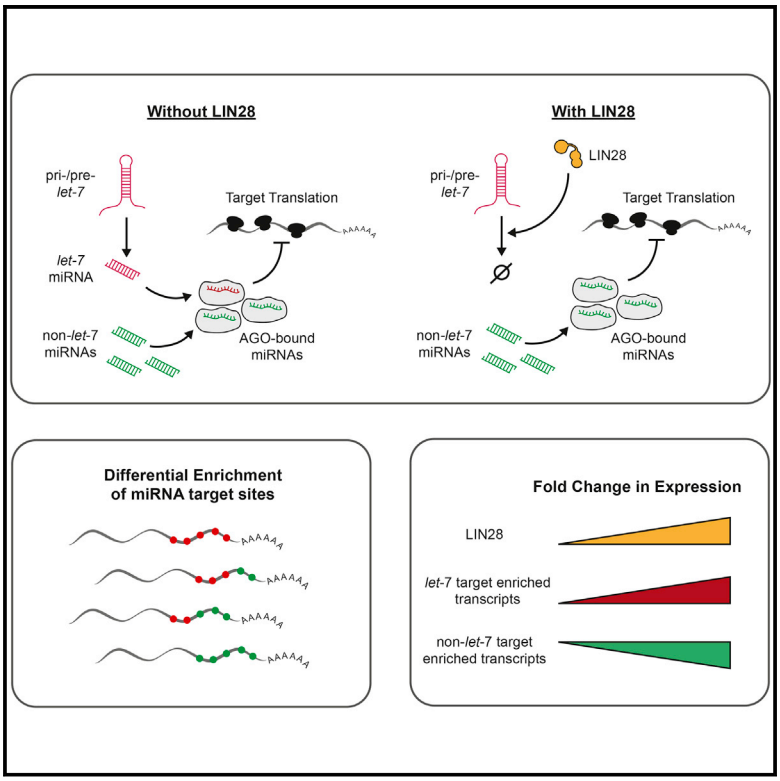
Article

Molecular Cell

# A Transcriptome-wide Translational Program Defined by LIN28B Expression Level

## Graphical Abstract




## Authors

Frederick E. Tan, Shashank Sathe, Emily C. Wheeler, Julia K. Nussbacher, Samson Peter, Gene W. Yeo

## Correspondence

geneyeo@ucsd.edu

## In Brief

Tan et al. show that changes in LIN28B expression can upregulate and downregulate gene expression. By suppressing *let*-7 family miRNA biogenesis, LIN28B liberates Argonaute to bind non-*let*-7 miRNA families with greater frequency, causing downstream changes because of redistributed miRNA activity. These effects affect a significant portion of the transcriptome.

## Highlights

- LIN28B and Argonaute are competitively bound by their RNA targets
- Blockade of *let*-7 biogenesis by LIN28 enhances the activity of non-*let*-7 miRNAs
- Post-transcriptional regulation varies with transcript-specific enrichment of miRNA sites
- Differences in miRNA site enrichment are conserved and predict LIN28 regulation

Tan et al., 2019, Molecular Cell *73*, 1–10
January 17, 2019 
https://doi.org/10.1016/j.molcel.2018.10.041

CellPress

Please cite this article in press as: Tan et al., A Transcriptome-wide Translational Program Defined by LIN28B Expression Level, Molecular Cell (2018), https://doi.org/10.1016/j.molcel.2018.10.041

Molecular Cell
Article



# A Transcriptome-wide Translational Program Defined by LIN28B Expression Level

Frederick E. Tan,[1,2,3,4] Shashank Sathe,[1,2,3,4] Emily C. Wheeler,[1,2,3] Julia K. Nussbacher,[1,2,3] Samson Peter,[1,2,3] and Gene W. Yeo[1,2,3,5,*]

[1]Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA, USA
[2]Stem Cell Program, University of California, San Diego, La Jolla, CA, USA
[3]Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA, USA
[4]These authors contributed equally
[5]Lead Contact
*Correspondence: geneyeo@ucsd.edu
https://doi.org/10.1016/j.molcel.2018.10.041

