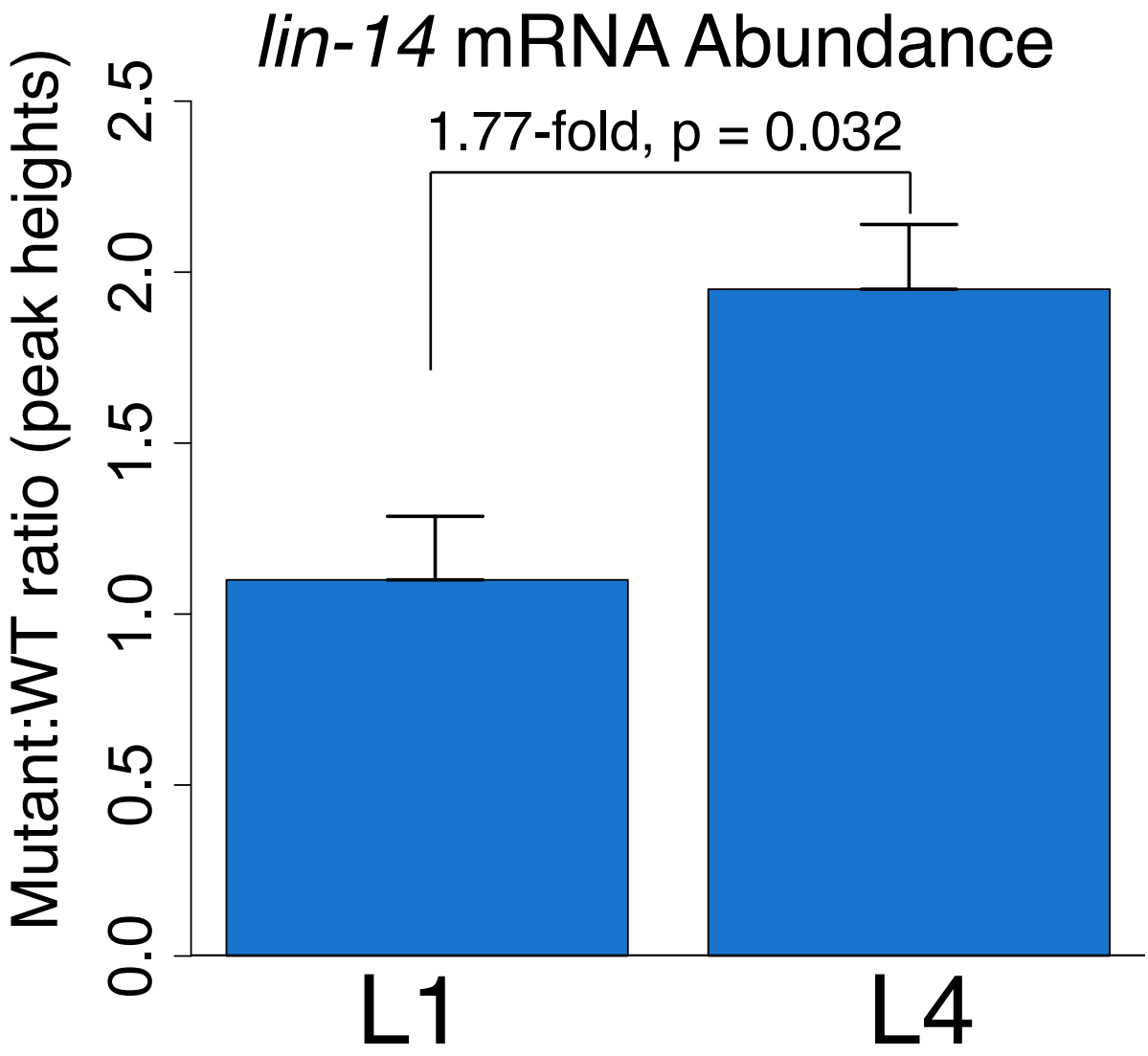


# Figure S3


A
L1
L4
Replicate 1
Replicate 2
Replicate 3
Mut:WT 1:2
Mut:WT 2:1
Mut:WT 1:1

B

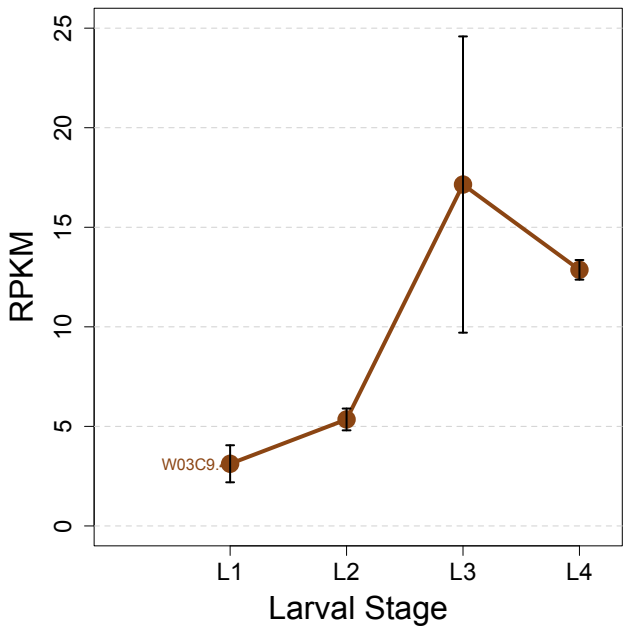
# Figure S3


*lin-14* mRNA Abundance
1.77-fold, p = 0.032
Mutant:WT ratio (peak heights)
0.0
0.5
1.0
1.5
2.0
2.5
L1
L4

