Supplementary Material
Target mRNA abundance dilutes microRNA and siRNA activity

Aaron Arvey, Erik Larsson, Chris Sander, Christina S. Leslie * and Debora S. Marks*

*Joint corresponding author

Contents:

Other factors correlated with decreased down-regulation

A+U-rich microRNAs may target more mRNAs (since many UTRs contain AU-rich elements), leaving open the possibility that the correlation we see is a result of A+U-rich microRNAs. Neither A+U content of the seed-region nor the entire small RNA are significantly correlated with mean down-regulation (Supplementary Figure 5).

The length of 3'-UTRs is weakly correlated with RNA-seq reads (Supplementary Figure 6) and mean down-regulation (Supplementary Figure 7). Since the targets of low-target abundance small RNAs tend to have longer 3'-UTRs and longer 3'-UTRs are associated with increased down-regulation, it could be argued that the length of 3'-UTR is the real determinant of mean down-regulation. We controlled for this factor using the shared target analysis presented in the main text (Fig 2) and further described in the next section.

Finally, we must consider bias arising from microarray artifacts where transcripts with higher expression levels may exhibit more consistent down-regulation. We found correlation between down-regulation and individual transcript abundance in 165/178 transfection experiments, 101 of which are significant at a level of $p<0.05$ (indicating that at least 95 experiments are affected) (Supplementary Figure 8). If a miRNA has thousands of targets (*e.g.* miR-16), the majority of these targets may be sampled from lower expressed genes, making the down-regulation seem smaller on average. We controlled for this effect by the shared target analysis presented in the main text (Fig 2) and described below.

Significance of differential down-regulation on sets of shared targets

Fig. 2 in the main text shows the differential down-regulation on a shared set of targets of two miRNAs. There are n=168 points from n=19 microarray experiments (Supplementary Table 4). This is the most statistically sound test we performed since it controls for transcript specific effects. However, since we are looking at mRNAs that two miRNAs target, the average 3' UTR length is nearly twice as long as an average 3'-UTR (2041nt versus 1167nt). While this may not introduce any additional artifacts, longer UTRs may be more susceptible to degradation mediated by small RNAs (Supplementary Figure 7, Jacobson et al, in preparation). We also note that the

background distribution nominally shows some correlation (Fig. 2d). The correlation in the background distribution could be a result of down-regulation attributable to targets that lack heptamer seed matches. Alternatively, the background correlation could reflect one of the transcript-specific biases we previously discussed. Regardless of the source of background correlation, the predicted targets show significantly increased correlation (Fig 2c,d).

We also ensured that lab-specific artifacts are not responsible for the observed correlation in down-regulation and target abundance of shared targets. We did this by performing an identical analysis using only independent miRNA transfections (Methods) from a single study. Namely, we examined the study of Grimson et al. (2007) in the shared target framework. We found a highly significant rank correlation in the 36 pairs of differential down-regulation and differential target abundance ($P < 10^{-6}$, ρ=0.74). We also found a highly significant empirical p-value ($P < 10^{-5}$) when we simulated shared targets as above but sampling only from the Grimson et al.

Our interpretation of the results assumes that microRNAs have equal proportions of false positive predicted targets. If false positives increased with increasing numbers of predicted targets *disproportionately,* we could see the anti-correlation between down-regulation and target abundance. For instance, if all microRNAs target exactly 50 genes (and these genes are always predicted as targets), then all additional predictions would be false positives. Thus, if all true targets are down-regulated 2-fold (log expression ratio of -1) and all false prediction targets are not down-regulated (log expression ratio of 0), then 100 total predictions would lead to a mean log expression ratio of -0.5 whereas a 500 total predictions would lead to a mean log expression ratio of -0.1. So even though the true targets are down-regulated to a similar extent, the mean is dictated exclusively by the false positives. This is an alternative hypothesis that could explain the experimental data, would still be interesting and could be examined in future work.

Michalis-Menten kinetics fit transcript degradation

In the main text, we describe a model for total transcript degradation rate. Here we describe the degradation reaction, assumptions behind the kinetic equations, alternative models, and statistical tests to determine the model that bests fits the data.

The mRNA transcript degradation reaction is mediated by small RNA loaded into RISC. This can be considered a classic multiple turnover enzyme (Haley & Zamore, 2004). Thus, we can consider the kinetics of the reaction

$$[E]+[S]\longleftrightarrow[ES]\longrightarrow[E]+[P],$$

where E represents the small RNA-loaded RISC, S represents the target transcripts, and P represents the degraded transcripts. We are interested in knowing [S] after the degradation reaction has proceeded for a fixed amount of time. Namely, let *x(T)* be the total abundance [S] of all targets at time *T,* so that *x(0)* is initial abundance and *x(T=1)* is the abundance at time of microarray assay. We model the time rate of change of the absolute number of molecules degraded (the velocity) $v(T)$ in each experiment using Michaelis-Menten kinetics to determine the initial velocity,

$$v(0)=\frac{dx(0)}{dt}=\frac{V_{\max}x(0)}{x(0)+K_m}.$$

Since we only have a single time point for most mi/siRNAs, we assume that velocity is constant over the time between transfection and microarray array and can be estimated from log expression change between T=0 (time of transfection) and T=1 day, and we only use $v=v(0)$ in the subsequent discussion.

The Michaelis-Menten model from the main text makes the following assumptions:

(i) No new production of S. We ignore target transcript production rates, which are likely to vary across sets of miRNA targets.

(ii) [E] remains constant, since RISC tends to be saturated by the transfected small RNA(Khan et al, 2009). Thus we ignore the association rates of small-RNA and RISC.

(iii) [P] is the amount of degradation that has occurred.

The model also asserts the standard assumptions of Michaelis-Menten kinetics, including free diffusion, and [ES] reaches a steady state (implying that there is always more target mRNAs to be degraded).

We explored three models for total transcript degradation: constant, linear, and Michaelis-Menten. The constant model supposes that *x(0) - x(T) = c*, where c is constant. The linear model

supposes that $x(0) - x(T) = mx(0) + b$. The Michaelis-Menten model supposes that $x(0) - x(T) = V_{max} x(0) / (K_m + x(0))$. We find highly significant linear and rank correlation between transcripts degraded and total transcript abundance ($P < 10^{-17}$, ρ=0.65, p=0.63). Thus a non-constant monotonic or linear model is a better fit than the constant model.

To determine the improvement of Michaelis-Menten over linear, we performed 10,000 iterations of determining the parameters (Vmax and Km OR m and b) on 37% of the data and computing sum of square error (SSE) on the remaining 63% of the data (similar in spirit to 0.632 bootstrapping (Efron, 1983)). We find that 9,170 iterations show Michaelis-Menten with lower SSE and 830 show the linear model with lower SSE. Thus, Michaelis-Menten kinetics are a better fit to transcript degradation amounts.

We also note that if transcript degradation were in fact linearly related to target abundance, our hypothesis that target concentration limits down-regulation would not hold. This is a result of

$$\log\left(\frac{x(T)}{x(0)}\right) = \log\left(\frac{x(0) - mx(0) - b}{x(0)}\right) = \log\left(1 - m - b/x(0)\right) \approx \log\left(1 - m\right).$$

The final approximation holds since b empirically tends to be much smaller than $x(0)$. Since m is a constant, linear degradation kinetics would suggest constant down-regulation across all experiments. However, since we find significant correlation between target concentration and mean down-regulation (Fig 1 in the main text), the degradation kinetics must be non-linear.

Microarray normalization

Microarray data is typically normalized when comparing expression change for a specific gene. However, standard normalization techniques may actually bias us to believe our hypothesis even if it is not true. For example, consider the null hypothesis that all target genes are down-regulated 3-fold irrespective of target abundance or number of target genes (Supplementary Fig 12a). If a small RNA has more target genes, then the normalization technique of mean centering and/or dividing by standard deviation results in a greater mean-shift of the target set. This in turn makes for a situation where we should have rejected our hypothesis, but we instead reject the null hypothesis (Supplementary Fig. 12b and 12c).

We thus avoid normalization of the microarrays in all of our experiments in the main text and supplementary material. In lieu of normalization, we evaluated other measures of down-regulation, such as the area between the CDF curves. However, even though we do not consider it valid, we still wanted to see if normalization made any difference to our final conclusions. Statistical tests for

i. mean down-regulation correlation with target abundance,
ii. shared targets differential down-regulation correlation with target abundance difference (both nominal and empirical tests), and
iii. siRNA primary target down-regulation correlation with target abundance

are all still significant at $P < 0.01$. However, due to the biases mentioned above, we do not consider these to be valid statistical tests in support of our hypothesis. We provide the results here for completeness.

