Re-analysis of genome wide data on mammalian microRNA-mediated suppression of gene expression

Ola Larsson¹ and Robert Nadon^{2,3}

¹Department of Oncology-Pathology; Cancer Center Karolinska; Karolinska Institutet; Stockholm, Sweden; ²Department of Human Genetics; McGill University; Montreal, QC Canada; ³McGill University and Genome Quebec Innovation Centre; Montreal, QC Canada

Keywords: microRNA, mRNA translation, mRNA stability, ribosome profiling, bioinformatics, statistics

Abbreviations: miRNA, microRNA; RPF, ribosome protected fragment; ARE, AU rich element

microRNAs are short endogenously expressed RNAs that regulate gene expression post-transcriptionally. Although both mRNA degradation and suppression of mRNA translation can mediate reduced protein levels following microRNA targeting of an mRNA, their relative contributions have remained elusive. A recent genome-wide study in mammals employing RNA-sequencing to measure microRNA effects on mRNA translation and stability concluded that 84–89% of microRNA-induced suppression of gene expression is due to degradation of target mRNAs. We re-analyzed this data set and applied a number of analysis modifications which revealed that the contribution of mRNA translation was likely underestimated for some mRNA subsets. Moreover, in contrast to the original analysis, our analysis indicated that suppression of mRNA translation precedes mRNA degradation upon microRNA targeting. Our findings thereby enhance our understanding of microRNA-mediated genome wide suppression of gene expression in mammals.

Introduction

MicroRNAs (miRNAs) are short, nuclear encoded RNAs that suppress gene expression through partial base pairing with target mRNAs, thereby regulating gene expression at the post transcriptional level.¹ In mammals, each miRNA typically targets hundreds of protein-encoding mRNAs and each mRNA is in turn often targeted by multiple miRNAs.¹ Both translational suppression and mRNA degradation have been proposed to explain how miRNAs suppress their target mRNAs. The relative contributions of these two mechanisms, however, remain unclear.^{2,3} Most genome-wide studies of miRNA suppression have measured levels of poly-A containing mRNAs, thereby focusing on the impact of miRNA suppression on mRNA stability.^{4–7} The role of translation has been technically more difficult to address at a genome-wide level. It is therefore only recently that the relative roles of mRNA stability and translational suppression have been studied on a genome-wide scale by measuring protein and mRNA levels using mass spectrometry and DNA microarrays, respectively.^{8,9} The conclusions from these studies were limited, however, to highly expressed mRNAs, whose mechanism of suppression may differ from mRNAs with lower expression levels. Moreover, comparisons between DNA microarray data and proteomics data are difficult to interpret because degradation of the

mRNA is measured at one specific time point whereas protein levels are measured over a time period. Thus if mRNA levels are measured pre-steady-state the direct comparison to protein levels will not be valid. Accordingly, data generated by the same technology may be preferred for detailed unambiguous genome-wide analysis. Other studies have used polysome-RNA preparations, a complex sample pooling scheme, DNA-microarrays, and mathematical models to calculate the contribution from mRNA translation.¹⁰ Following the methodological breakthroughs in RNA sequencing to measure translational activity,¹¹ Guo et al. recently published on the relative contribution of translational suppression and mRNA degradation to miRNA suppression using RNA-sequencing data to estimate both translation and stability. They applied the ribosome protected fragment (RPF) technique to measure translation¹² whereby translating ribosomes are immobilized and the protected RNA fragments are isolated and sequenced (generating sequencing reads corresponding to RPFs). RPF data therefore represent a snapshot of where ribosomes are positioned on the mRNA and allows for comparisons of relative ribosome-mRNA association under different conditions. To assess the contribution of RNA degradation, Guo et al. generated in-parallel RNA-sequencing data from randomly fragmented poly-A RNA.¹²

Correspondence to: Ola Larsson and Robert Nadon; Email: ola.larsson@ki.se and robert.nadon@mcgill.ca

Submitted: 01/22/13; Revised: 03/22/13; Accepted: 04/04/13

Citation: Larsson O, Nadon R. Re-analysis of genome wide data on mammalian microRNA-mediated suppression of gene expression. Translation 2013; 1: e24557; <http://dx.doi.org/10.4161/trla.24557>

A fundamental problem arises in this approach when dissecting the relative contribution of translation and RNA stability because the two data types are not independent.¹¹ Any difference between two treatments observed in the poly-A data should also be observed in RPF data, if the mRNA is translated, because RPF data are influenced by both differential mRNA levels and differential translation. Thus, to calculate how much translation contributes to the miRNA-mediated suppression of gene expression, the RPF data need to be corrected for the contribution of RNA stability.¹³ In an attempt to accomplish this, Guo et al. subtracted (in log scale) the poly-A miRNA vs. mock effects from the RPF miRNA vs. mock effects: $(\log_2[\text{RPF}_{\text{miRNA}}] - \log_2[\text{RPF}_{\text{mock}}]) - (\log_2[\text{polyA}_{\text{miRNA}}] - \log_2[\text{polyA}_{\text{mock}}])$. Such correction is potentially associated with spurious correlations;¹³ however, a better alternative approach¹² could not be used due to the lack of experimental replication (i.e., $n = 1$ for all conditions). Guo et al. also performed an additional normalization step motivated by a potential contribution from off-target effects to observed stability and translation effects. Off-target effects include de-repression of miRNA target-sites for miRNAs that are not targeted in the experimental model¹⁴ and other post-transcriptional regulation. In contrast, transcription would affect both poly-A and RPF data similarly if the mRNA is translated. This normalization is therefore needed only if post-transcriptional off-target effects act differently on stability and translation. This additional correction was performed by subtracting observed median stability and translation effects of the no-site control gene population (i.e., those that do not carry the experimentally targeted miRNA target site) from stability and translation effects, respectively, of each target mRNA.