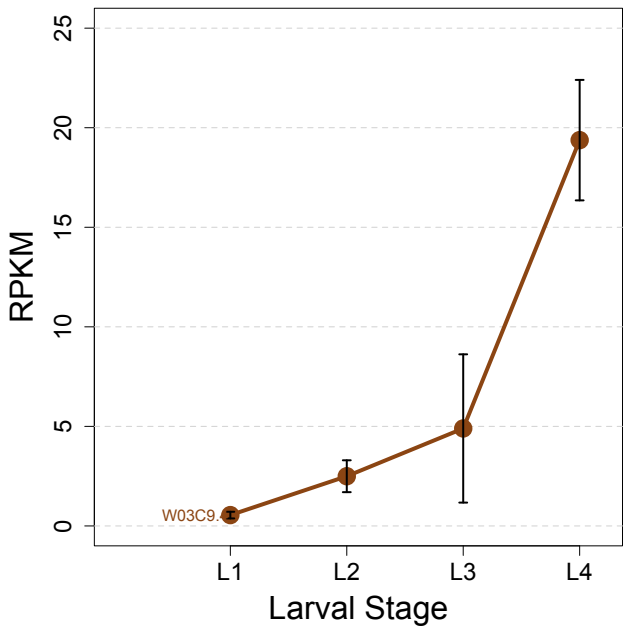
# Figure S4

## LIN-29 regulation during larval development

### mRNA-seq




### RPFs




### Ribosome loading