Note that the normalization we perform in the context of siRNA primary targets does not suffer from the bias described above since we are only looking at a single target and normalizing across experiments (in contrast to normalizing single experiments separately).

References

Efron B (1983) Estimating the Error Rate of a Prediction Rule: Improvement
on Cross-Validation. *Journal of the American Statistical Association* **78:** 316-331

Jacobson A, Marks D, Krogh A (Under Review) Systematic discovery of novel motifs which modulate microRNA regulation. *In Preparation.*

Khan A, Betel D, Miller M, Sander C, Leslie C, Marks D (2009) Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. *Nat Biotech* **27:** 549-555

Supplementary Figure Legends

Supplementary Figure 1. Alternative methods for quantifying down-regulation and target abundance. (a) Mean down-regulation and area between cumulative distribution curves are highly correlated. (b) Mean down-regulation and percentage of target molecules degraded are highly correlated. (c) Predicted target abundance by RNA-seq (RPN) and array fluorescence are highly correlated. The arrays used by Anderson et al. have different range and are shown separately from other experiments. (d) Predicted target abundance by RNA-seq (RPN) and number of genes targeted are highly correlated.

Supplementary Figure 2. Multiple target predictions methods yield similar results. The methods examined include (a) conserved heptamers, (b) TargetScan, and (c) miRanda.

Supplementary Figure 3. Transfection experiments in HCT-116 cells show mean log expression ratio is correlated with concentration of predicted targets, as estimated using array fluorescence.

Supplementary Figure 4. Transfection experiments followed by quantitative proteomics show that mean down-regulation in protein expression is anti-correlated with the number of predicted targets.

Supplementary Figure 5. AU-content is not significantly correlated with average down-regulation. The mean down-regulation is plotted with respect to AU content of the seed and entire miRNA. The correlation measures (Pearson's and Spearman's) are ***not*** statistically significant at a threshold of $P<0.05$.

Supplementary Figure 6. The length of RefSeq 3' UTR is weakly but significantly correlated with RPN.

Supplementary Figure 7. The average length of the 3' UTR for predicted targets is correlated with mean down-regulation.

Supplementary Figure 8. Scatter plots for each microarray experiment showing individual target transcript abundance versus log expression ratio. The majority of plots show significant negative correlation, indicating that highly expressed transcripts are more consistently down-regulated.

Supplementary Figure 9. Off-target abundance determines knock-down efficiency of primary targets by siRNA. (a) The same as Fig. 1f. (b) A t-test of the data shown in (a). (c-f) The raw values prior to normalization and agglomeration of the data sets.

Supplementary Figure 10. Estimation of kinetic parameters assuming a linear or a Michaelis-Menten model for velocity of degradation. The Michaelis-Menten model is a better fit with sum of squared residuals of 6974.2. There is a significant dependence on molecule abundance as determined by rank correlation.

Supplementary Figure 11. Selection of time point for transfection experiments. When we look at individual datasets and two time points (12h and 24h) are available, we only take the most down-regulated time point. While this is usually the later time point, there are instances where it is the earlier time point.

Supplementary Figure 12. Possible biases introduced by normalization techniques.

Supplementary Table Legends

Supplementary Table 1. Description of datasets analyzed. Columns: small RNA, Family, Seed, Sense, Hour, Accession, Array, File, Reference.

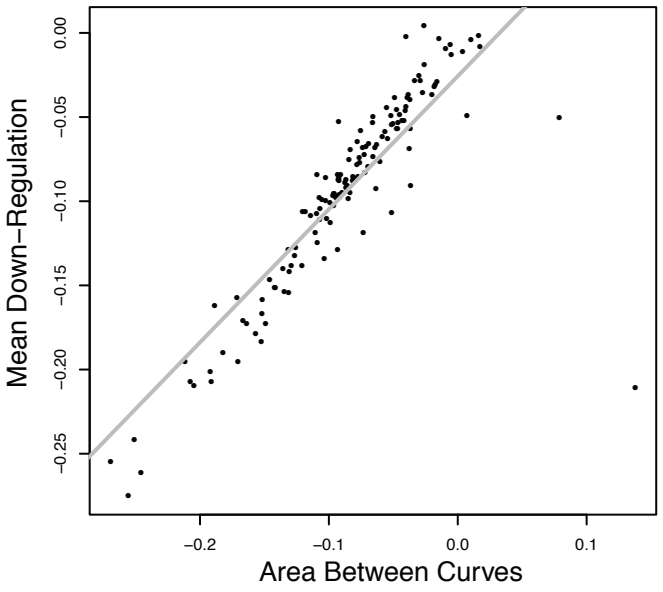
Supplementary Table 2. Quantification of transcript abundance and down-regulation exerted by small RNA. Columns: small RNA, Hour, Accession, File, RPK, Area between curves, Mean down-regulation.

Supplementary Table 3. The expression ratio and RPN for all targets in each transfection experiment. For each experiment, we give the small RNA, hour, accession, filename, total target RPN, average down regulation, and area between curves. We provide the gene name, expression ratio, and RPN for each individual target.

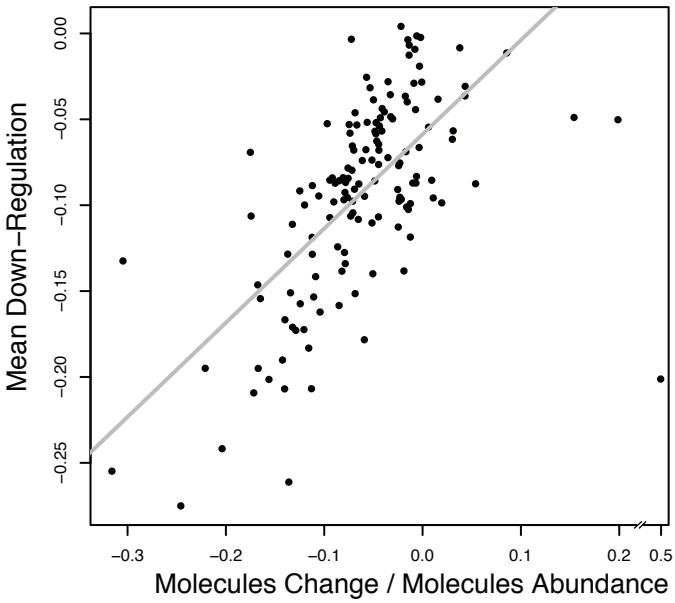
Supplementary Table 4. The change in expression of individual genes that are predicted to be targets of two microRNAs. Each table shows the predicted targets of two microRNAs along with the change in expression of each target and the target abundance.