Using this approach, Guo et al. reported that miRNA suppress gene expression primarily by degrading their target mRNAs inasmuch as only 11–16% of miRNA-mediated suppression was reported to occur through translation. They also concluded that the data did not support that suppression of mRNA translation precedes degradation of the target mRNA following introduction of miRNA.¹² Because of these unexpected conclusions, given abundant empirical single mRNA studies indicating substantial contributions from translation,¹⁵ we revisited the data set. Our re-analysis suggests that the conclusion that 84–89% of microRNA-induced suppression of gene expression is due to degradation of target mRNAs is not unequivocally supported by the data and that the importance of translation was likely underestimated for some subsets of miRNA targets. Moreover, in contrast to Guo et al., our analysis indicates that mRNA translation precedes degradation of the target mRNA during miRNA-mediated suppression of gene expression in mammals.

Results

Potential bias from normalization of stability- and translation-effects. Guo et al. assumed that translation and stability were affected differentially by off-target effects and that off-target effects could be estimated from a set of no-site control genes (i.e., those mRNAs that were not targeted by the experimentally tested miRNA). Comparison between a population of target mRNAs and a population of no-site controls is a common strategy to assess

target dependent regulation in genome-wide experiments (e.g. 16). Normalization using a no-site control population, however, has not to our knowledge been used to normalize effects of different mechanisms, here stability and translation, to compare their relative magnitude. The potential pitfall of using the no-site population for normalization in this context is that if that population is not affected by the same off-target mechanisms as the population that is being normalized, the resulting normalized data will at least partially reflect the impact of other mechanisms which act only on the no-site controls. Indeed, in DNA microarray gene expression data, a number of early bias-correction attempts that appeared reasonable from biological or technical perspectives failed in practice (by increasing bias and/or variability). They include background correction of 2-color cDNA data,^{17,18} the MAS 5.0 algorithm for Affymetrix data,¹⁹ and the use of house-keeping genes for normalization.²⁰ Therefore, an alternative strategy that is commonly used in analysis of genome-wide data are to assume that off-target effects will on average act similarly and that attempts to normalize are therefore best avoided. For these reasons, we were interested in assessing the validity of the no-site control normalization approach in the present study.

First we sought to assess whether the assumptions of the no-site control normalization holds. One assumption that can be tested is whether the no-site controls efficiently estimate off-target effects arising from post-transcriptional mechanisms, such as de-repression.¹⁴ To this end we compared the no-site control genes to the target genes on number of miRNA target sites not targeted by the introduced miRNA. The median number of miRNA target sites (for miRNAs that are expressed in HeLa cells²¹) among no-site controls was smaller than for experimentally-targeted mRNAs (Fig. S1-a; $p < 1e-8$ for all no-site control to target site category comparisons [Wilcoxon's rank sum test]; we used a rank sum test and show the medians in Figure S1a due to the skewed distribution of the number of target sites per mRNA). Other RNA elements could also mediate off-target effects and should be evenly distributed between the no-site control and the target gene categories (e.g., frequencies of AU-rich elements [AREs]). No-site control genes showed lower frequencies of ARE (Fig. S1-b; $p < 0.01$ for all no-site control to target site category comparisons [Fisher's exact test]). We also compared 3' UTR lengths to assess overall potential for regulation by other post-transcriptional mechanisms which could act as off-target effects. The no-site control genes had shorter 3'UTRs, indicating that they likely harbor fewer other potential off-target effect mediating RNA-elements (Fig. S1-c; $p < 1e-14$ for all no-site control genes to target site category comparisons [Student's t-test]). Thus, no-site control mRNAs differed substantially from target gene categories in key RNA features. We therefore performed our analysis under the assumption that off target effects will on average be similarly active in translation and stability; we also monitored how conclusions based on this normalization approach differed from those reached by Guo et al.

Analysis of non-overlapping subsets indicates a wider range for the contribution of mRNA translation to miRNA mediated suppression of gene expression. In the original study,¹² two cellular models were used: an in vitro model of HeLa cells studied at 12 and 32 h after transfection with mock, miR-1 or miR-155 and a

mouse neutrophil knockout model for miR-223 (mock and miR-223 knockout). The original estimate of how much translation contributed to the observed differential levels of RPF data was presented for the 32 h time point and the neutrophil model (it was concluded that the 12 h time point data supported these experiments). In light of empirical findings that translational repression is more pronounced early after miRNA targeting and is followed by RNA degradation,³ we re-analyzed both time points. Also, the original analysis¹² included assessment of three groups of miRNA-targeted mRNAs. Those mRNAs carrying the miRNA target site but where the protein was not detected in a proteomics study on the impact of miRNAs on protein levels^{8,9} - “site-no-protein category”; those mRNAs with target sites where the synthesized protein could be detected using proteomics but where there was no support for reduced protein levels following miRNA transfection - “site-protein-no-support category”; and those where the level of the synthesized protein from the targeted mRNA changed following miRNA perturbation - “site-protein-support category.”