## SUMMARY

**LIN28 RNA binding proteins are dynamically expressed throughout mammalian development and during disease. However, it remains unclear how changes in LIN28 expression define patterns of post-transcriptional gene regulation. Here we show that LIN28 expression level is a key variable that sets the magnitude of protein translation. By systematically varying LIN28B protein levels in human cells, we discovered a dose-dependent divergence in transcriptome-wide ribosome occupancy that enabled the formation of two discrete translational subpopulations composed of nearly all expressed genes. This bifurcation in gene expression was mediated by a redistribution in Argonaute association, from *let*-7 to non-*let*-7 microRNA families, resulting in a global shift in cellular miRNA activity. Post-transcriptional effects were scaled across the physiological LIN28 expression range. Together, these data highlight the central importance of RBP expression level and its ability to encode regulation.**

## INTRODUCTION

The expression level of RNA-binding proteins (RBPs) is a fundamental variable in post-transcriptional control because binding sites are highly degenerate and often present in excess (Singh and Valcárcel, 2005). Competition for RBP binding may exist, and changes in RBP expression level can have a strong effect on target regulation. However, the extent to which changes in RBP expression level account for changes in post-transcriptional regulation remains unclear and may have unique readouts for each protein.

Here we examine the effect of varying LIN28B's expression level on gene regulation. LIN28B is an RBP critical for development and is implicated in human disease. Dynamic changes in LIN28B expression are observed throughout prenatal and perinatal mammalian development, becoming fully downregulated in cells as they terminally differentiate (Yang and Moss, 2003). Deficiencies in LIN28B or its paralog LIN28A lead to severe developmental abnormalities that result from a reduced capacity for growth and regeneration in numerous organ systems (Shinoda et al., 2013; Shyh-Chang and Daley, 2013; Zhu et al., 2011). In contrast, increasing LIN28 expression has the opposite effect because it enhances tissue growth and repair and, for some cell types, a reversion to fetal-like gene expression signatures (Copley et al., 2013; Shyh-Chang et al., 2013; Yuan et al., 2012). This is best exemplified during cellular reprogramming, where forced LIN28 expression helps transition differentiated cells to pluripotency (Yu et al., 2007; Zhang et al., 2016). LIN28 proteins are also frequently re-activated and variably upregulated in many growth-related pathologies, such as cancer (Zhou et al., 2013). These observations indicate that LIN28 proteins have dose-dependent effects on development and disease phenotypes.

Changes in LIN28 expression level can influence gene expression through direct and indirect mechanisms (Figure 1A). LIN28 proteins are best known for their ability to suppress microRNA (miRNA) biogenesis. A key target of this regulation is the *let*-7 miRNA family (Viswanathan et al., 2008), a highly conserved regulator of growth, differentiation, and metabolism (Roush and Slack, 2008). Regulation of miRNA biogenesis involves the binding of terminal pri- or pre-miRNA hairpin loops to block Drosha- and Dicer-mediated processing as well as enable 3′ end oligouridylation to promote pre-miRNA degradation (Heo et al., 2009; Loughlin et al., 2011; Nam et al., 2011; Van Wynsberghe et al., 2011). Post-transcriptional regulation can also be mediated independent of miRNAs through LIN28's direct binding to mRNA targets (Cho et al., 2012; Graf et al., 2013; Hafner et al., 2013; Madison et al., 2013; Wilbert et al., 2012). LIN28 binds a significant fraction of transcripts in several profiled cell types, with binding largely localized to coding and 3′ UTR sequences. However, the functional significance of this direct binding remains unclear (Madison et al., 2013). How these molecular interactions set the dynamic range of LIN28 regulation remains unexplored.