Supp Fig 1

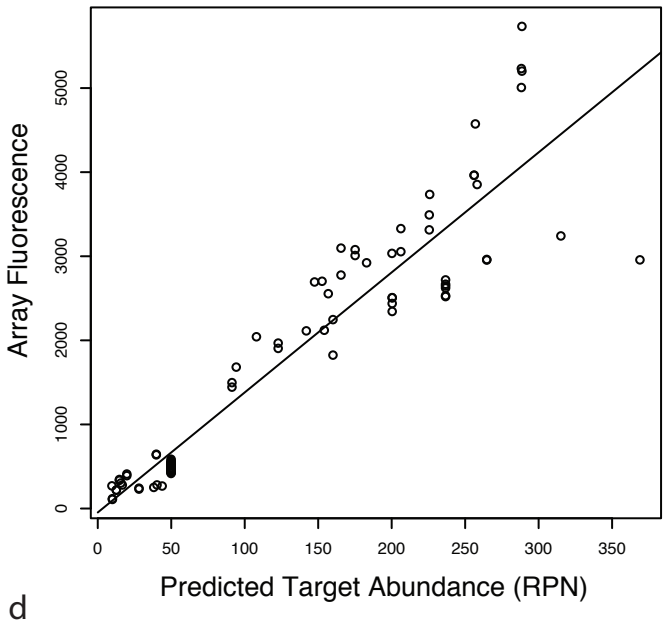
a

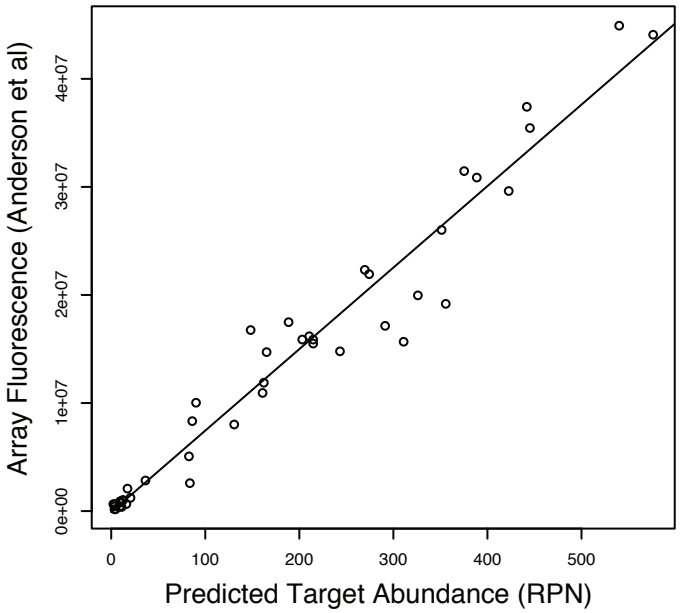


b

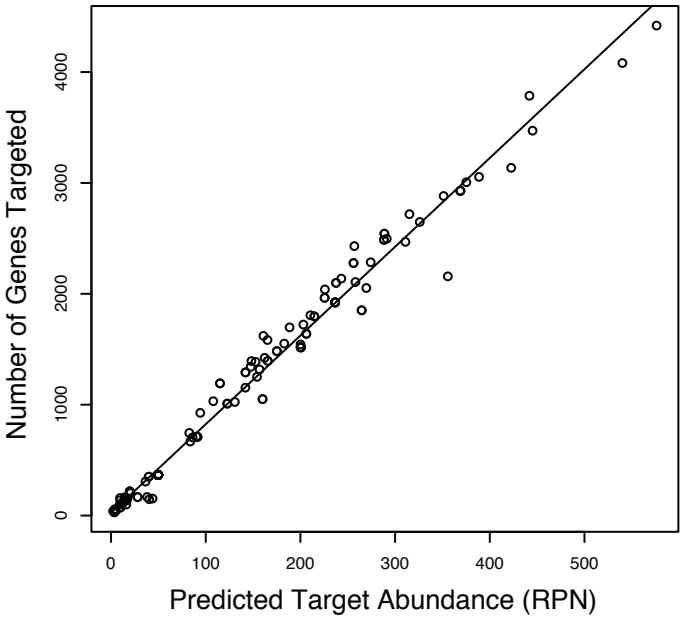


c

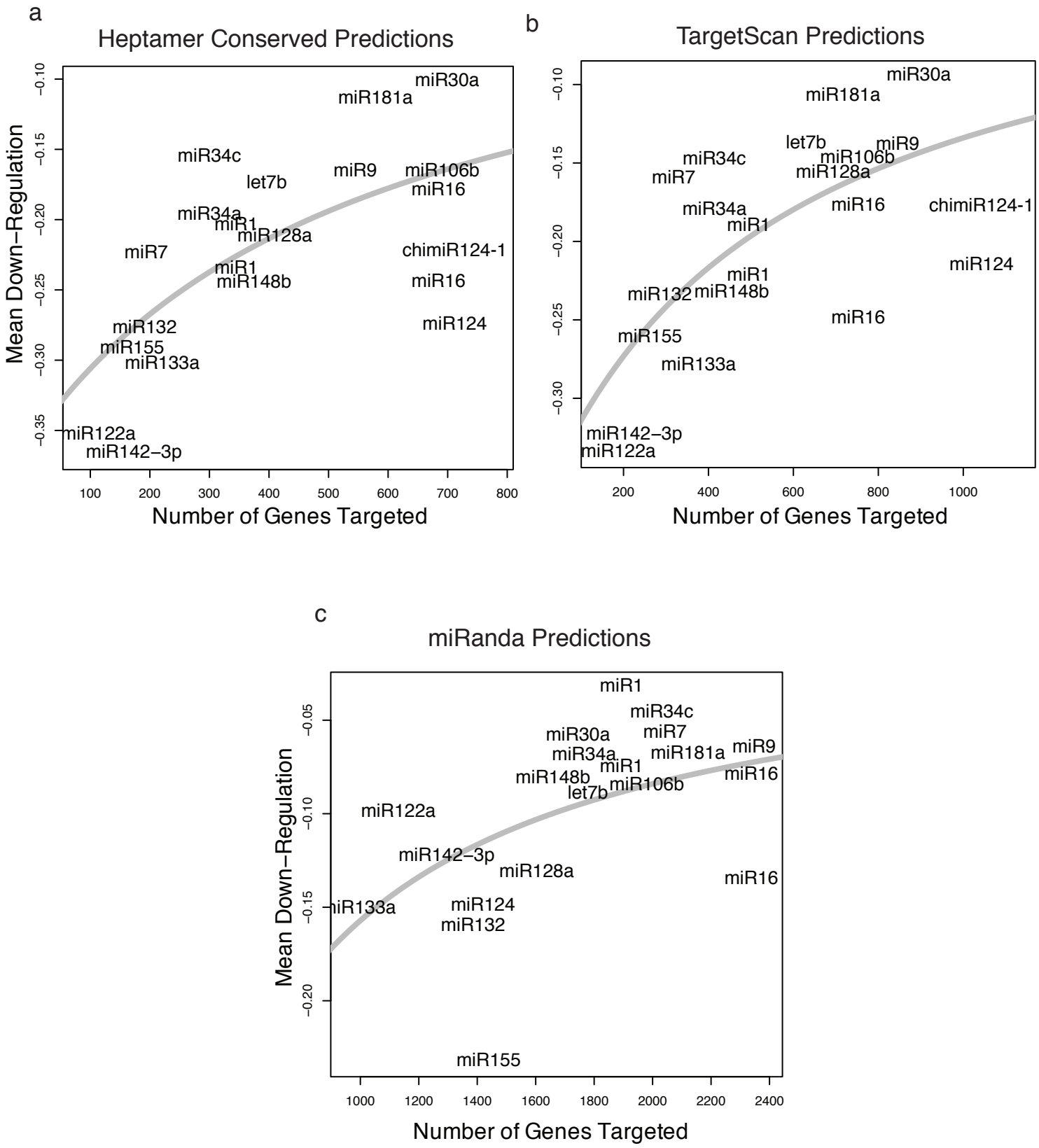




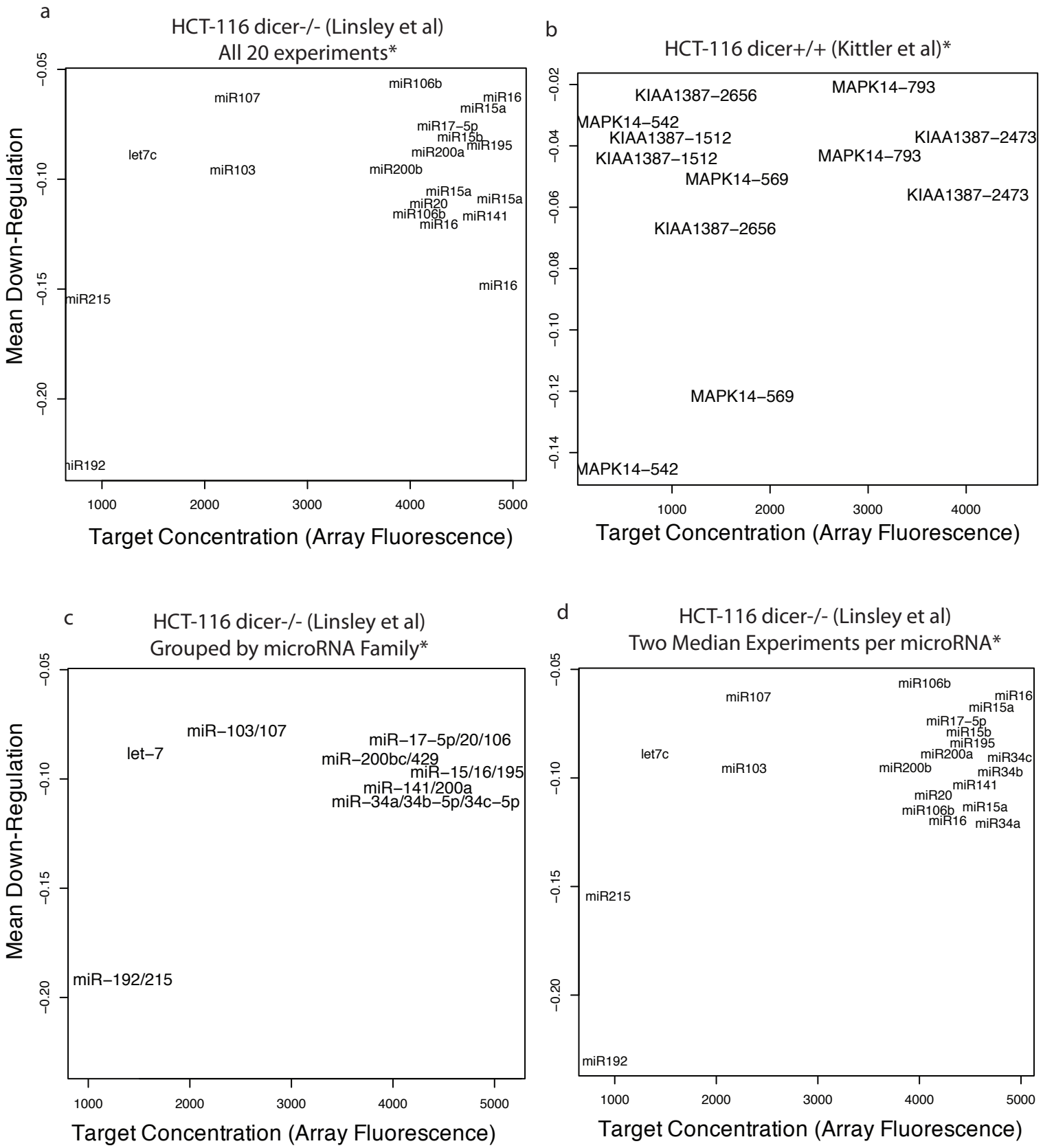
d



Supp Fig 2

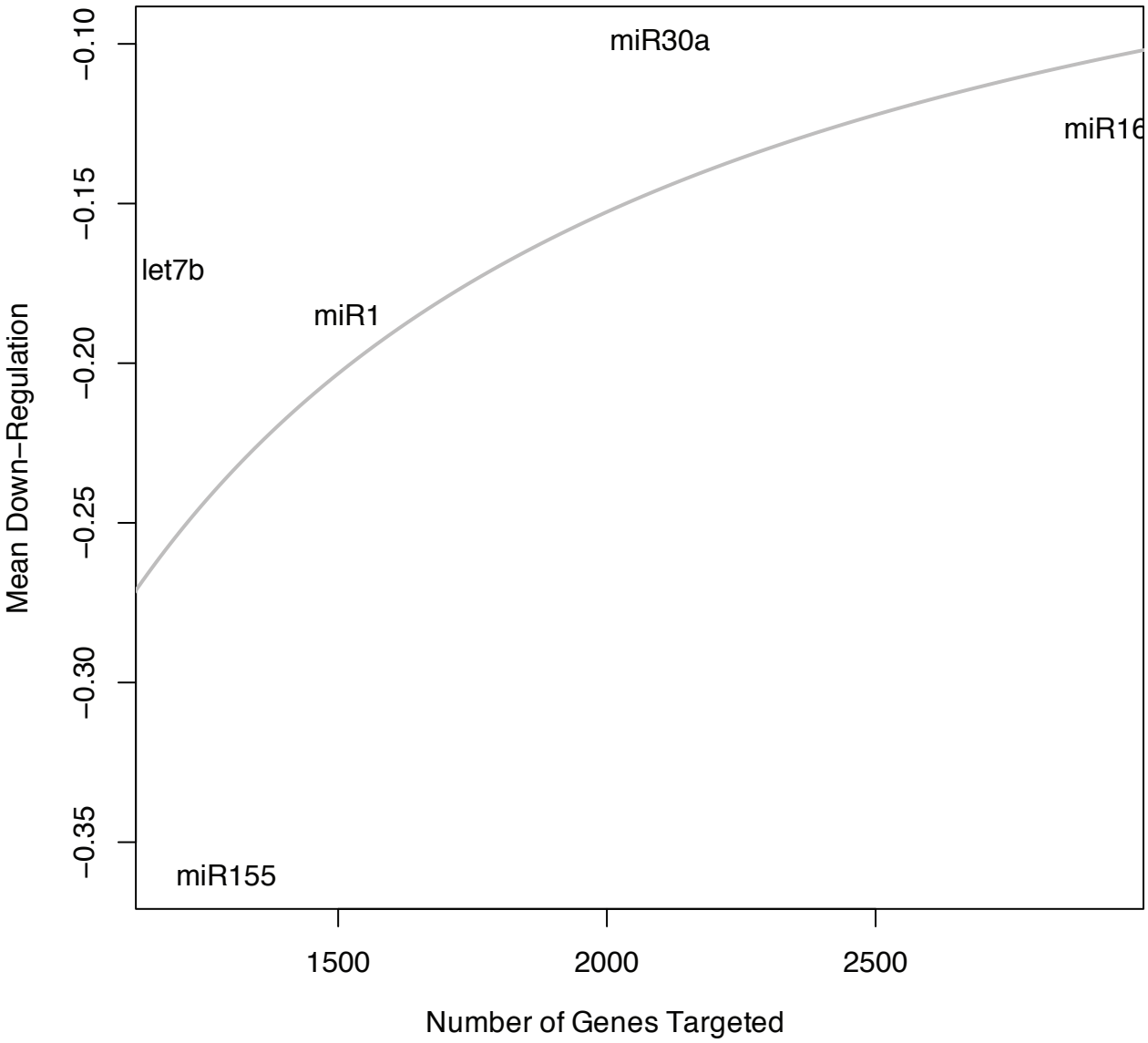

a
Heptamer Conserved Predictions
Mean Down−Regulation
Number of Genes Targeted
miR30a
miR181a
miR34c
miR9
miR106b
let7b
miR16
miR34a
miR1
miR128a
miR7
chimiR124-1
miR1
miR148b
miR16
miR132
miR124
miR155
miR133a
miR122a
miR142−3p
b
TargetScan Predictions
Number of Genes Targeted
miR30a
miR181a
let7b
miR9
miR34c
miR106b
miR7