In the original analysis of the data, the miRNA target categories were non-orthogonal.¹² That is, miRNA targets belonging to the site-protein-no-support and the site-protein-support categories were included in the site-no-protein category. Similarly, miRNA targets belonging to the site-protein-support category were included in the site-protein-no-support category (Fig. 1A). This procedure made the comparisons between the partially overlapping categories difficult to interpret and invalidated their associated p-values. To avoid this complication in our re-analysis, we generated three orthogonal groups (Fig. 1B).

Guo et al. concluded that because target mRNAs whose expressed proteins were detected using proteomics (i.e., site-protein-no-support and site-protein-support categories) showed mRNA translation contributions similar to those of mRNAs with sites (i.e., the site-no-protein, site-protein-no-support and the site-protein-support category), the site-no-protein category added no additional information and was not evaluated in detail. Because of the overlapping categories, however, differences between target mRNAs that were or were not detected at the protein level could have been obscured (Fig. 1A–B).

To assess the contribution of mRNA translation Guo et al. calculated the mean difference between RPF and poly-A effects ($\log_2[\text{RPF}_{\text{miR}}] - \log_2[\text{RPF}_{\text{mock}}] - (\log_2[\text{polyA}_{\text{miR}}] - \log_2[\text{polyA}_{\text{mock}}])$) divided by the mean RPF effect ($\log_2[\text{RPF}_{\text{miR}}] - \log_2[\text{RPF}_{\text{mock}}]$). This is interpreted as the percentage of the observed miRNA mediated suppression that is caused by translational control. Without biological replication, however, the variability (random, technical and biological) of the estimates is unknown. We therefore used the two time points (12 and 32 h) from the mock condition in the HeLa cell model as proxies for replicates in a simulation to estimate the variability of the mean RPF effect, the mean poly-A effect, and the mean RPF effect corrected for the poly-A effect for each miRNA (miR-1 or miR-155), time point (12 or 32 h) and miRNA target site category (site-no-protein, site-protein-no-support or site-protein-support) (see Materials and Methods for details). We then compared target mRNAs that were detected at the protein level (site-protein-no-support and site-protein-support categories) to those that were

Figure 1. Analysis using non-overlapping categories differs from analysis with overlapping categories. (A) In the original analysis, the miRNA target site categories were not analyzed independently as mRNAs harbouring miRNA targets overlapped between the categories. (B) In our re-analysis the categories were created to allow an independent analysis of the three categories. (C) Percentage contribution from translation to observed miRNA mediated suppression between mRNAs carrying target sites that were not (site-no-protein) or were (protein-detected) detected by proteomics. P-values for KS and Wilcoxon rank sum tests are indicated.

not detected (site-no-protein category) by calculating the percentage of the regulation attributed to translation and the associated 95% empirical confidence intervals (from the simulations). This analysis identified a ~1.7 fold difference for the contribution of mRNA translation between mRNAs (with miRNA target sites) whose proteins were or were not detected (Fig. 1C, KS and Wilcoxon p-values < 2.2e-16) for both miR-1 and miR-155 at the 32 h time-point. The same analysis using data normalized to the no-site control generated similar results (KS and Wilcoxon p-values < 2.2e-16). The difference in results between our analysis and that of Guo et al. is therefore primarily due to the use of overlapping vs. non-overlapping categories and not to the normalization approach. This finding prompted re-analysis of the independent categories.

We performed detailed analysis of each of the non-overlapping categories using the simulation approach described above; the resulting densities in Figure S2 represent the variability of the mean effects. This analysis showed that the effects observed for the site-protein-no-support category were so small that further analysis was difficult to interpret. Although Guo et al. also excluded the site-protein-no-support category, the almost complete lack of regulation was not apparent when assessing overlapping categories (i.e., the combined site-protein-no-support and site-protein-support categories vs. the site-protein-support category). Similarly to Guo et al., we found that the site-protein-support

A			
% contribution from translation without no-site control normalization			
miR-1	12 hours	32 hours	n
site-no-protein	58 (43, 71)	31 (18, 43)	467
site-protein-support	21 (8, 32)	12 (3, 20)	99
miR-155	12 hours	32 hours	
site-no-protein	78 (26, 121)	31 (13, 47)	408
site-protein-support	58 (23, 86)	13 (-1, 25)	121
B			
% contribution from translation with no-site control normalization			
miR-1	12 hours	32 hours	n
site-no-protein	41 (25, 54)	29 (17, 41)	467
site-protein-support	13 (0, 24)	12 (3, 20)	99
miR-155	12 hours	32 hours	
site-no-protein	-6 (-101, 52)	36 (18, 51)	408
site-protein-support	18 (-24, 53)	16 (3, 28)	121

Figure 2. Relative contributions of translation and stability to miRNA-mediated suppression of gene expression depend on time and miRNA target site category. (A) Shown is the percentage contribution from translation without normalization to the no-site control to observed miRNA mediated suppression with associated 95% confidence interval at each condition. The number (n) of mRNAs in each category is indicated. (B) Same as (A) but using data normalized to the no-site control.