Here we show that LIN28B influences gene expression by mediating transcriptome-wide changes in miRNA pathway activity. Rate-limiting expression of LIN28B and Argonaute ensures that they are competitively bound by their RNA targets. As a result, selective depletion of the abundant *let*-7 miRNA family leads to a shift in cellular miRNA activity to all other miRNAs

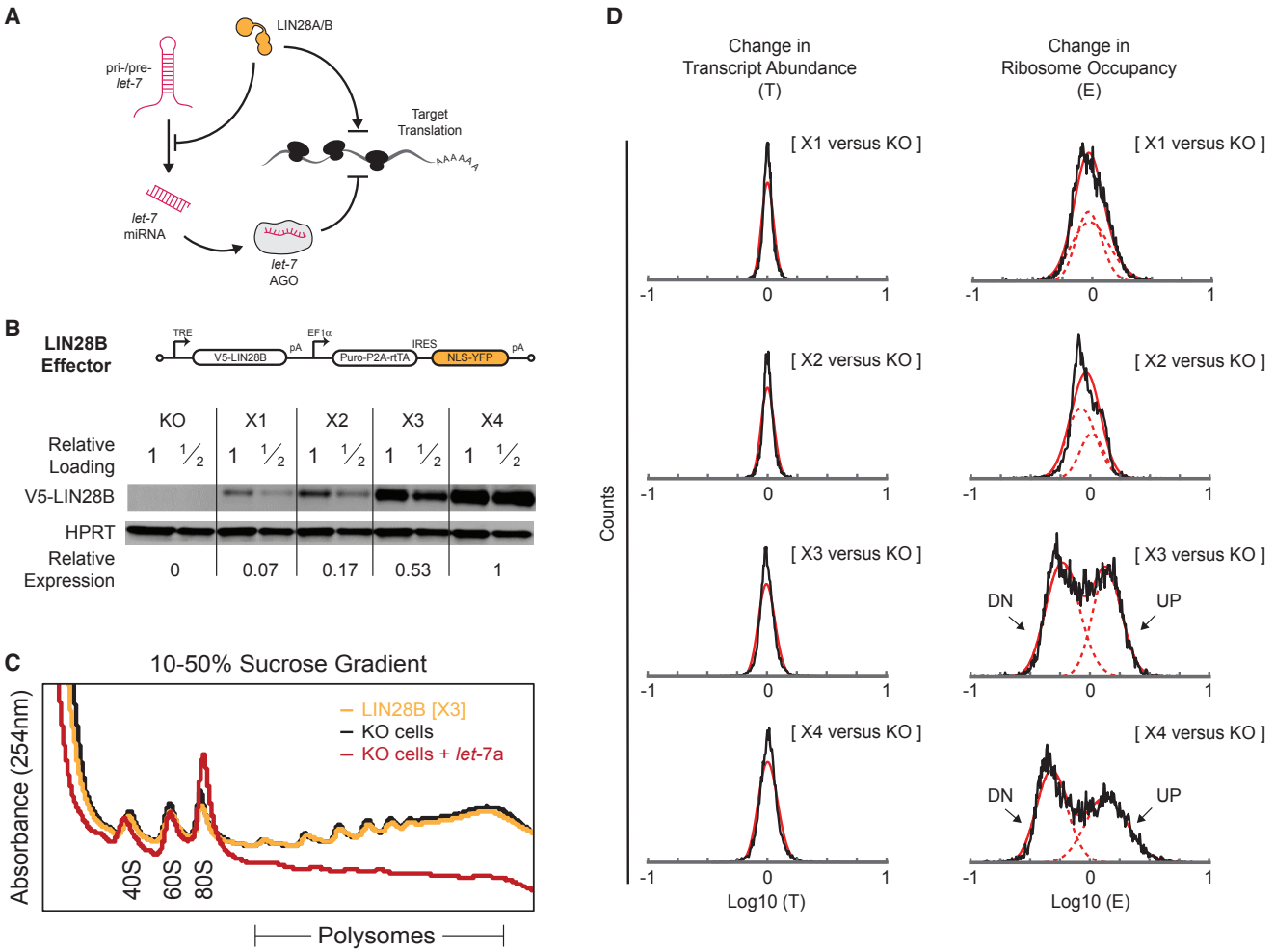



**Figure 1. LIN28B Expression Separates All Genes into Two Translational Subpopulations**

(A) Cartoon depicting the current model of LIN28 regulation. LIN28 regulates protein translation by controlling *let*-7 family miRNA biogenesis and directly binding target mRNA transcripts.