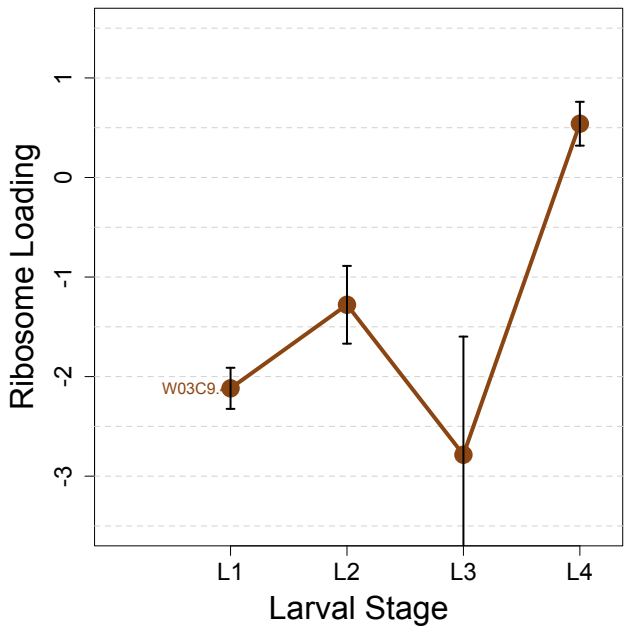


Figure S5

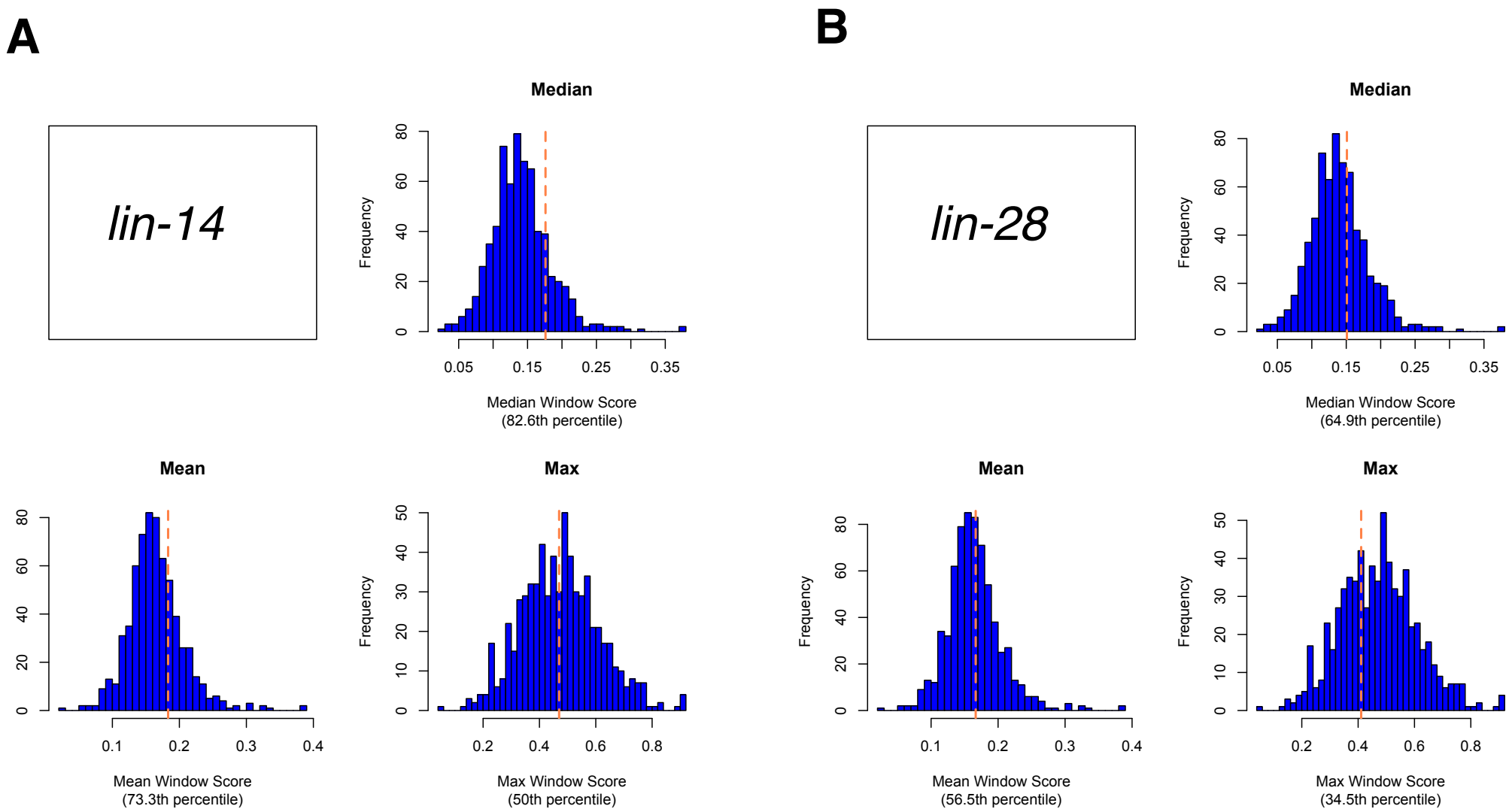
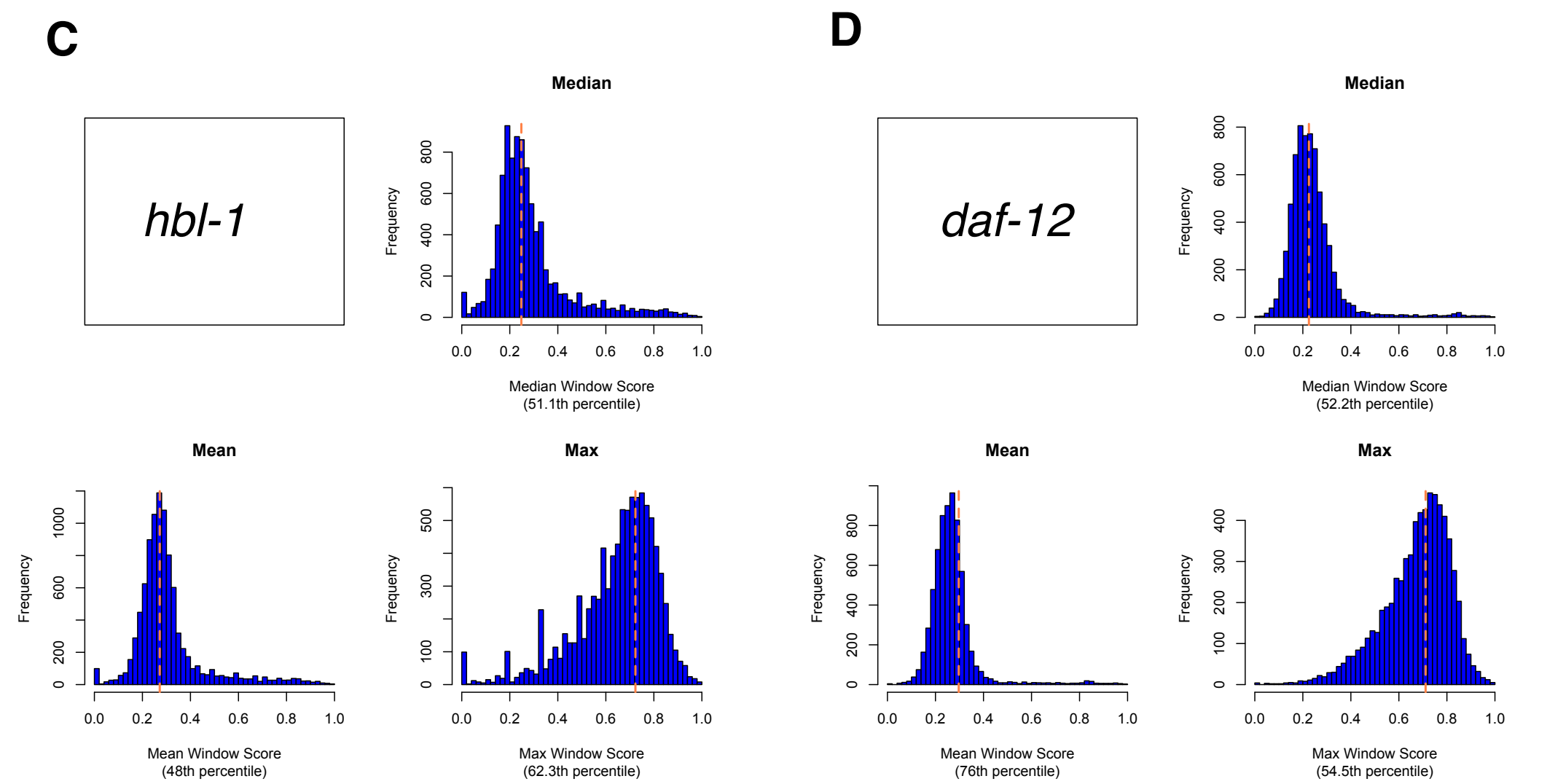
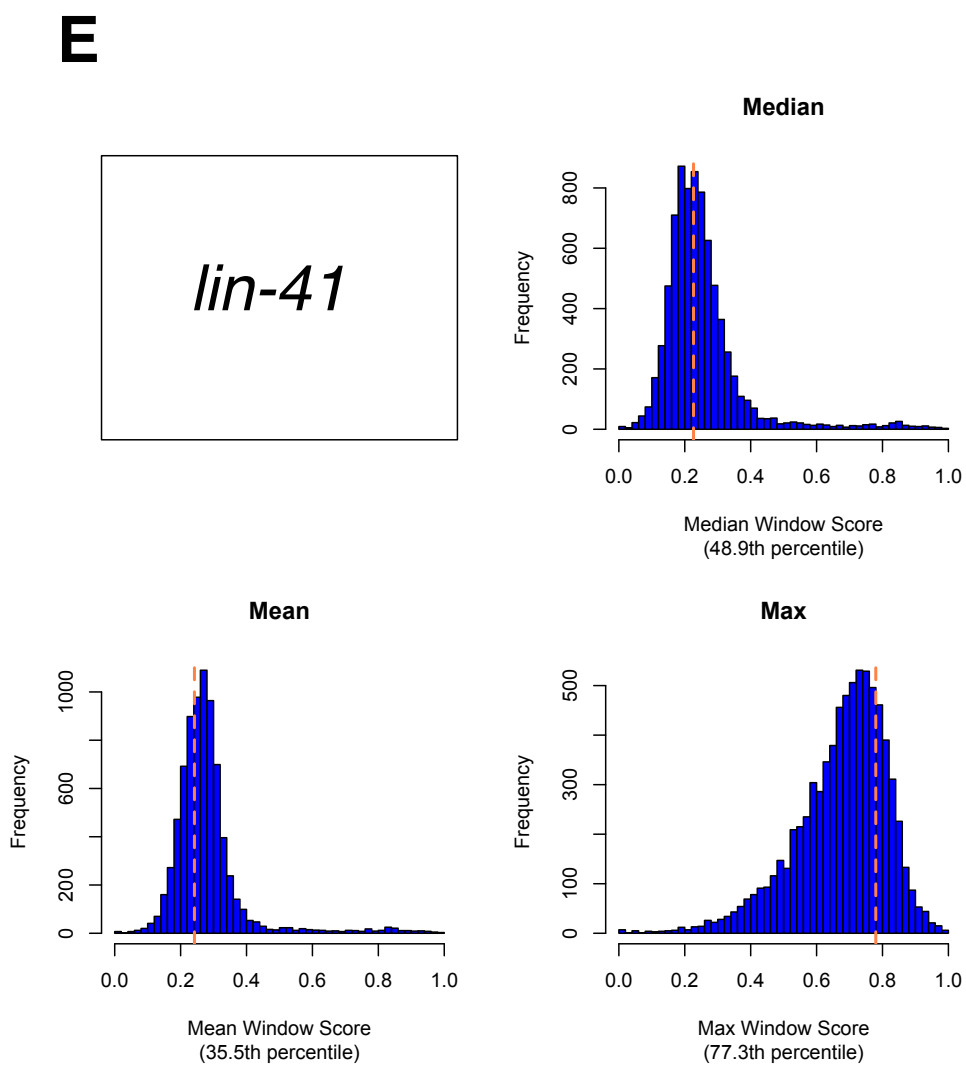