category showed robust effects at most time-points which was also the case for the site-no-protein category. We therefore continued our analysis using the site-no-protein and the site-protein-support categories only. Figure 2A summarizes the contribution of translation in each of the conditions by calculating the percentage of regulation attributed to translation (similar to Figure 1C). The miR-155 12 h conditions, in particular the site-no-protein category, notably showed substantially larger confidence intervals than all other conditions (Fig. 2A). This is a result of the smaller observed effects because the percentage contribution of translation can become unstable across simulations when one small effect is divided by another small effect (Fig. S2). The remaining conditions showed comparable variability as judged by the width of the confidence intervals (Fig. 2A). After excluding the unreliable miR-155 12 h condition, one of the two 12 h time point miRNA categories suggested that translation contributes at least as much as stability to miRNA mediated suppression ($\geq 50\%$), while at 32 h two of four conditions suggested substantial contribution of translation ($\geq 30\%$). The contribution of translation as presented in Figure 2A (21–58% at 12 h [after excluding the miR-155 condition] and 12–31% at 32 h, depending on category and miRNA) is substantially different from the 11–16% contribution suggested by Guo et al.¹²

We next assessed the impact of the Guo et al.'s no-site control normalization approach on these conclusions. We re-did the above

analyses using no-site control normalization and obtained effects with similar variability as judged by the width of the empirical confidence intervals (compare Figure 2A–B). The no-site control gene normalized data showed percentage contributions from translation of 13–41% after 12 h (after excluding the miR-155 condition) and 12–36% after 32 h (Fig. 2B), again substantially different from 11–16%. The relative contribution of translation under most conditions (5 out of 6, after excluding the miR-155 12 h condition) using the no-site control gene normalized effects were within the 95% confidence intervals from the analysis of non no-site control gene normalized data (Fig. 2A–B). This suggests that although the normalization approach to some extent affected mean-contributions of mRNA translation (compare Figure 2A–B), our analysis of independent non-overlapping categories had a larger impact on the results. Thus, by separating the miRNA categories into independent groups at different time points, different conclusions regarding the relative contribution of translation and stability for miRNA-mediated target suppression emerge.

We extended this analysis to the chronic miR-223 model, which could be mechanistically different from the acute HeLa model and which does not allow for temporal analysis. Due to the lack of replication, the analysis presented in Figure S2 for HeLa cells was not possible for the miR223 model because such analysis relies on empirical confidence intervals. Instead, we compared mRNA effect distributions rather than the category mean effects shown in Figure S2. In the miR-223 model, neither the site-no-protein nor the site-protein-no-support category showed any regulation (Fig. S4). The remaining 77 mRNAs belonging to the site-protein-support category showed small effects and indicated that stability contributed more than translation to observed miRNA mediated suppression. Notably, the distribution of the differences between the RPF and the poly-A effects overlapped substantially with the distribution of the poly-A effects and $> 50\%$ of the mRNAs showed larger poly-A effects than RPF effects ($> 40\%$ when the data had been normalized to the no-site controls). These aspects highlight the limitations of the miR-223 model.

Analysis of non-overlapping categories suggests that suppression of mRNA translation precedes degradation. There appeared to be a larger contribution of mRNA translation after 12 h compared with 32 h for miR-1 (Fig. 2A) which motivated us to reassess the conclusion of Guo et al. that translation was not more active at 12 as compared with 32 h. Indeed there was a significant reduction of the contribution of mRNA translation at 32 as compared with 12 h based on the simulated distributions for both the miR-1 site-no-protein (1.9-fold) and site-protein-support (1.8-fold) categories (all KS and Wilcoxon p-values $< 2.2e-16$). The same analysis using data normalized to the no-site control generated a similar result for the site-no-protein category (KS and Wilcoxon p-value $< 2.2e-16$) while the contribution of translation for the site-protein-support category was not further reduced from 13%. Thus in contrast to the conclusions of Guo et al., our reanalysis of non-overlapping categories suggests that suppression of mRNA translation precedes mRNA degradation for at least some of the miRNA target categories. The effect was also observed for the miR-155 data, although the 12 h time-point is based on effects too small to allow analysis (Fig. 2A). Although this conclusion may also be affected by

the difference score approach used to calculate translation effects (which is potentially associated with spurious correlations) it nevertheless illustrates the impact of analyzing non-overlapping categories.

Translation effects are potentially limited by the floor of the dynamic range. An important insight when evaluating this data set using the present approach is that for the translation effect to be, for example, equal to that of mRNA stability, the RPF effect must be twice the poly-A effect (i.e., a 2-fold larger effect for RPF data are needed to conclude a 50% relative contribution of translation compared with stability as calculated in **Figure 2A**). Therefore, the larger the poly-A effect, the larger the required dynamic range for the RPF data to conclude 50% relative contribution from translation. This aspect could become critical because the signals could approach the floor of the dynamic range and therefore be unable to decrease further (i.e., be unable to detect translation effects). Guo et al. used a heuristic cut off for mRNA inclusion of 100 sequence reads in the mock conditions to address this potential problem (both polyA and RPF). We sought an approach to evaluate if the contribution of translation, using this heuristic threshold, could have been underestimated due to lack of dynamic range for RPF data.