(B) Using yellow fluorescent protein (YFP) fluorescence, four cell populations were derived ([X1], [X2], [X3], and [X4]) for bulk profiling studies. Each population contains a different number of stably integrated PiggyBAC transgenes, which allow LIN28B to be expressed at different levels after Doxycycline (DOX) induction (10 ng/mL).

(C) Cellular lysates separated on a 10%–50% sucrose gradient were fractionated and resolved at 254 nm. The profile of ribosome-bound transcripts for LIN28B KO cells (KO, black), LIN28B-expressing cells ([X3], yellow), and KO cells transfected with mature *let*-7a miRNA at 10 nM (KO + *let*-7a, red) are overlaid.

(D) Changes in transcript abundance (T) or ribosome occupancy (E) for each gene in each LIN28B expression condition are presented relative to their value in LIN28B KO cells. For each transcript abundance distribution, a solid red line represents a single gamma distribution fit to sample data. For each (E) distribution, dashed red lines approximate the shape of UP and DN subpopulations as single gamma distributions, and solid red lines represent their sum, as a best fit to the sample data. These distributions show changes in the regulation of any gene of more than 1 reads per kilobase transcripts, per million mapped reads (RPKM) (N = 11,342 genes).

present in a cell that are unaffected by LIN28B. These changes in regulation underlie a separation of human transcripts into discrete translational subpopulations.

## RESULTS

### Transcriptome-wide Bifurcation in Ribosome Occupancy

To develop a cellular system that could achieve the physiological spectrum of LIN28 expression states, we stably integrated an inducible LIN28B transgene cassette into a human LIN28B knockout cell line (HEK293A; LIN28B knockout [KO] cells) using PiggyBAC shuttle vectors (Figure 1B). Transient transfection of PiggyBAC vectors with transposase permits variable copy number integration among single cells, which can yield stable differences in LIN28B protein expression upon doxycycline induction (Figures S1A and S1B). From this heterogeneous cell culture, we derived four populations expressing different LIN28B levels (sorted based on yellow fluorescent protein [YFP] fluorescence, named [X1] through [X4]) spanning the physiological range of

expression, which we used for bulk profiling experiments (Figure 1B; Figure S1C).

We first examined global alterations in ribosome occupancy via sucrose gradient fractionation. Surprisingly, even though changes in translation were expected, we did not find any significant difference in bulk poly-ribosome distribution in response to LIN28B expression (Figure 1C). By contrast, transfection of mature *let*-7a miRNAs at 10 nM into KO cells significantly reduced occupancy in poly-ribosome fractions (Figure 1C). To examine gene-specific changes in regulation, we prepared sequencing libraries from total and poly-ribosome fraction-derived RNA to measure changes in RNA abundance and changes in ribosome occupancy (STAR Methods). Changes in transcript abundance (T) were limited, approximately 2-fold or less with increasing LIN28B expression (Figure 1D). In contrast, the distribution showing changes in occupancy within poly-ribosome fractions (E) was dramatically altered (Figure 1D). Nearly all expressed transcripts bifurcated into translationally upregulated (UP) or downregulated (DN) subpopulations after normalization to External RNA Controls Consortium (ERCC) spike-in controls, indicating the existence of progressive and widespread effects on translation.

We validated changes in regulation using qRT-PCR for three genes (*EEF2*, *HMGA2*, and *SRP9*), which were chosen based on the characteristics of their regulation (Figure S2). From our high-throughput sequencing studies, *EEF2* and *SRP9* are translationally DN and UP by LIN28B, respectively, but exhibit very little change in transcript abundance. By contrast, the known *let*-7 target gene *HMGA2* changes in abundance, but it is less affected translationally because it is located between the DN and UP subpopulations. qRT-PCR measurements using unfractionated total RNA samples corroborate observed changes in transcript abundance in response to LIN28B expression (Figure S2C). qRT-PCR measurements using RNA derived from individual sucrose gradient fractions also corroborate changes in overall poly-ribosome occupancy (Figures S2D and S2E). A neutral redistribution of *HMGA2* transcripts occurs among polysome fractions (9 through 20). By contrast, transcripts for *EEF2* and *SRP9* become depleted and enriched in polysome fractions, respectively.