A
lin-14
Median
Frequency
Median Window Score
(82.6th percentile)
Mean
Mean Window Score
(73.3th percentile)
Max
Max Window Score
(50th percentile)
B
lin-28
Median
Median Window Score
(64.9th percentile)
Mean
Mean Window Score
(56.5th percentile)
Max
Max Window Score
(34.5th percentile)

Figure S5

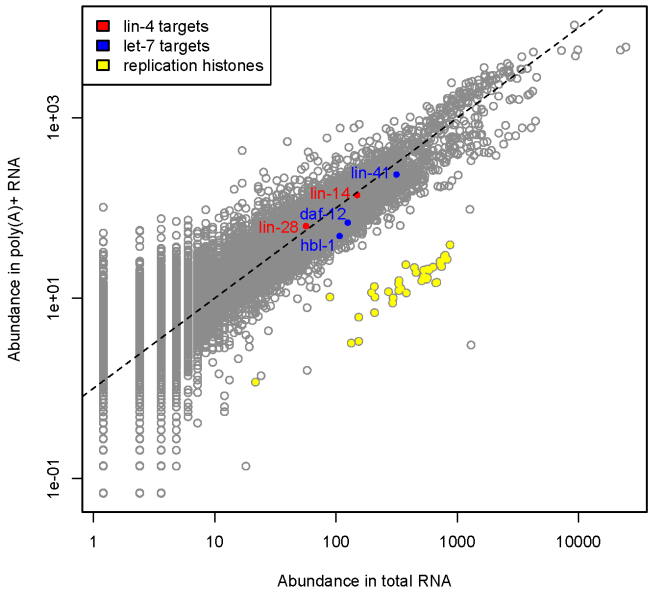

C
hbl-1
Median
Frequency
Median Window Score
(51.1th percentile)
Mean
Frequency
Mean Window Score
(48th percentile)
Max
Frequency
Max Window Score
(62.3th percentile)
D
daf-12
Median
Frequency
Median Window Score
(52.2th percentile)
Mean
Frequency
Mean Window Score
(76th percentile)
Max
Frequency
Max Window Score
(54.5th percentile)

# Figure S5

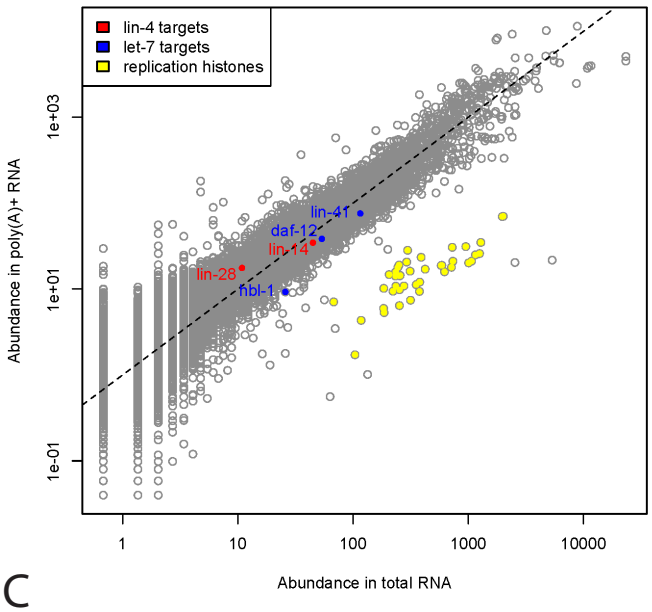

E
lin-41
Median
Frequency
0
200
400
600
800
0.0
0.2
0.4
0.6
0.8
1.0
Median Window Score
(48.9th percentile)
Mean
Frequency
0
200
600
1000
Mean Window Score
(35.5th percentile)
Max
Frequency
0
100
300
500
Max Window Score
(77.3th percentile)

# Figure S6

A




B




C

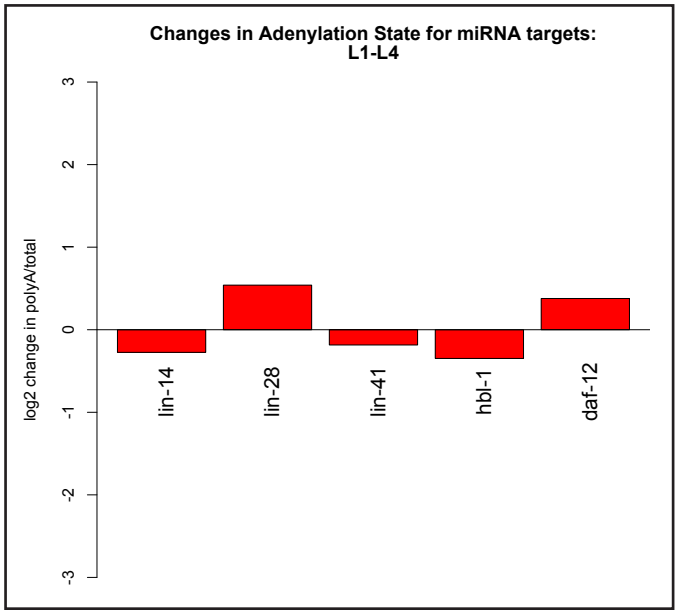
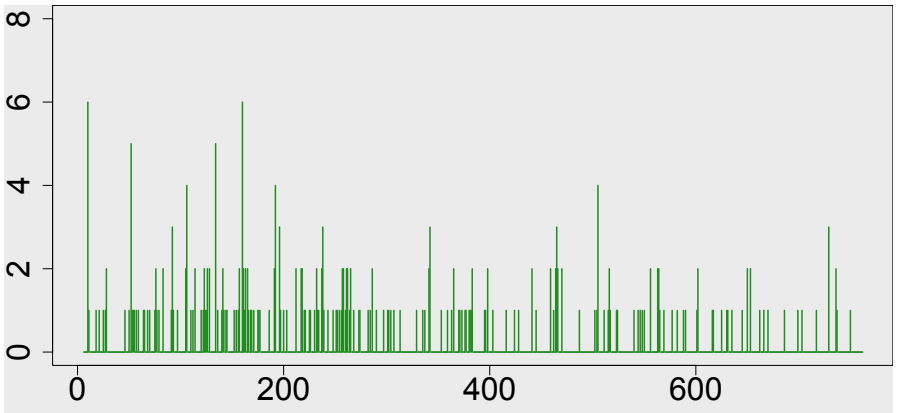
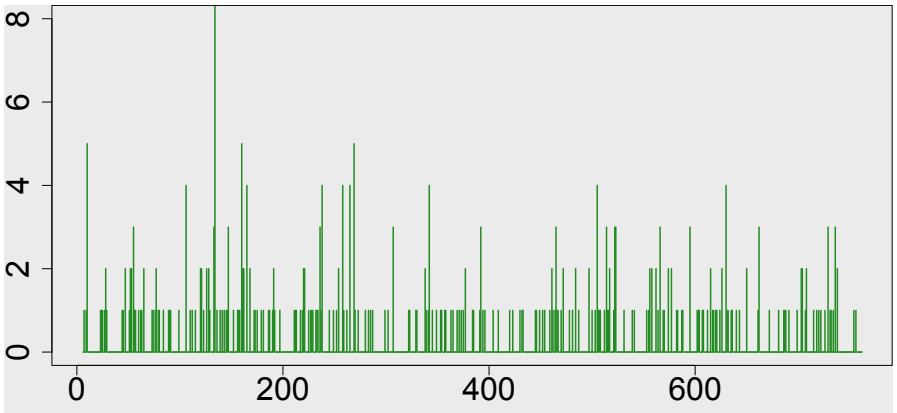