Variances often increase at the floor of the dynamic range of a log₂ scale due to the additivity of the noise close to the detection limit.²² Thus, by comparing the log₂ signals to the standard deviations one can assess if the data are approaching the floor of the dynamic range. We obtained the standard deviations from each experimental condition across the 12 and 32 h time points only from mRNAs that were not miRNA targets in each specific experiment. This represents a good estimate of what the standard deviation is across the signal range as thousands of mRNAs are included. We then examined the relationship between the expression level and the standard deviations for the site-no-protein and the site-protein-support categories at 32 h. **Figure 3A** shows that at 32 h both the site-no-protein and the site-protein-support categories had shifted to a position in the signal range for RPF data where the standard deviations increase rapidly, indicating that a substantial proportion of the RPF data are at or near the floor of the dynamic range. This was true for both the miR-1 and the miR-155 experiments but was not evident to the same extent for the poly-A data in part because poly-A data were less variable (this analysis was not possible for the miR-223 model due to the lack of replication). Because a gene with a large poly-A effect requires an even larger dynamic range for RPF data to show translation effects, a lack of dynamic range would lead to larger poly-A effects (i.e., more negative) being associated with lower percentage contribution from mRNA translation. We therefore compared the percentage contribution from mRNA translation to the poly-A effect (i.e., stability effect). Strikingly, larger poly-A effects (i.e., more negative) were associated with a lower contribution of mRNA translation (**Fig. 3B**; as expected, this relationship was also found when the data had been normalized to the no-site control [**Figure 3C**]). Moreover, this analysis suggests that the difference between the categories (site-no-protein and site-protein-support) in terms of relative contribution of mRNA translation (**Fig. 2A**) is that

more mRNAs in the site-protein-support category are associated with larger poly-A effects than in the site-no-protein category (**Fig. 3B**) and therefore potentially more biased toward low contribution of mRNA translation due to the lack of dynamic range. The finding is also consistent with the observation that suppression of mRNA translation precedes degradation of target mRNA in as much as suppression of mRNA translation is more pronounced among mRNAs that are not substantially degraded (**Fig. 3B**) and also that miRNA targets that enjoy more regulation at the level of stability are primarily targeted by stability at this time point. However, these hypotheses cannot be distinguished from effects due to a lack of dynamic range.

Mechanisms for suppression of single miRNA target-mRNAs are largely unknown. Our and the original analyses attempt to derive systems-wide effects of translation and stability by estimating miRNA target site category mean effects. **Figure 3B** suggests, however, that there might be considerable heterogeneity between mRNAs. To assess across-mRNA differential mechanistic involvement, a per-mRNA rather than a per-category analysis is needed. **Figure 4A** shows the per mRNA proportional contribution of translation relative to all regulation (calculated in a manner similar to that in **Figure 2A** but per mRNA instead of means across mRNAs) for each miRNA (miR-1 and miR-155), time point (12 and 32 h) and category (site-no-protein and site-protein-support). This analysis suggests large inter-mRNA variation for the relative contribution of translation and stability for all stratifications of the analysis (this was also the case when data had been normalized to the no-site control, **Fig. S5-a**). It is noticeable that the ratios sometimes are larger than 1 (indicating > 100% translation) and sometimes less than 0 (indicating > 100% stability) (similar results were obtained for miR-223 [**Fig. S5-b**]). A comparison of the per mRNA effects for poly-A and RPF data indicates that this measurement difficulty is common in the present data: poly-A effects are larger than RPF effects for 30–40% of the mRNAs (**Fig. S5-c**; > 50% mRNAs show larger poly-A effects compared with RPF effects for miR223 data [not shown]). These values are similar to those generated by the no-site control normalized data (miR-223 > 40% [not shown], miR-1 30–40% and miR-155 30–50% [**Fig. S5-d**]), indicating that normalization to no-site controls cannot overcome this issue. Another approach for per mRNA analysis is to compare the RPF effect to the doubled poly-A effect as this corresponds to the situation when the two mechanisms contribute equally to miRNA mediated suppression. Such analysis, using the same stratification as in **Figure 4a**, supports the conclusion that there are mRNAs which are primarily regulated either by translation or stability (**Fig. 4B**; similar results for miR-223 [**Fig. S5-e**]). As expected this was the case also when the data had been normalized to the no-site control (**Fig. S5-f**). These two approaches do not, however, take per mRNA variance into account – something that is needed when attempting to obtain mRNA specific conclusions regarding mechanisms. A per-mRNA linear regression provides a statistical model for analysis and indicates that it is unclear whether translation or stability is more important for miRNA mediated suppression for most individual mRNAs (**Fig. 4C**). Although

Figure 3. Translation effects are potentially limited by a lack of dynamic range for RPF data. **(A)** For each data type (RPF or poly-A), treatment (miR-1 or miR-155) and target site category (site-no-protein or site-protein-support) the density of signals at 32 h from mRNAs belonging to that subset is indicated (for this analysis we used mean centered counts (to 0) as signals to avoid obscuring the relationship between signal and standard deviation by differences in mRNA length). A line representing the lowest regression of the standard deviation (obtained from mRNAs not targeted by the miRNA across the two time points) to the signal is shown together with the empirical 95% confidence interval for the standard deviation as a function of the signal. The floor of the dynamic range is approached as standard deviations increase, and is related to signal intensities. **(B–C)** Relationship between poly-A effects and the percentage contribution from mRNA translation at 32 h stratified by miRNA site (miR-1 or miR-155) and category (site-no-protein or site-protein-support) using non-normalized **(B)** or no-site control normalized **(C)** effects. Genes passing filtering thresholds are shown (RPF effect $< [-0.5]$; relative contribution of mRNA translation between 0 and 100%; and poly-A effect < 0). The line represents a lowest regression.

normalization to the no-site control cannot be incorporated, applying this approach to the no-site control genes generated the expected number of significant findings under the null hypothesis (i.e., 5%, **Figure 4C**) indicating that the approach is reasonable. Thus at a single mRNA level, the mechanism of suppression cannot be deduced from the current data set. This likely results from the lack of biological replication leading to insufficient statistical power.