miR128a
miR34a
miR16
chimiR124-1
miR1
miR124
miR1
miR132
miR148b
miR16
miR155
miR133a
miR142−3p
miR122a
c
miRanda Predictions
Mean Down−Regulation
Number of Genes Targeted
miR1
miR34c
miR30a
miR7
miR34a
miR181a
miR9
miR1
miR148b
miR16
miR106b
let7b
miR122a
miR142−3p
miR128a
miR16
miR124
miR132
miR155

Supp Fig 3



* Points have been moved slightly to avoid overlapping.
Statistics in text are calculated from original positions.

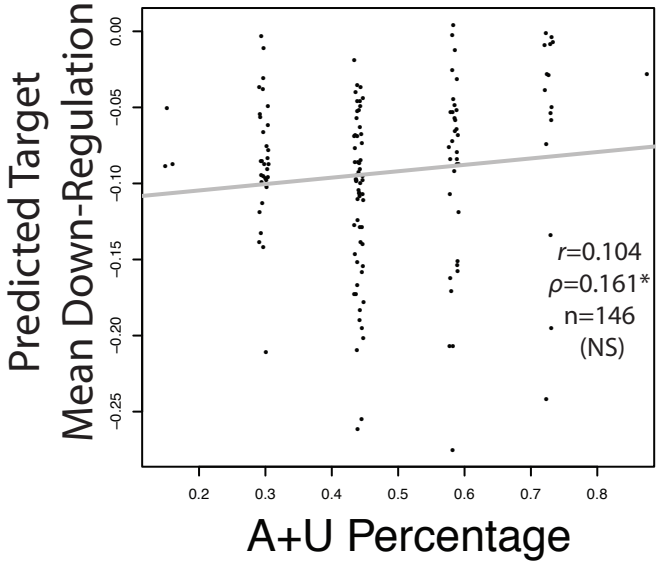
Supp Fig 4


Mean Down−Regulation
−0.10
−0.15
−0.20
−0.25
−0.30
−0.35
miR30a
miR16
let7b
miR1
miR155
1500
2000
2500
Number of Genes Targeted