Figure S7

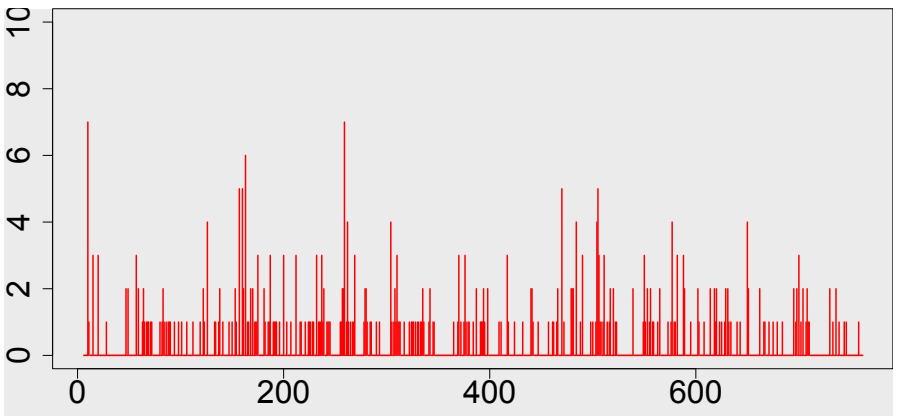
*daf-12*

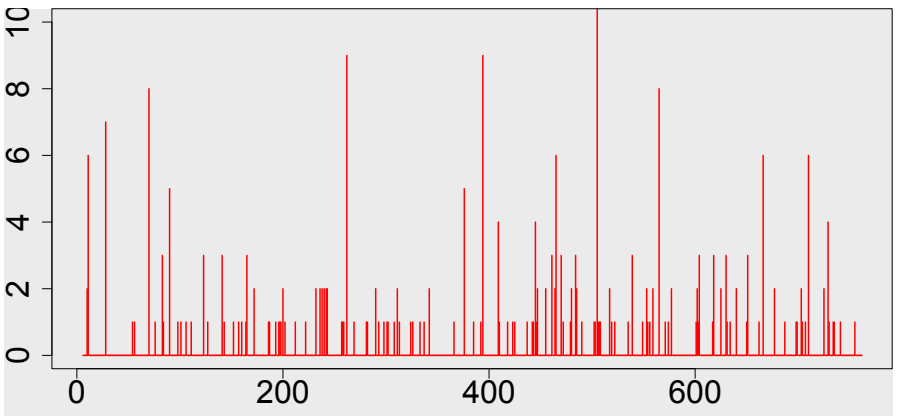
L2



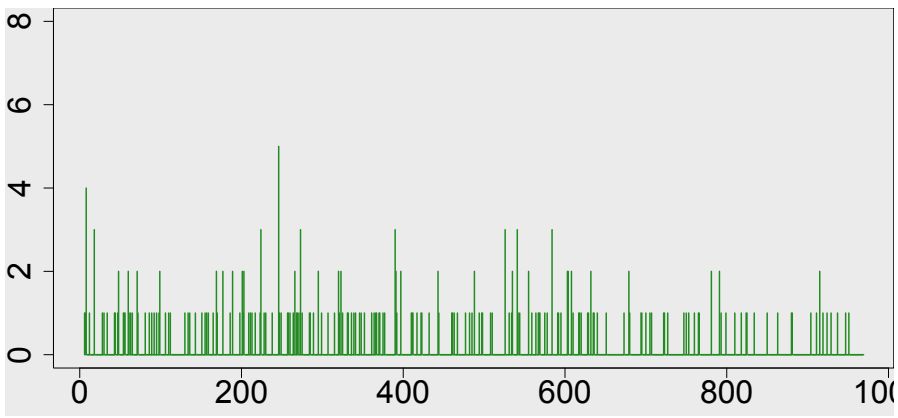

L3

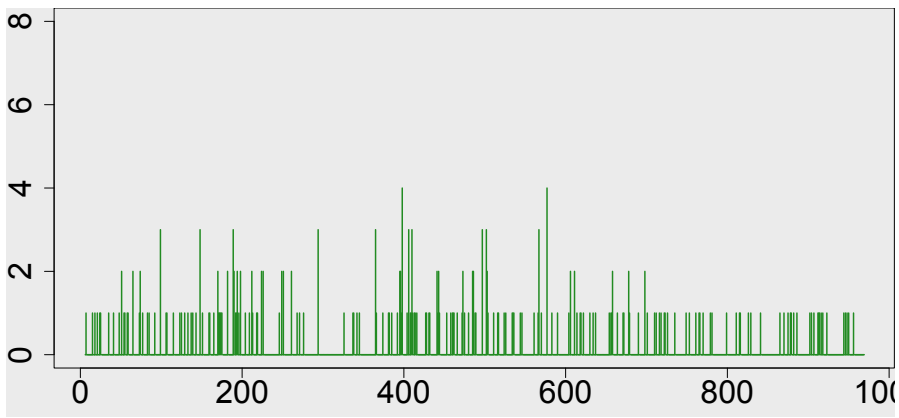


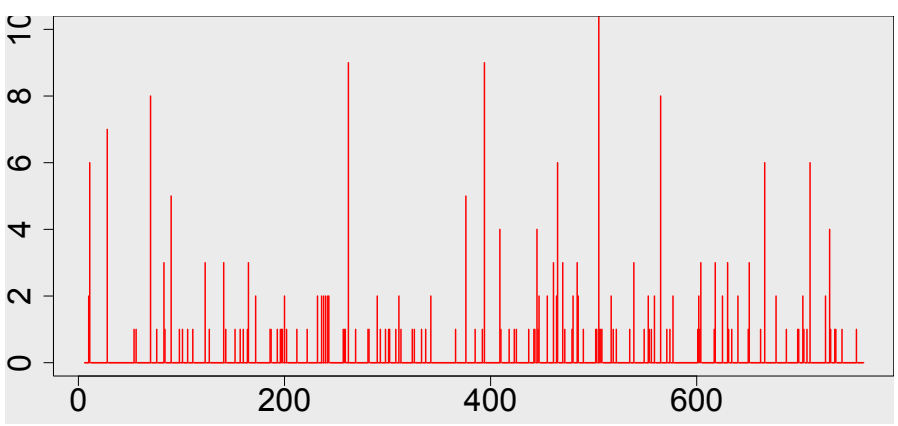
# *hbl-1*

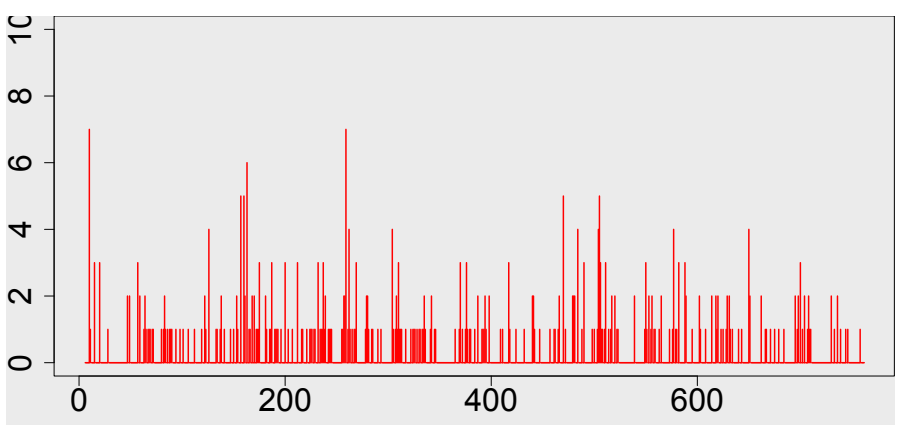
L2


0
2
4
6
8
0
200
400
600
800


0
2
4
6
8
0
200
400
600
800

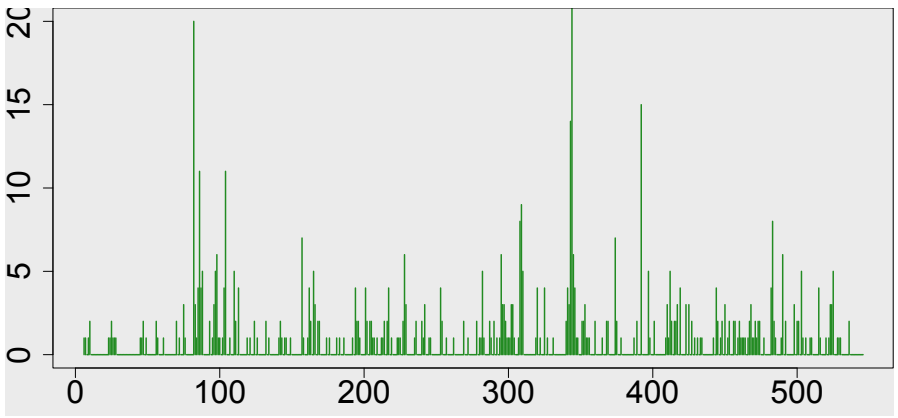
L3

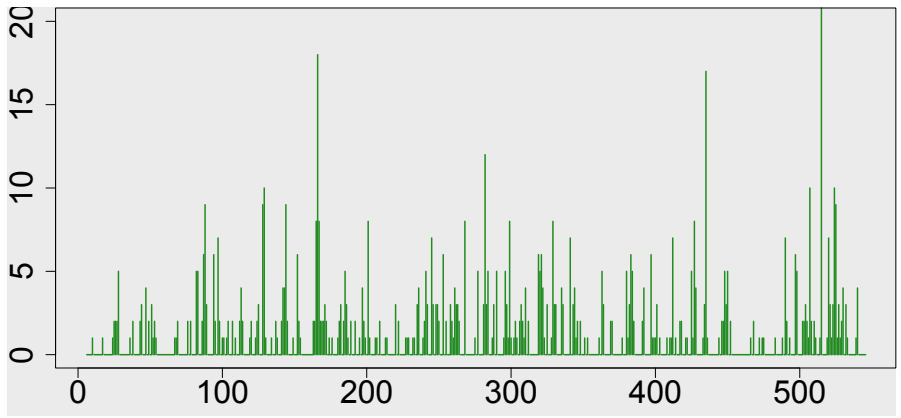

0
2
4
6
8
0
200
400
600


0
2
4
6
8
0
200
400
600

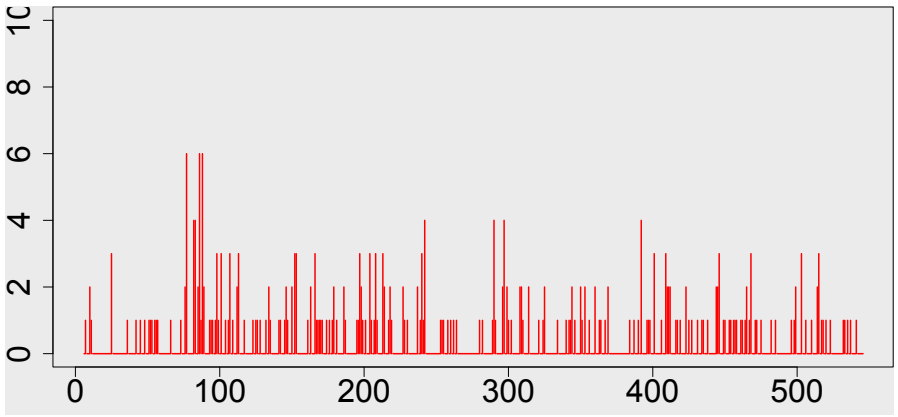