Discussion

We have presented data that calls for a re-assessment of the magnitude of the relative contribution of translation and stability to observed miRNA mediated suppression of gene expression. We show that the conclusion from Guo et al. that 84–89% of microRNA-induced suppression of gene expression is due to degradation of target mRNAs was influenced by how the miRNA

target categories were constructed and how the time points were analyzed. Moreover, both the Guo et al. analysis and our analysis are potentially influenced by lack of sufficient dynamic range for RPF data and the resulting likely underestimation of translation (Fig. 3A–B). Although the choice of normalization approach (no-site control normalization or not) influenced relative contribution from mRNA translation, discrepancies stemmed mostly from whether the analyses were performed on non-orthogonal (Guo et al.) or orthogonal (the present analysis) miRNA categories.

Our re-analysis also suggests that translational control may precede mRNA degradation in mammals (Fig. 2A), which is in agreement with a large proportion of the current empirical literature,^{3,15} a recent study applying the RPF technique in zebrafish²³ and detailed recent studies of kinetics.^{24,25} Although Guo et al. did not design their experiment to primarily assess regulation of single targets, it is important to note that the data are insufficient in most cases to make conclusions about the mechanisms of regulation at the single mRNA level (Figs. 3B–C and 4A–C). Possible heterogeneity or varying combinatorial utilization of mechanism (e.g., depending on miRNA identity, mRNA identity and biological context) therefore needs to be considered whenever an investigator intends to examine the impact of specific mRNA – miRNA interactions on gene expression. One aspect that also needs to be considered is the limitations of the experimental setup for conclusions on how single miRNA targets are regulated. A plausible model is that once the miRNA is introduced there is a probability that an mRNA with a target site is suppressed by translational control or RNA degradation that could be cumulative over time. If an mRNA is targeted by translational repression there is also a probability for that mRNA to be degraded which could also be cumulative over time. Assessing these probabilities will allow determination of miRNA-target specific roles of mRNA translation and stability for suppression of gene expression. Such studies will need to take kinetics into consideration and include more sequence reads to avoid problems relating to analyzing diminishing effects at the floor of the dynamic range. Replication is needed to obtain mRNA-by-mRNA estimates of random error for statistical tests and for analysis without confounding spurious correlations.¹³ Small replication such as triplicates enables analysis without spurious correlations, offers a substantial increase in measurement precision, and enables use of variance shrinkage methods to increase the power of statistical tests beyond what low replication normally provides.²⁶

Material and Methods

Data set preparation. Data were downloaded from the Gene Expression Omnibus (GEO) database (accession number GSE22004 and GSE22001). We used the data that were extracted and normalized by the authors of the original presentation using the Reads Per Kilobase of exon model per Million mapped reads (RPKM) approach. Similar to the original presentation we used

Figure 4. A per mRNA analysis of the relative contribution of translation and stability to miRNA-mediated suppression of gene expression is inconclusive. (A) Across mRNAs box plots of proportions of the regulation attributed to translation (similar interpretation as Figure 2A). The lines indicate (from top) 100%, 75%, 50% and 25% contribution from translation. (B) Box plots comparing the per mRNA contribution from translation and stability to miRNA suppression by subtracting twice the poly-A effect from the RPF effect. 0 (dotted line) indicates equal contribution of translation and stability. Negative values indicate more contribution from translation while positive values indicate more contribution of stability. (C) Per mRNA statistical analysis using linear models of mechanisms contributing to miRNA mediated suppression. The % mRNAs targeted primarily by translation or stability; or both translation and stability in the HeLa model stratified by treatment (miR-1 or miR-155) and miRNA target site category (no-site, site-no-protein or site-protein-support) is shown.

the log2(RPKM) values as signals. The miRNA target site category identities (“site-no-protein,” “site-protein-no-support” and “site-protein-support” as shown in Figure 1A–B) were also downloaded as part of GSE22004 and GSE22001.

Comparison of RNA features between no-site control genes and target categories. We annotated for all miRNAs that are expressed in HeLa cells to assess de-repression by comparing abundance of non-experimentally targeted miRNA sites.²¹ We used conserved sites from TargetScan for miRNA to mRNA

target site annotation^{5,27} and counted the number of non-experimentally targeted miRNA target sites per mRNA for expressed miRNAs. We used Wilcoxon's rank sum test to compare the group of no-site control genes to each of the target site categories for each miRNA transfection (miR-1 or miR-155) separately. ARE annotation was obtained from ARED²⁸ and we scored mRNAs dichotomously as either harboring ARE or not. Fisher's exact test was used to compare the proportion of ARE carrying mRNAs between the group of no-site control genes to each of the target site categories for each miRNA (miR-1 or miR-155) separately. We used data from the RefSeq database²⁹ for comparison of 3' UTR lengths. We used Student's t-test to compare the group of no-site control genes to each of the target site categories for each miRNA treatment (miR-1 or miR-155) separately.