The number of genes belonging to each subpopulation was comparable, with 3,864 genes showing upregulation (E > 1) and 6,129 genes showing downregulation (E < 1). This balanced bifurcation in ribosome occupancy could explain why the bulk distribution of ribosome-bound mRNAs remained largely unchanged (Figure 1C). Importantly, the existence of a sizeable DN subpopulation was unexpected and inconsistent with much of LIN28's regulatory biology (Figure 1A). We therefore sought mechanisms that could mediate these dose-dependent and transcriptome-wide effects.

## LIN28B Alters the Composition of Argonaute-Bound miRNAs

LIN28 directly represses the biogenesis of specific miRNA families. To examine the role of miRNA regulation in the transcriptome-wide bifurcation of ribosome occupancies, we calculated, for each gene, the relative abundance of sites for each miRNA family using TargetScan predictions (conserved and non-conserved sites; 6- to 8-nt seeds) (Agarwal et al., 2015). We rank-ordered genes according to their change in ribosome occupancy in E[X4 versus KO], a distribution where discrete subpopulations are observed, and discovered a similar transcriptome-wide divergence in potential miRNA targeting (Figure 2A). Based on small RNA sequencing of HEK293A cells, LIN28B-regulated miRNA families include *let*-7 miRNAs and miR-98. Expectedly, there was a high correlation between increased ribosome occupancy and the presence of *let*-7 and miR-98 sites (Figure 2A, pink and red lines). Unexpectedly, ribosome occupancy was negatively correlated with sites for all other "non-*let*-7" family miRNAs (Figure 2A). The overrepresentation of non-*let*-7 miRNA sites among DN genes may reflect an increase in their activity in response to LIN28B expression. For this analysis, we considered the enrichment of all non-*let*-7 miRNAs together, even though LIN28 proteins are known to regulate the biogenesis of a few non-*let*-7 miRNAs (Peters et al., 2016; Warrander et al., 2016).

A potential mechanism that could mediate a global upregulation of non-*let*-7 miRNA activity is if, like LIN28B, Argonaute (AGO) protein levels are rate-limiting in post-transcriptional regulation. When mature miRNAs are present in excess, suppressing the biogenesis of abundant *let*-7 and miR-98 shifts AGO occupancy to the remaining non-*let*-7 miRNA species present in the cell. The magnitude of this regulation would scale with LIN28B and AGO expression as well as the relative expression level of *let*-7 and miR-98.

As measured by small RNA sequencing with spike-in controls, *let*-7 and miR-98 account for ~10% of mature miRNA expression in LIN28B KO cells. Expressing LIN28B to the level [X3] reduced their relative abundance to less than 1% of mature miRNA expression, which allowed for an ~1.2-fold increase in the apparent availability of non-*let*-7 miRNA species (Figure 2B). Importantly, this fold increase in availability occurred with very little disruption of the absolute abundance of non-*let*-7 miRNAs (Figure 2B). We discovered that changes in the composition of miRNAs co-precipitating with AGO2 reflected the change in relative miRNA abundance, not absolute abundance, which is consistent with miRNA regulation in the presence of limiting quantities of AGO protein (Figure 2B).

We further tested this model by increasing AGO levels, which should enhance post-transcriptional regulation by allowing additional miRNA to be bound but also reduces the magnitude of LIN28-dependent regulation. We constructed cell lines that expressed a fixed additional amount of AGO2 (KO + AGO2 and X4 + AGO2), which augmented expression by ~1.6-fold (Figure 2C). Total and poly-ribosome fraction-derived RNA were prepared, similar as before, and changes in mRNA abundance and ribosome occupancy were measured. We found that increasing AGO2 expression globally reduced the abundance of transcripts bound as polysomes (Figure 2D). Moreover, the presence of additional AGO2 diminished the effect of LIN28B at expression level X4 (Figure 2E).