# *lin-14*

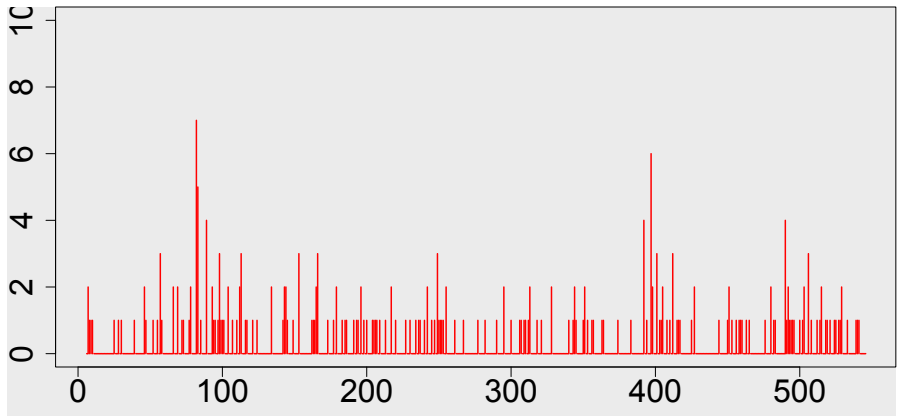
L1


0
5
10
15
20
0
100
200
300
400
500


0
5
10
15
20
0
100
200
300
400
500

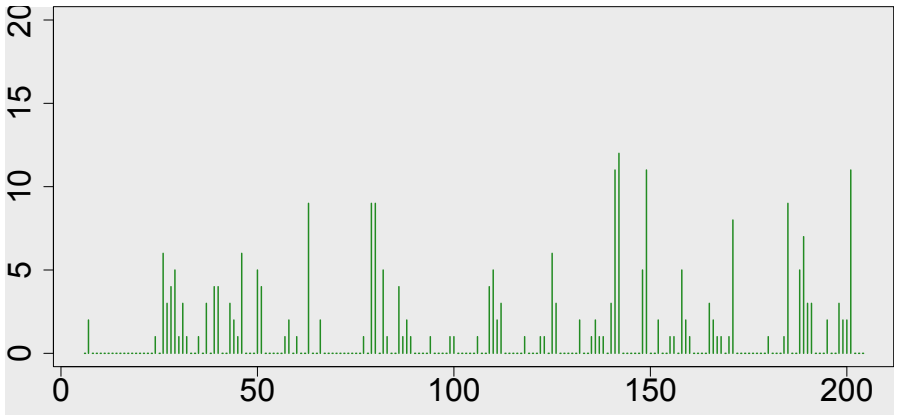
L2

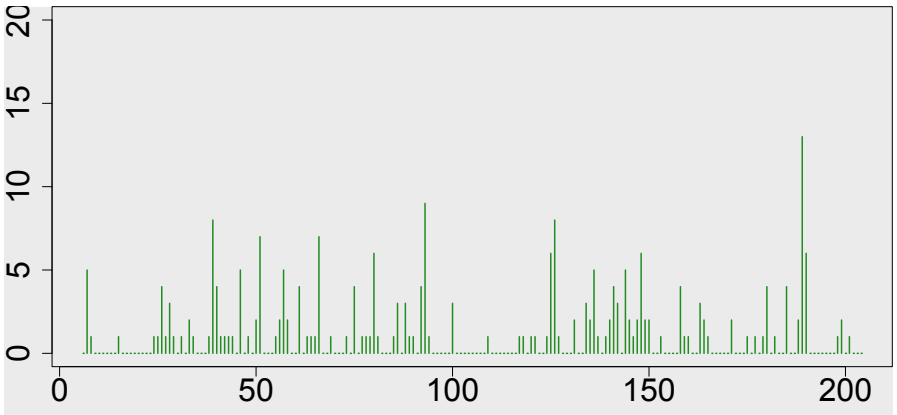

0
2
4
6
8
10
0
100
200
300
400
500


0
2
4
6
8
10
0
100
200
300
400
500

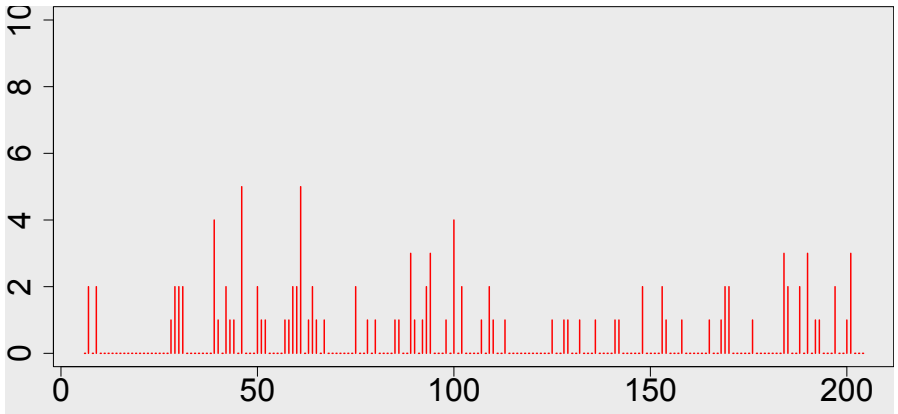
# *lin-28*

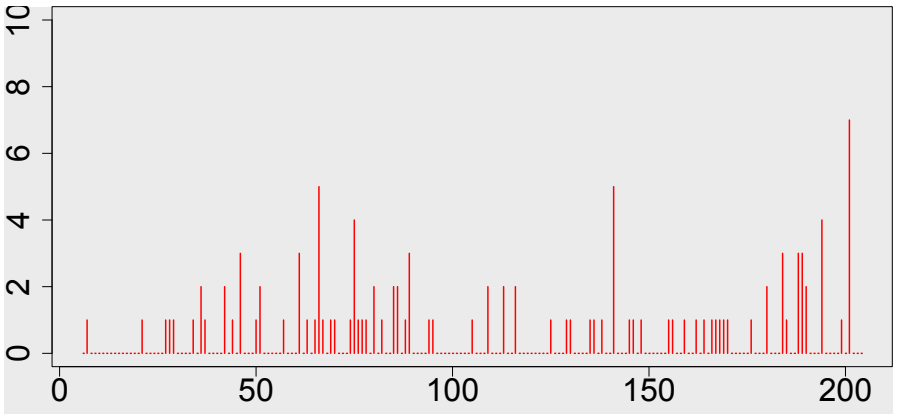
L1


0
5
10
15
20
0
50
100
150
200


0
5
10
15
20
0
50
100
150
200

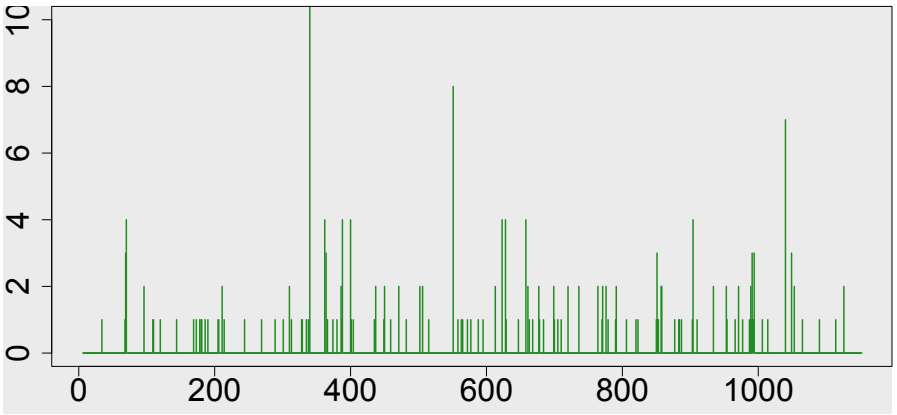
L2

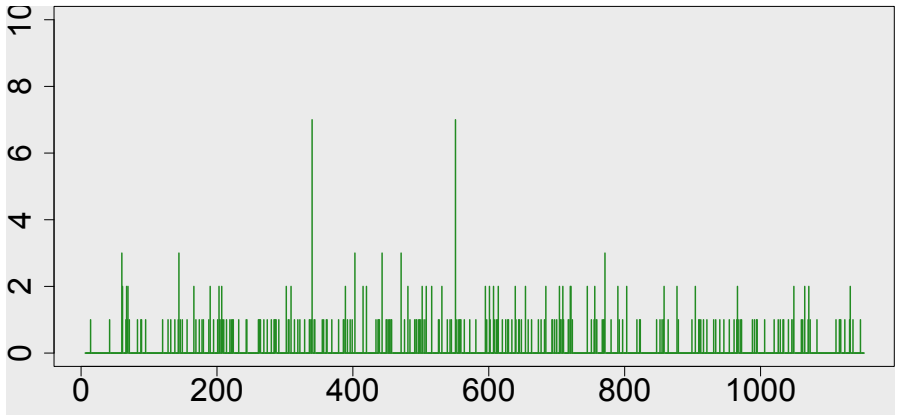

0
2
4
6
8
10
0
50
100
150
200


0
2
4
6