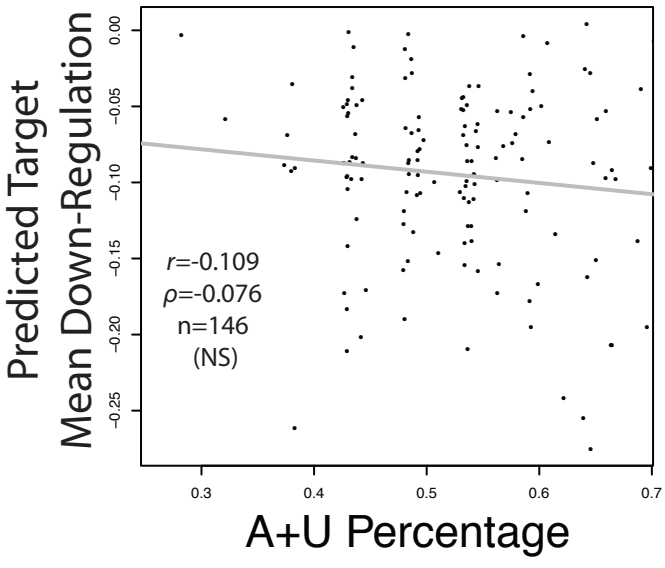
Supp Fig 5

A+U Content is ***not*** significantly correlated (Pearson or Spearman) with mean down-regulation

Seed A+U Content



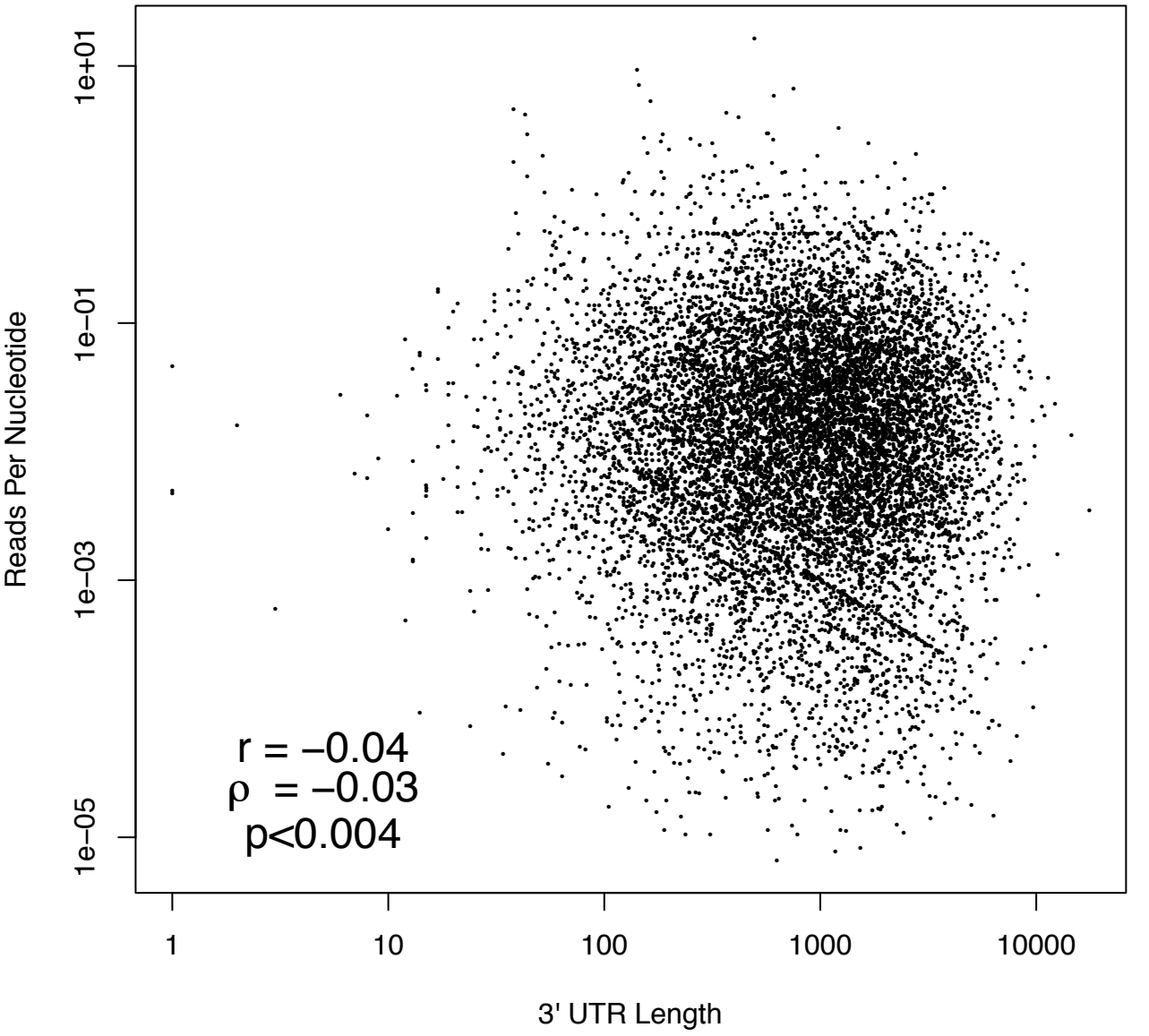
Whole miR A+U Content



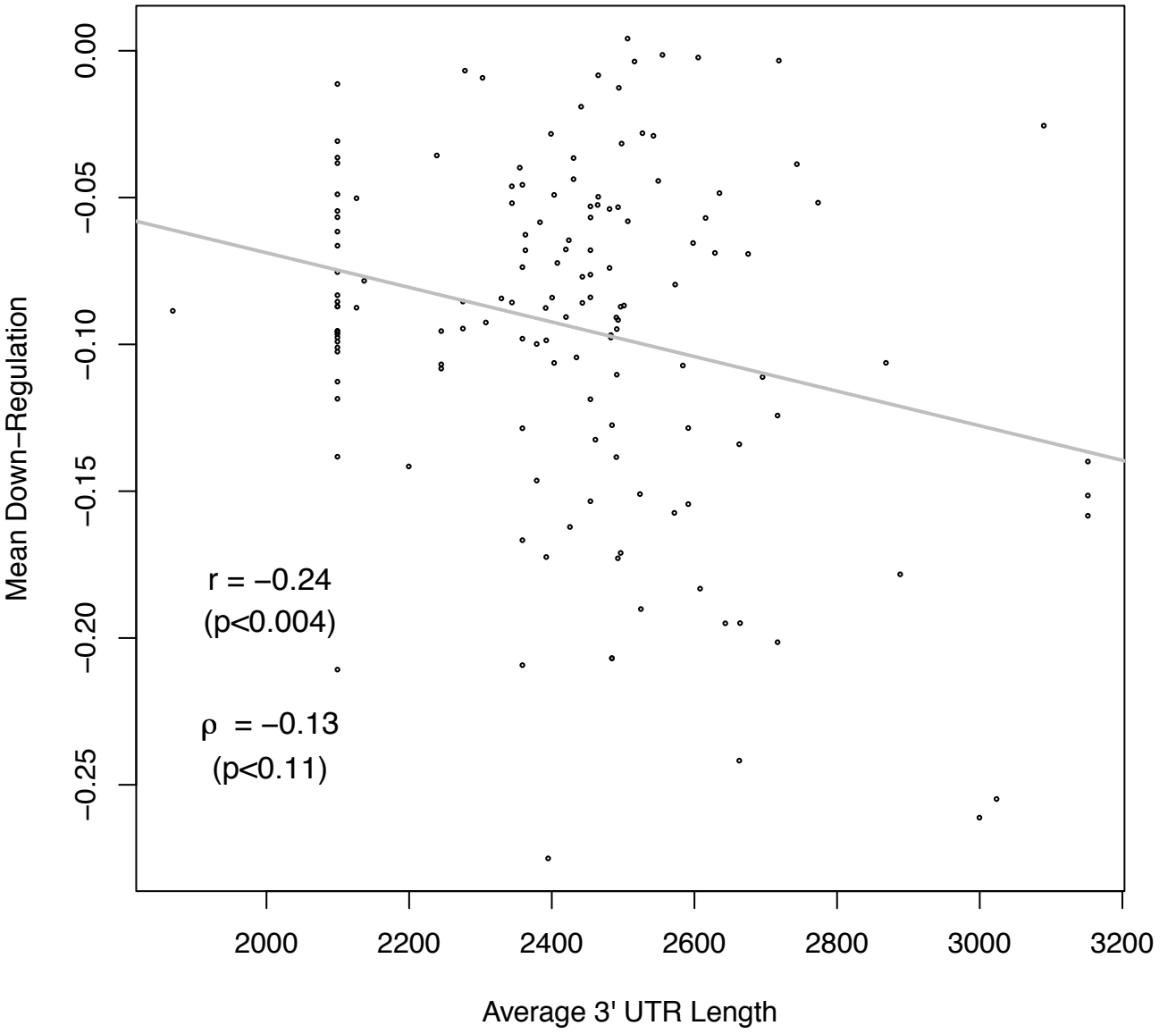
NS = Not statistically significant at a threshold of 0.05

* Spearman rank correlation is unstable due to many "ties" in A+U Percentage rank across the small RNAs. There are only 6 possible ranks for n=146 small RNAs and 4 ranks have the majority of data points (n=142).