We constructed fluorescent reporters to quantify changes in miRNA activity and correlated the response of each reporter to increasing LIN28B expression (Figure 2F). These reporters contained a fully complementary miRNA binding site in their 3′ UTR. We found that increasing LIN28B levels enhanced the expression of reporters with complementary *let*-7a or miR-98 target sites. Consistent with an increase in the activity of non-*let*-7

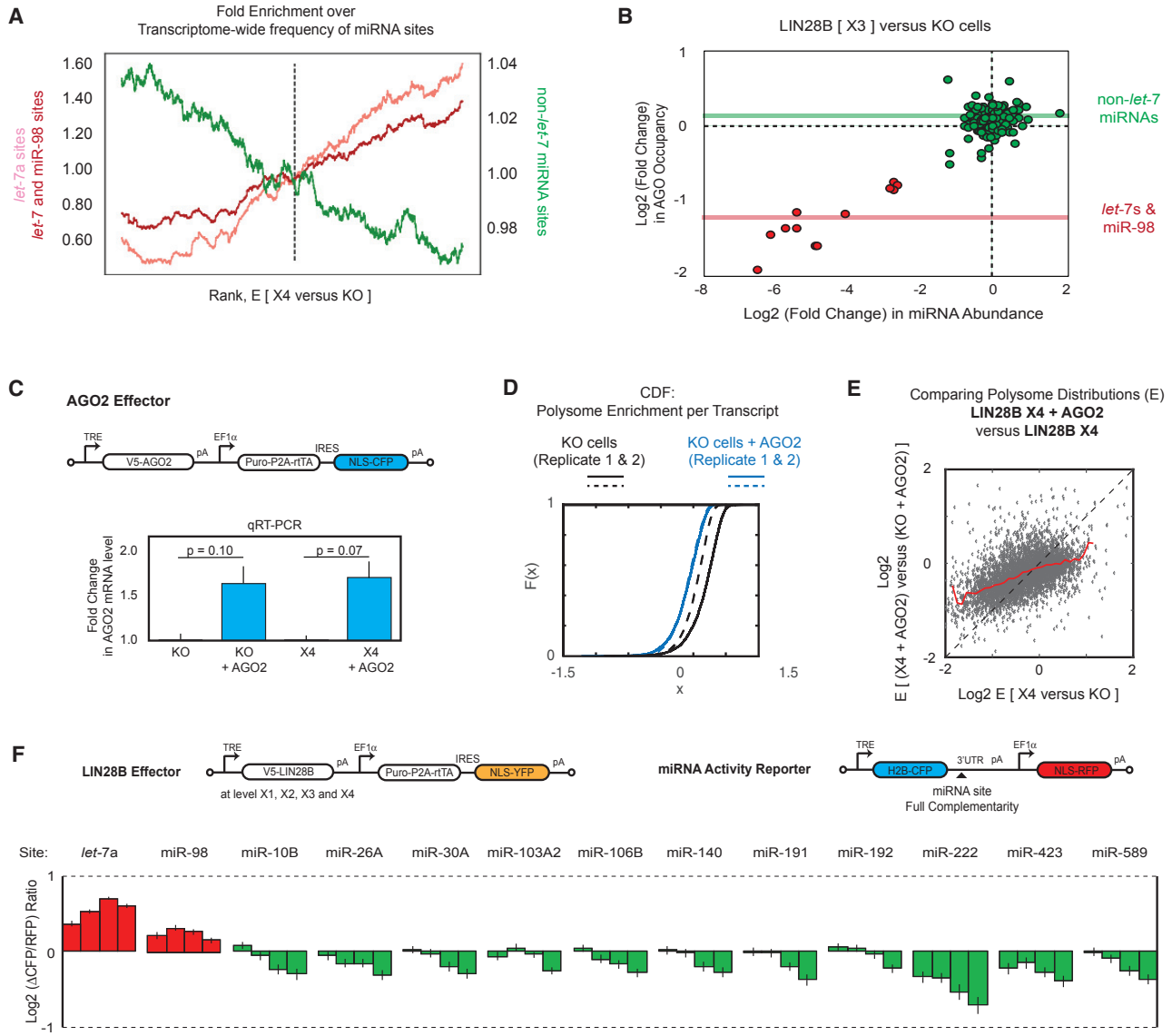



**Figure 2. LIN28B Shifts AGO Activity from *let-7* to Non-*let-7* miRNA Families**

(A) Relationship between miRNA site enrichment and the response to LIN28B expression (at level [X4]). Trends map out the enrichment of specific miRNA target sites per transcript relative to their average frequency among genes, within a moving window of 1,000 genes, from lowest to highest ribosome occupancy in E[X4 versus KO] (N = 11,342 genes). Highlighted are the trends for *let-7a* sites only (pink), all *let-7* and miR-98 sites (red), and all other unregulated miRNA families (green). The position corresponding to median ribosome occupancy is denoted (vertical black dashed line).

(B) The fold change in AGO2 occupancy resulting from LIN28B expression [X3] plotted against the fold change in miRNA abundance resulting from LIN28B expression [X3]. Non-*let-7* miRNA species (green) account for a greater percentage of total expressed miRNAs after the suppression of *let-7* and miR-98 (red) biogenesis by LIN28B (level [X3]). Solid red and green lines reflect the average fold change in *let-7* and non-*let-7* AGO occupancy after LIN28B expression (level [X3]).

(C) AGO2 effector construct and qRT-PCR validation of fold change in *AGO2* expression in LIN28B KO cells and LIN28B [X4] cells (±SEM, n = 4).

(D) Cumulative distribution of polysome enrichment per transcript in KO cells and KO + AGO2 cells.

(E) Increasing AGO2 expression suppresses LIN28B-dependent changes in ribosome occupancy. Ribosome occupancy distributions for LIN28B E[X4 versus KO] in the presence and absence of additional AGO2 are compared. The red line plots the average trend in (X,Y) values between these datasets.