8
10
0
50
100
150
200

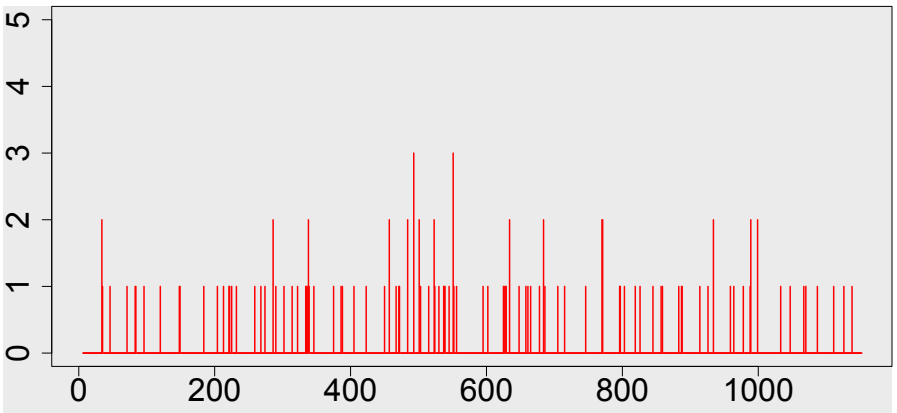
# *lin-41*

L3


0
2
4
6
8
10
0
200
400
600
800
1000


0
2
4
6
8
10
0
200
400
600
800
1000

L4


0
1
2
3
4
5
0
200
400
600
800
1000

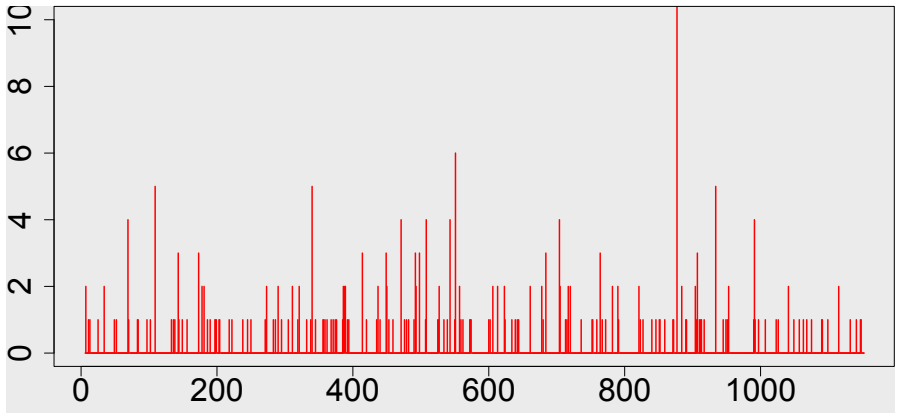

0
2
4
6
8
10
0
200
400
600
800
1000

Figure S8

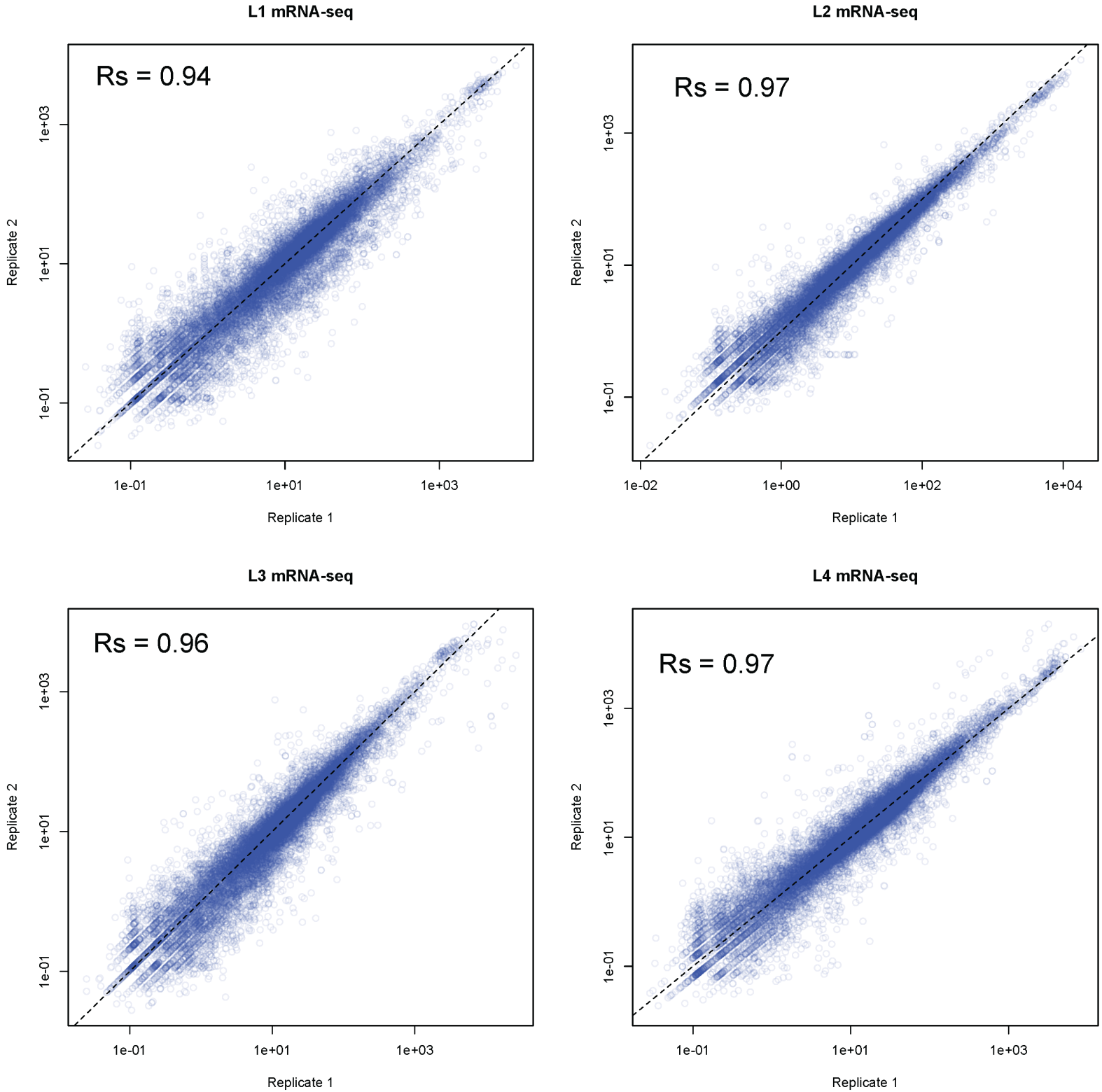

L1 mRNA-seq
Rs = 0.94
Replicate 2
Replicate 1
1e-01
1e+01
1e+03
L2 mRNA-seq
Rs = 0.97
Replicate 2
Replicate 1
1e-02
1e+00
1e+02
1e+04
L3 mRNA-seq
Rs = 0.96
Replicate 2
Replicate 1
L4 mRNA-seq
Rs = 0.97
Replicate 2
Replicate 1