Simulation of translation and stability effects. The two time points (i.e., 12 and 32 h following transfection) could not be used as replicates for estimating mRNA and condition-specific standard deviations because effects (for mRNAs targeted by miRNAs) often increased with time. Instead, we estimated mRNA-specific standard deviations from the mock conditions across the 12 and the 32 h time points (for RPF and poly-A data separately). We simulated 1000 signals for each mRNA in each of the 6 conditions (i.e., mock, miR-1 or miR-155 both 12 or 32 h post-transfection) by sampling from a normal distribution (rnorm function in R) using the empirical signal as a mean and each mRNA's standard deviation from the mock condition as standard deviation (RPF or poly-A according to the simulated condition). We calculated per mRNA and per simulation effects $[\log_2(\text{miR}) - \log_2(\text{mock})]$ for RPF data and poly-A data; and the difference between the RPF effects and the poly-A effects. We then generated per simulation means of the RPF effects, poly-A effects and the differences between RPF and poly-A effects, stratified on treatment (miR-1 or miR-155), time point (12 or 32 h) and miRNA target site category (site-no-protein, site-protein-no-support or site-protein-support). This generated 1000 mean effects per condition and effect type which were plotted as densities in Figure S2. Empirical 95% confidence intervals in Figure 2A–B were obtained from calculations of relative contributions from translation $(\log_2[\text{RPF}_{\text{miR}}] - \log_2[\text{RPF}_{\text{mock}}]) - (\log_2[\text{polyA}_{\text{miR}}] - \log_2[\text{polyA}_{\text{mock}}])$ divided by total regulation $\log_2[\text{RPF}_{\text{miR}}] - \log_2[\text{RPF}_{\text{mock}}]$ for each of the 1000 simulations.

Assessments of dynamic range. We estimated standard deviation (noise) to signal relationships for each treatment and RNA type (i.e., miR-1 and miR-155 for both RPF and poly-A data) to examine dynamic range (Fig. 3A). Because standard deviations of miRNA-targeted mRNAs potentially reflect both noise and treatment, standard deviations were estimated from non-targeted mRNAs across time for each treatment (miR-1 or miR-155) and RNA (RPF or poly-A). Note that this is different from using no-site control genes for normalization in that no-site control genes are used solely for assessing the signal to variance relationship for the experiment. We thereby obtained four standard deviation to signal relationships (RPF miR-1; RPF miR-155; poly-A miR-1; and poly-A miR-155). These were used as comparisons to the distributions of the signals 32 h after miRNA transfection (miR-1 or miR-155) for each RNA (RPF

or poly-A) and miRNA target site category (site-no-protein and site-protein-support) separately.

Per mRNA analysis using linear regression models. Linear regression models without replication are saturated, leaving no degrees of freedom for the error term when using the full model. A common solution is to obtain estimates of residual error from higher order interactions (in the present case, the 3-way interaction) under the assumption that they estimate error rather than effects. We used a full linear model (with RNA [RPF or poly-A], treatment [mock or miR] and time [12 or 32 h]); the 3-way interaction effects for individual mRNAs were evenly distributed roughly around 0 for these analyses, suggesting that they could provide reasonable estimates of error (Fig. S6). We then used a linear model without the 3-way interaction to calculate the 2-way interaction, and its associated confidence interval, between RNA (RPF or poly-A) and treatment (miR or mock) separately for each treatment (miR-1 or miR-155) and target site category (site-no-protein or site-protein-support). We tested the null hypothesis that translation and stability were equally active. We considered an mRNA specific null hypothesis derived from the poly-A effect $(\log_2[\text{miR}] - \log_2[\text{mock}])$ but rejected that possibility due to the lack of information about the error of that estimate per mRNA. Instead we used the per treatment (miR-1 or miR-155) and target site category (site-no-protein or site-protein-support) median poly-A effect as the null hypothesis. We then classified mRNAs as primarily targeted by translation (when the confidence interval for the RNA x treatment interaction was negative compared with the null hypothesis) or stability (when the confidence interval for the RNA x treatment interaction was positive compared with the null hypothesis). There was insufficient evidence to reject the null hypothesis for the majority of the mRNAs, indicating that conclusions about the relative importance of translation and stability could not be made for these mRNAs. The percentages belonging to each class for each treatment (miR-1 or miR-155) and miRNA target site category (no-site, site-no-protein or site-protein-support) are shown in Figure 4C.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We would like to thank Nahum Sonenberg (McGill University) for suggestions and comments on both the analysis and the manuscript. We would also like to thank Andrew Fernandes (The University of Western Ontario), David Rocke (University of California, Davis), Rickard Sandberg (Karolinska Institute), Ivan Topisirovic (McGill University) and Kristian Wennmalm (Karolinska Institute) for insightful comments on the manuscript. This work was supported by the Swedish Research Council; the Jeansson Foundation; the Åke-Wiberg Foundation; a Marie-Curie international reintegration grant; the Swedish Cancer Foundation; and the Swedish Cancer Society in Stockholm (all to O.L.).