(F) Changes in miRNA activity were profiled using fluorescent reporters. Post-transcriptional regulation was measured as changes in cyan fluorescent protein (CFP) fluorescence relative to unregulated RFP fluorescence (ΔCFP/RFP). These reporters were transfected into LIN28B expression cell lines (KO, [X1], [X2], [X3], and [X4]), and the ΔCFP/RFP ratio was computed relative to KO cells. These data are shown as a set of bar plots for each miRNA family (bars within each set from left to right: [X1] versus KO, [X2] versus KO, [X3] versus KO, and [X4] versus KO; ±SE of cells).

miRNAs, the expression of reporters with complementary binding sites for non-*let*-7 miRNA families were reduced. Changes in reporter expression were gradual, indicating that a global redistribution of miRNA targeting responded sensitively to changes in LIN28B level. Regulatory effects of greater magnitude are expected for endogenous transcripts because native 3′ UTRs are targeted by multiple miRNAs simultaneously.

We also show that regulation of miRNA target sites by this mechanism is not saturated by instead transfecting mature *let*-7a (at 10 nM) into LIN28B KO cells, which mimics knockdown of LIN28B expression (Figure S3A). Transfection of *let*-7a increases the relative abundance of *let*-7a from 0.2% to 3.1% of total mature miRNA expression in HEK293A cells. Transcripts more enriched for *let*-7a target sites (predominantly found in the UP subpopulation in E[X4 versus KO]) are now found to be DN, and transcripts more enriched for non-*let*-7 miRNA target sites (predominantly found in the DN subpopulation in E[X4 versus KO]) are now found to be UP (Figure S3A). The correlation between ribosome occupancy and miRNA site enrichment is similarly reversed (Figure S3B).

These data indicate that LIN28B expression affects miRNA activity both positively and negatively, which may explain why changes in mRNA abundance are limited. Nonetheless, changes in abundance are strongly correlated with changes in translation efficiency (Figure S3C), indicating that these post-transcriptional effects cooperate to modify gene expression. By contrast, regulation by *let*-7 alone cannot reproduce the observed changes in regulation because its activities mainly destabilize and translationally repress its direct targets (Figures S3D and S3E). Together, these data further underscore the significance of non-*let*-7 miRNA activity in defining the characteristics of LIN28B regulation.