Figure S8

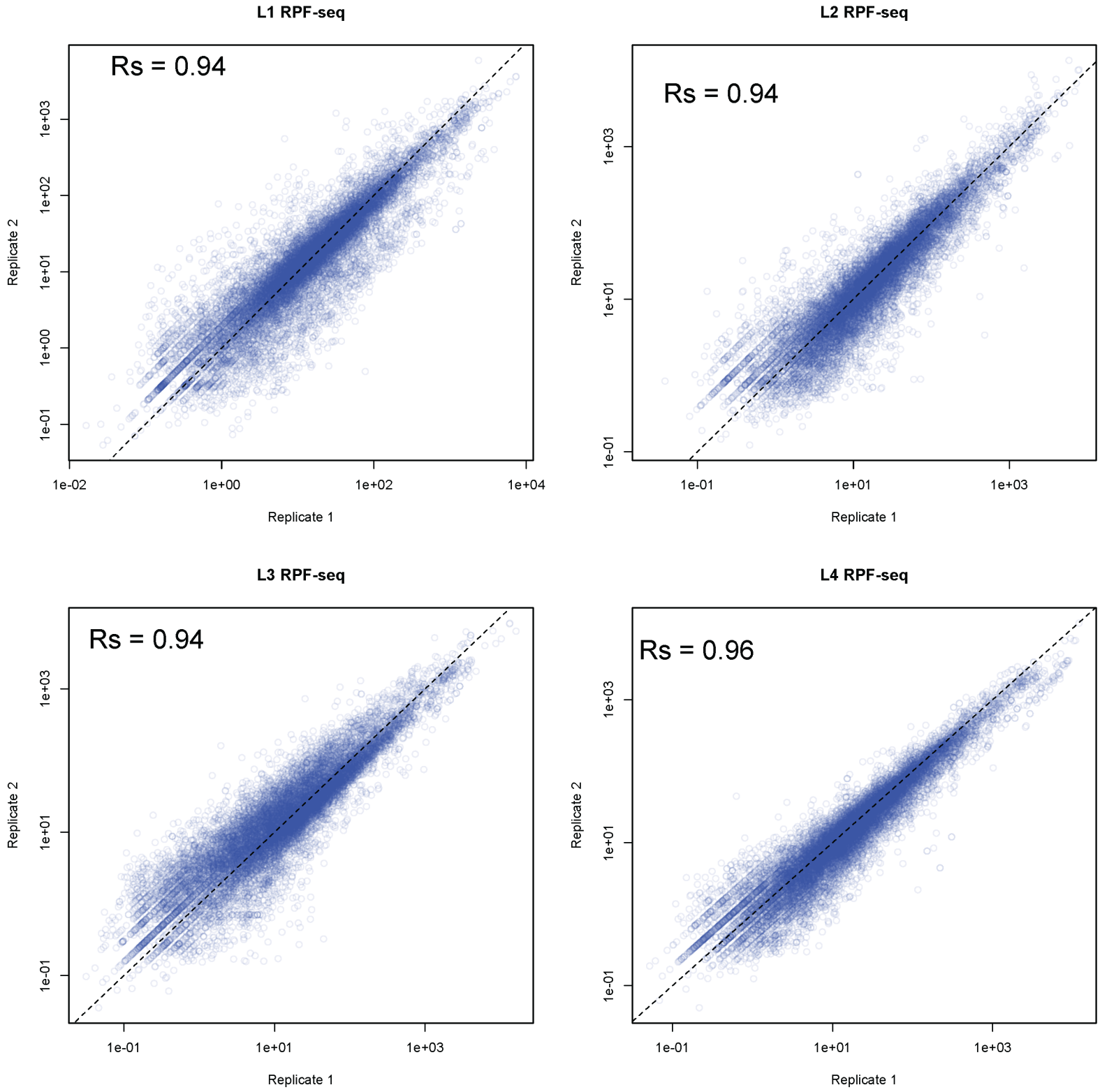
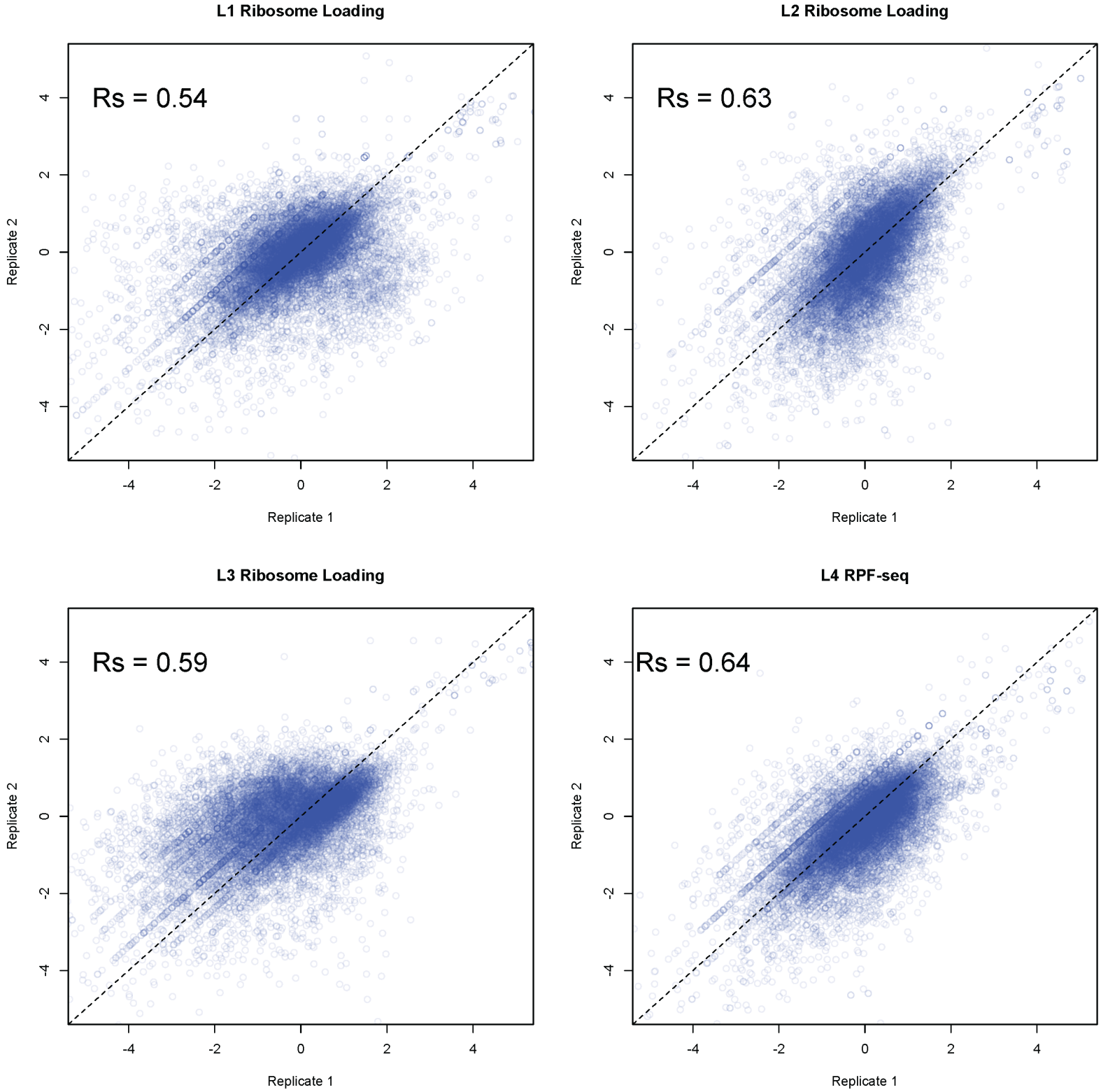
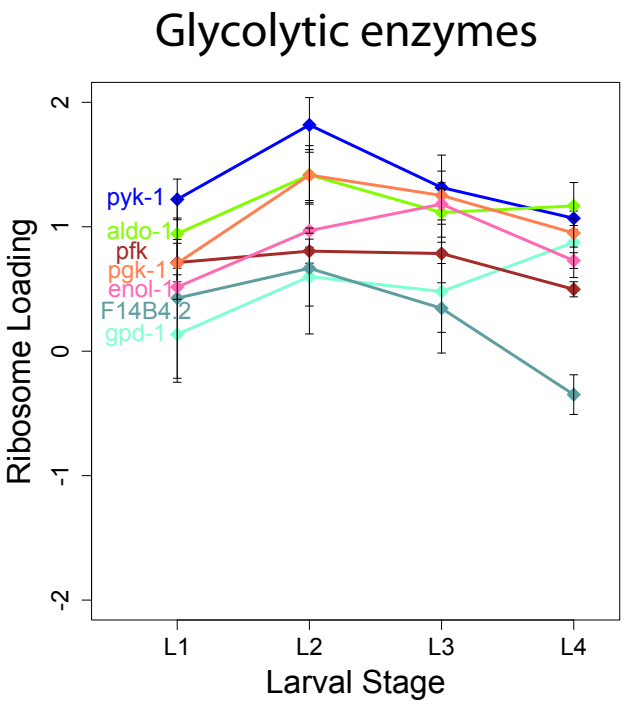
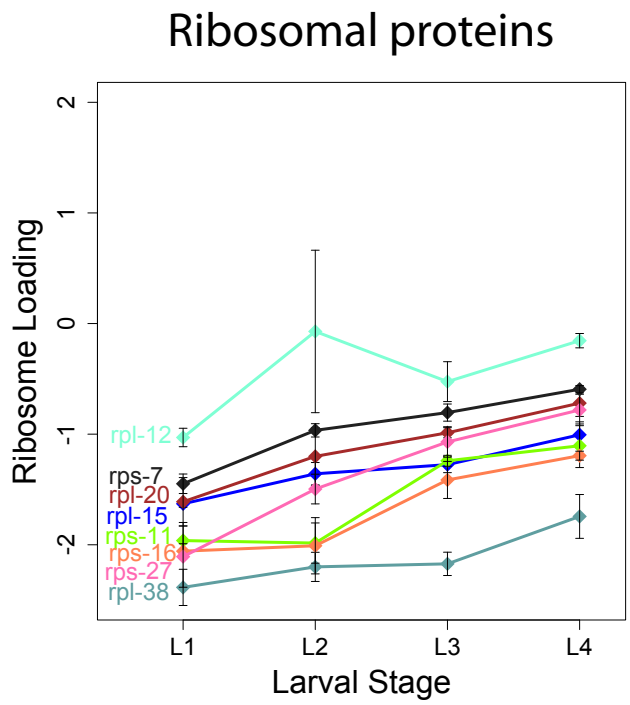

L1 RPF-seq
Rs = 0.94
Replicate 2
Replicate 1
1e-02
1e+00
1e+02
1e+04
1e-01
1e+00
1e+01
1e+02
1e+03
L2 RPF-seq
Rs = 0.94
Replicate 2
Replicate 1
1e-01
1e+01
1e+03
L3 RPF-seq
Rs = 0.94
Replicate 2
Replicate 1
1e-01
1e+01
1e+03
L4 RPF-seq
Rs = 0.96
Replicate 2
Replicate 1
1e-01
1e+01
1e+03

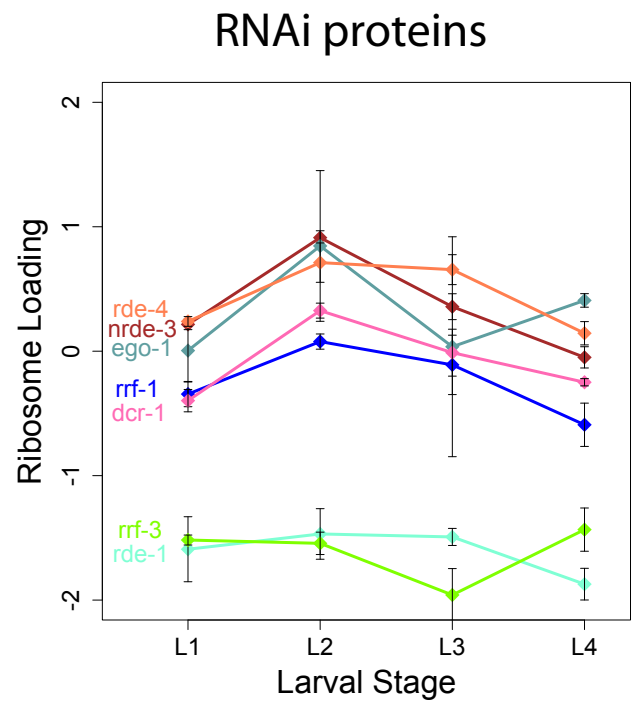
Figure S8


L1 Ribosome Loading
Rs = 0.54
L2 Ribosome Loading
Rs = 0.63
L3 Ribosome Loading
Rs = 0.59
L4 RPF-seq
Rs = 0.64
Replicate 1
Replicate 2
-4
-2
0
2
4

# Figure S9


Glycolytic enzymes
Ribosome Loading
Larval Stage
L1
L2
L3
L4
pyk-1
aldo-1
pfk
pgk-1
enol-1
F14B4.2
gpd-1


Ribosomal proteins
Ribosome Loading
Larval Stage
L1
L2
L3
L4
rpl-12
rps-7
rpl-20
rpl-15
rps-11
rps-16
rps-27
rpl-38


RNAi proteins
Ribosome Loading
Larval Stage
L1
L2
L3
L4
rde-4
nrde-3
ego-1
rrf-1
dcr-1
rrf-3
rde-1