Supp Fig 6


Reads Per Nucleotide
1e+01
1e−01
1e−03
1e−05
r = −0.04
ρ = −0.03
p<0.004
1
10
100
1000
10000
3' UTR Length

Supp Fig 7

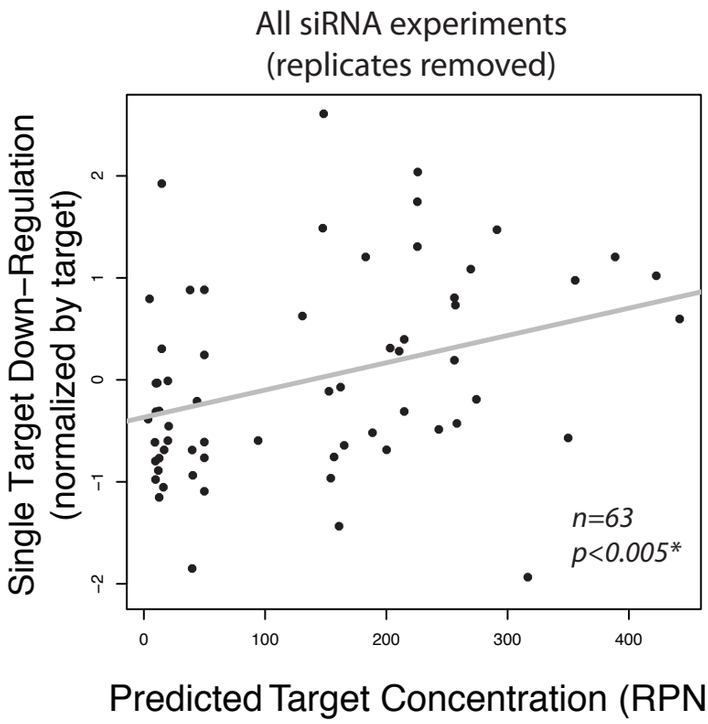

Mean Down−Regulation
0.00
−0.05
−0.10
−0.15
−0.20
−0.25
r = −0.24
(p<0.004)
ρ = −0.13
(p<0.11)
2000
2200
2400
2600
2800
3000
3200
Average 3' UTR Length

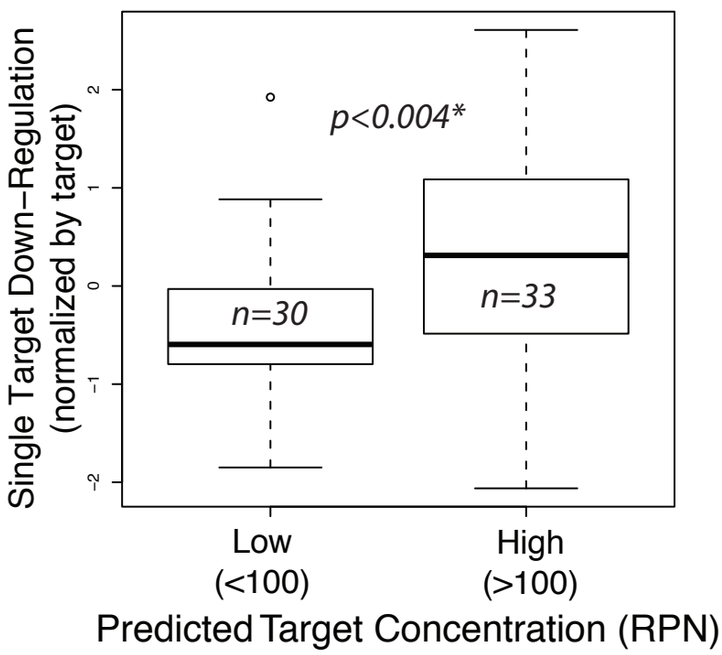
Supp Fig 8

See additional file for Supp Fig 8.

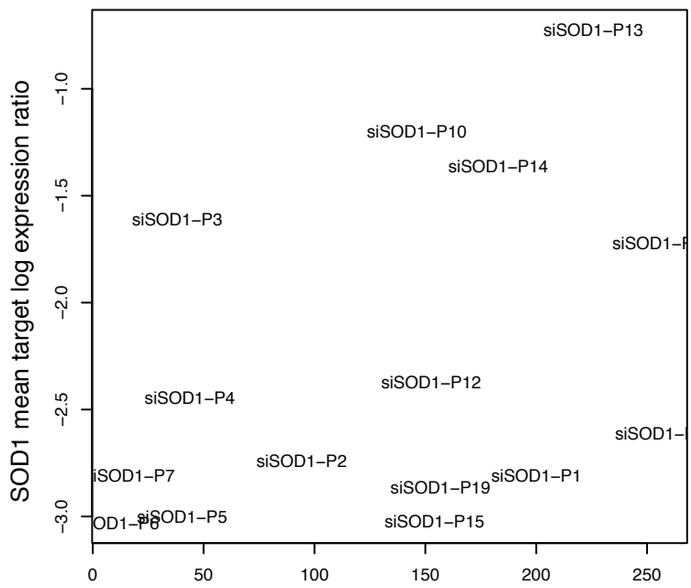
Analysis of siRNA effect on primary target (in contrast to off-targets)