Supplemental Material

Supplemental material may be found here:
www.landesbioscience.com/journals/translation/article/24557

References

- Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell* 2009; 136:215-33; PMID:19167326; <http://dx.doi.org/10.1016/j.cell.2009.01.002>.
- Fabian MR, Sonenberg N, Filipowicz W. Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem* 2010; 79:351-79; PMID:20533884; <http://dx.doi.org/10.1146/annurev-biochem-060308-103103>.
- Djuranovic S, Nahvi A, Green R. A parsimonious model for gene regulation by miRNAs. *Science* 2011; 331:550-3; PMID:21292970; <http://dx.doi.org/10.1126/science.1191138>.
- Lim LP, Lau NC, Garrett-Engle P, Grimson A, Schelter JM, Castle J, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 2005; 433:769-73; PMID:15685193; <http://dx.doi.org/10.1038/nature03315>.
- Grimson A, Farh KK, Johnston WK, Garrett-Engle P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 2007; 27:91-105; PMID:17612493; <http://dx.doi.org/10.1016/j.molcel.2007.06.017>.
- Nielsen CB, Shomron N, Sandberg R, Hornstein E, Kitzman J, Burge CB. Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA* 2007; 13:1894-910; PMID:17872505; <http://dx.doi.org/10.1261/rna.768207>.
- Krützfeldt J, Rajewsky N, Braich R, Rajeev KG, Tuschl T, Manoharan M, et al. Silencing of microRNAs in vivo with 'antagomirs'. *Nature* 2005; 438:685-9; PMID:16258535; <http://dx.doi.org/10.1038/nature04303>.
- Back D, Villén J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. *Nature* 2008; 455:64-71; PMID:18668037; <http://dx.doi.org/10.1038/nature07242>.
- Selbach M, Schwanhäusser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. Widespread changes in protein synthesis induced by microRNAs. *Nature* 2008; 455:58-63; PMID:18668040; <http://dx.doi.org/10.1038/nature07228>.
- Hendrickson DG, Hogan DJ, McCullough HL, Myers JW, Herschlag D, Ferrell JE, et al. Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. *PLoS Biol* 2009; 7:e1000238; PMID:19901979; <http://dx.doi.org/10.1371/journal.pbio.1000238>.
- Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 2009; 324:218-23; PMID:19213877; <http://dx.doi.org/10.1126/science.1168978>.
- Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 2010; 466:835-40; PMID:20703300; <http://dx.doi.org/10.1038/nature09267>.
- Larsson O, Sonenberg N, Nadon R. Identification of differential translation in genome wide studies. *Proc Natl Acad Sci U S A* 2010; 107:21487-92; PMID:21115840; <http://dx.doi.org/10.1073/pnas.1006821107>.
- Khan AA, Betel D, Miller ML, Sander C, Leslie CS, Marks DS. Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. *Nat Biotechnol* 2009; 27:549-55; PMID:19465925; <http://dx.doi.org/10.1038/nbt0709-671a>.
- Filipowicz W, Bhattacharyya SN, Sonenberg N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* 2008; 9:102-14; PMID:18197166; <http://dx.doi.org/10.1038/nrg2290>.
- Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 2010; 141:129-41; PMID:20371350; <http://dx.doi.org/10.1016/j.cell.2010.03.009>.
- Qin LX, Kerr KE. Empirical evaluation of data transformations and ranking statistics for microarray analysis (vol 32, pg 5471, 2004). *Nucleic Acids Res* 2004; 32.
- Ritchie ME, Silver J, Oshlack A, Holmes M, Diyagama D, Holloway A, et al. A comparison of background correction methods for two-colour microarrays. *Bioinformatics* 2007; 23:2700-7; PMID:17720982; <http://dx.doi.org/10.1093/bioinformatics/btm412>.
- Wu ZJ, Irizarry RA. Preprocessing of oligonucleotide array data. *Nat Biotechnol* 2004; 22:656-8, author reply 658; PMID:15175677; <http://dx.doi.org/10.1038/nbt0604-656b>.
- Lee PD, Sladek R, Greenwood CM, Hudson TJ. Control genes and variability: absence of ubiquitous reference transcripts in diverse mammalian expression studies. *Genome Res* 2002; 12:292-7; PMID:11827948; <http://dx.doi.org/10.1101/gr.217802>.
- Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* 2007; 129:1401-14; PMID:17604727; <http://dx.doi.org/10.1016/j.cell.2007.04.040>.
- Rocke DM. Design and analysis of experiments with high throughput biological assay data. *Semin Cell Dev Biol* 2004; 15:703-13; PMID:15561590.
- Bazzini AA, Lee MT, Giraldez AJ. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* 2012; 336:233-7; PMID:22422859; <http://dx.doi.org/10.1126/science.1215704>.
- Djuranovic S, Nahvi A, Green R. miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science* 2012; 336:237-40; PMID:22499947; <http://dx.doi.org/10.1126/science.1215691>.
- Béthune J, Artus-Revel CG, Filipowicz W. Kinetic analysis reveals successive steps leading to miRNA-mediated silencing in mammalian cells. *EMBO Rep* 2012; 13:716-23; PMID:22677978; <http://dx.doi.org/10.1038/embor.2012.82>.
- Malo N, Hanley JA, Cerquozzi S, Pelletier J, Nadon R. Statistical practice in high-throughput screening data analysis. *Nat Biotechnol* 2006; 24:167-75; PMID:16465162; <http://dx.doi.org/10.1038/nbt1186>.
- Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005; 120:15-20; PMID:15652477; <http://dx.doi.org/10.1016/j.cell.2004.12.035>.
- Halees AS, El-Badrawi R, Khabar KS. ARED Organism: expansion of ARED reveals AU-rich element cluster variations between human and mouse. *Nucleic Acids Res* 2008; 36(Database issue):D137-40; PMID:17984078; <http://dx.doi.org/10.1093/nar/gkm959>.
- Pruitt KD, Tatusova T, Klimke W, Maglott DR. NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res* 2009; 37(Database issue):D32-6; PMID:18927115; <http://dx.doi.org/10.1093/nar/gkn721>.