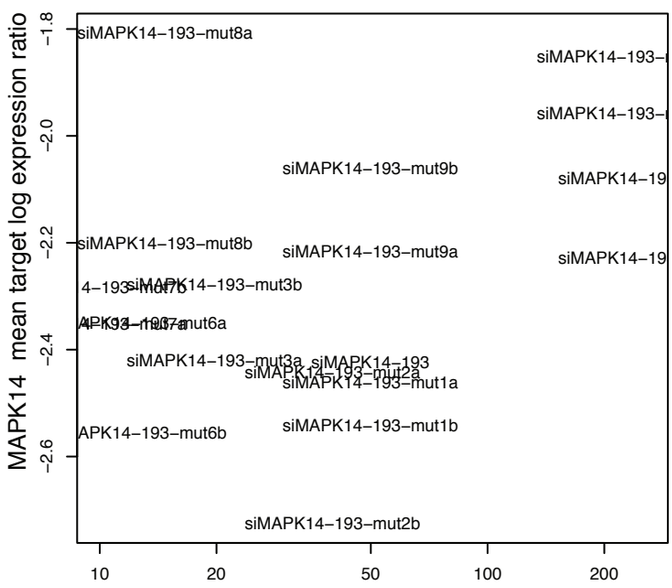
Supp Fig 9

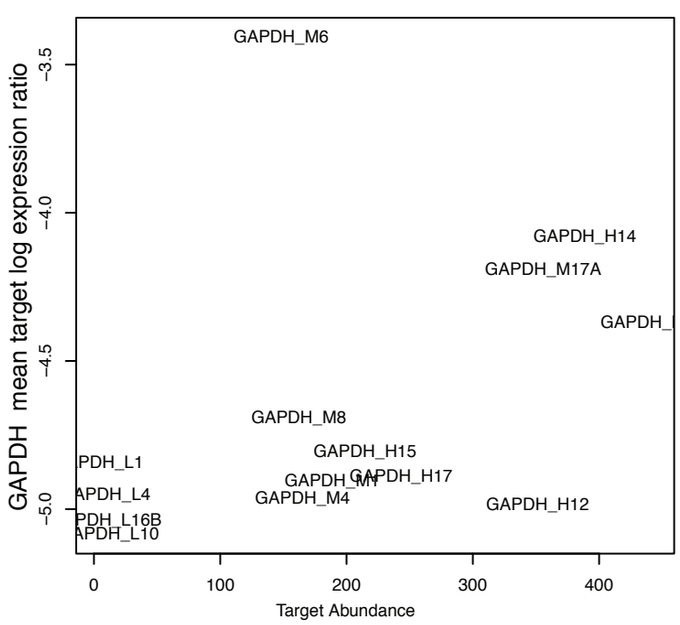


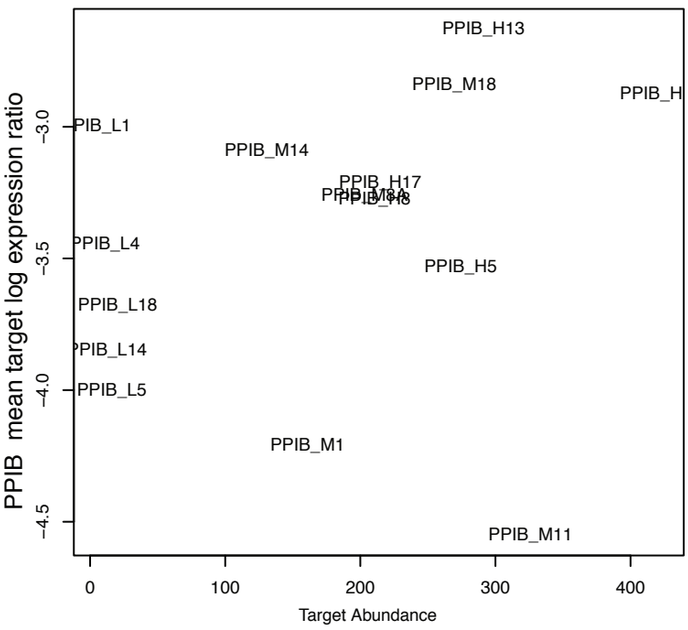


Raw data prior to normalization

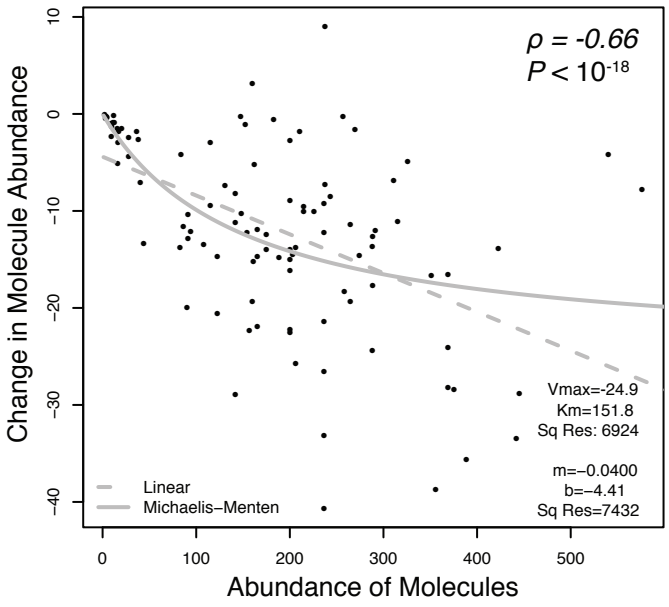




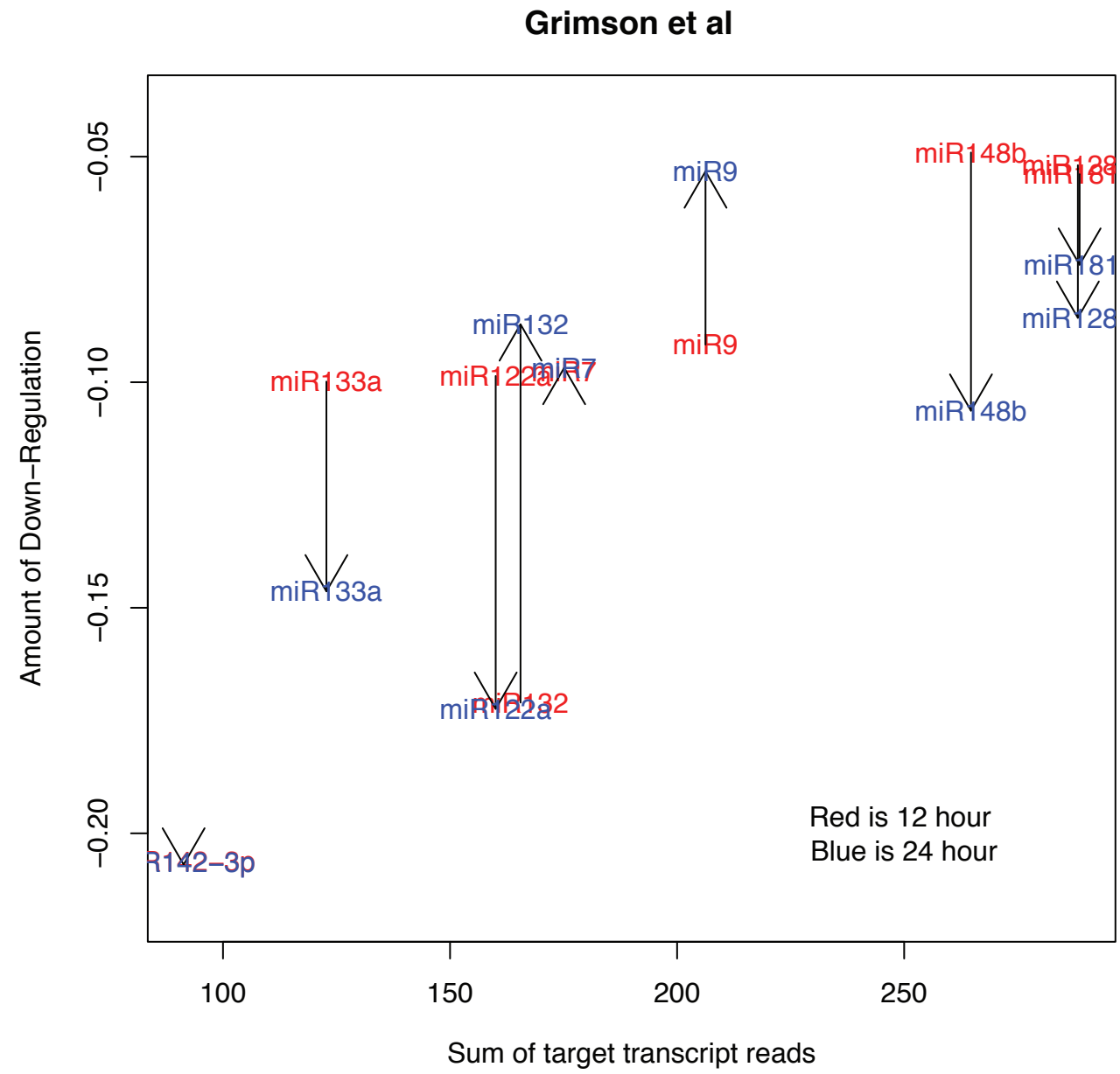


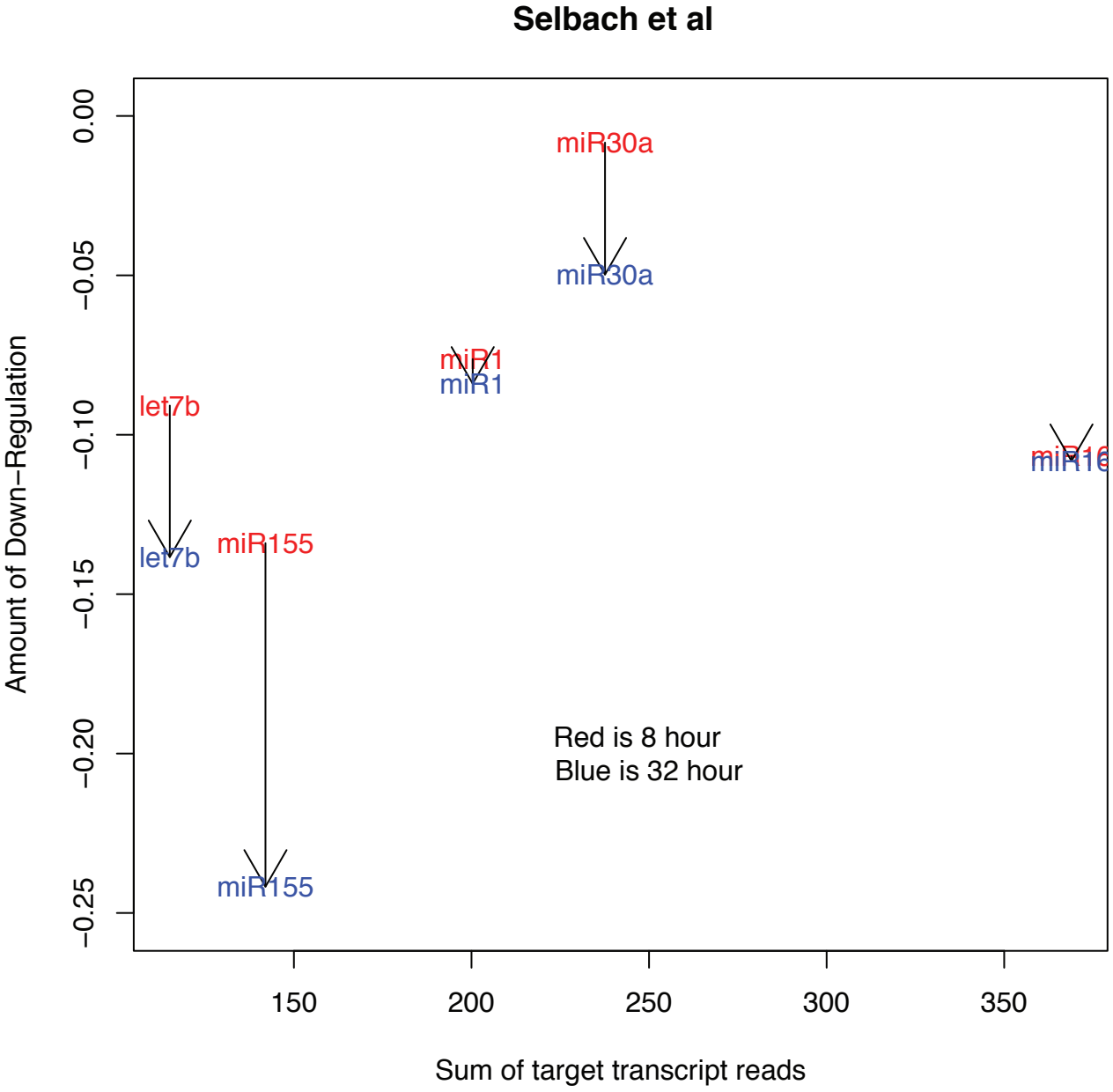


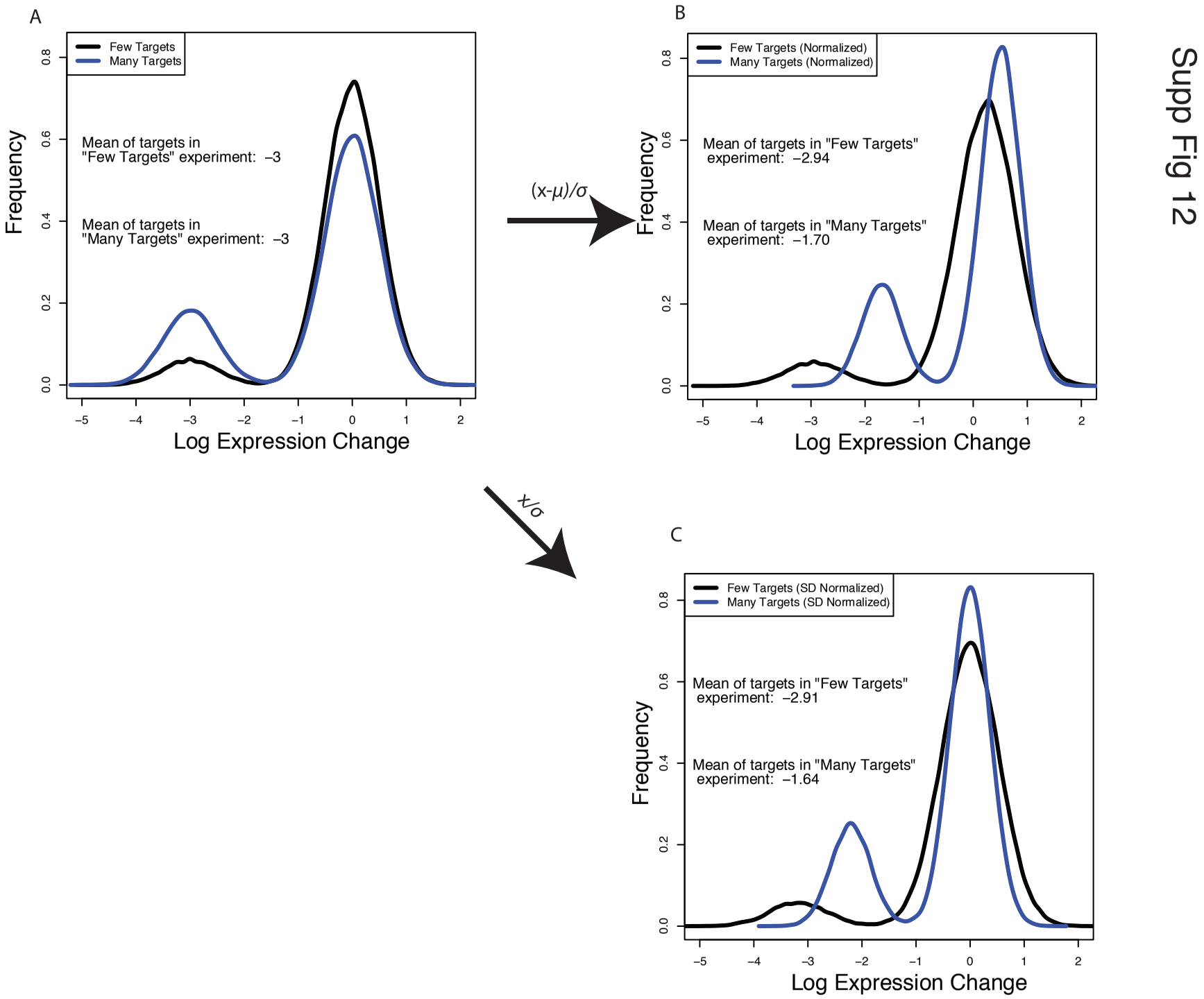
Supp Fig 10


ρ = -0.66
P < 10^-18
Vmax=-24.9
Km=151.8
Sq Res: 6924
m=−0.0400
b=−4.41
Sq Res=7432
Linear
Michaelis−Menten
Change in Molecule Abundance
Abundance of Molecules
10
0
−10
−20
−30
−40
0
100
200
300
400
500

Supp Fig 11


Grimson et al
Amount of Down−Regulation
−0.05
−0.10
−0.15
−0.20
miR9
miR9
miR148b
miR148b
miR181
miR128
miR132
miR7
miR133a
miR133a
miR122a
miR122a
miR132
R142−3p
Red is 12 hour
Blue is 24 hour
100
150
200
250
Sum of target transcript reads


Selbach et al
Amount of Down−Regulation
0.00
−0.05
−0.10
−0.15
−0.20
−0.25
miR30a
miR30a
miR1
miR1
let7b
let7b
miR155
miR155
Red is 8 hour
Blue is 32 hour
150
200
250
300
350
Sum of target transcript reads


A
Few Targets
Many Targets
Mean of targets in
"Few Targets" experiment: −3
Mean of targets in
"Many Targets" experiment: −3
Frequency
Log Expression Change
(x-μ)/σ
B
Few Targets (Normalized)
Many Targets (Normalized)
Mean of targets in "Few Targets"
experiment: −2.94
Mean of targets in "Many Targets"
experiment: −1.70
Frequency
Log Expression Change
x/σ
C
Few Targets (SD Normalized)
Many Targets (SD Normalized)
Mean of targets in "Few Targets"
experiment: −2.91
Mean of targets in "Many Targets"
experiment: −1.64
Frequency
Log Expression Change

Supp Fig 12