## LIN28B Regulation via Opposing Post-Transcriptional Effects

Coordinated changes in miRNA activity thus allow LIN28B to both support and antagonize translation. By directly suppressing the biogenesis of *let*-7 family miRNAs, LIN28B frees AGO to bind non-*let*-7 miRNAs, leading to a transcriptome-wide shift in miRNA activity (Figures 3A and 3B). The magnitude of this post-transcriptional effect is constrained by the level of LIN28B, AGO, and *let*-7 miRNA expression in a cell.

Interestingly, the UP and DN subpopulations can be clearly identified in a principal component space that visualizes genome-wide patterns of miRNA site enrichment per gene (Figure 3C; Figure S4A). In this visualization, the first principal component (PC1) reflects the number of different miRNA families that potentially regulate a gene, which is a measure of cross-regulation between miRNA families (Figure S4B). The second principal component (PC2) maps out the segregation in frequency between *let*-7 and non-*let*-7 miRNA target sites (Figures 3C and 3D). The clear delineation between UP and DN genes indicates that each transcript exhibits specific patterns of miRNA site enrichment that ensures the dominance of either *let*-7 or non-*let*-7 miRNA effects.

These disparities in miRNA site enrichment are a signature of LIN28 regulation. Given the central importance of LIN28 in animal development, we wondered whether the same disparities were also selected for in other species. Using TargetScan predictions (conserved and non-conserved sites, 6- to 8-nt seeds) (Agarwal et al., 2015), we discovered that differences in the enrichment of *let*-7 versus non-*let*-7 miRNA target sites are broadly conserved (Figure 3E; Figure S4C). We believe that such differences potentially reflect and can be used to predict LIN28-dependent post-transcriptional control.

## Accurate Prediction of LIN28B Regulation

To determine whether we have captured the salient features of LIN28 regulation and to simultaneously test the evolutionary conservation of the above mechanism, we built machine learning classifiers. These classifiers were designed to examine the per-gene enrichment of miRNA sites within human datasets and look for similar patterns of enrichment in different genomes. For these classifiers, enrichment statistics from the most extremely regulated UP and DN transcripts from HEK293A cells were used for training (Figure 4A), and their accuracy in reevaluating post-transcriptional regulation was validated (STAR Methods). Classifiers that performed well in predicting regulation in human cells were then used to process a test gene set from a different mammalian species for *de novo* predictions (Figure 4A). Successful inter-species classification and validation would indicate regulatory conservation.

Classifiers trained on miRNA site enrichment statistics were able to correctly re-assign post-transcriptional regulation in human cells and exhibited high prediction accuracies (Figures 4B and 4C). This remained true even when enrichment patterns from any LIN28B expression condition were considered (Figure 4D). The reassignment accuracy for a validation gene set varied between 78% to 97%. As expected, classifiers performed poorly when gene identifiers in the training set were shuffled or when classifiers were trained on datasets not subject to LIN28B regulation (i.e., polysome distribution in KO cells) (Figure 4D, right, bar plots). These results indicated that classifiers trained on human data were suitable for *de novo* predictions in other species.

We next generated predictions for mouse coding transcripts and performed an identical set of LIN28B perturbations using immortalized mouse fibroblasts (NIH 3T3) to experimentally validate these predictions. Sequencing libraries prepared from total and poly-ribosome fraction-derived RNA revealed widespread changes in gene regulation resulting from LIN28B expression (Figure 5A). Changes in transcript abundance were more prominent in mouse NIH 3T3 compared with human HEK293A cells. The distribution of poly-ribosome occupancies was also bimodal. UP and DN subpopulations were resolved, although not as well separated (Figure 5A). The smaller separation distance between translational subpopulations in NIH 3T3 cells may be due to the less significant difference in *let*-7 miRNA site enrichment between UP and DN genes (Figure 5B).

Nonetheless, when considering the most extremely regulated genes in the NIH 3T3 transcriptome (i.e., the 10% most UP and the 10% most DN in response to LIN28B expression), classifiers trained on human data were up to 86% accurate in predicting post-transcriptional regulation in the mouse (Figure 5C, left). Moreover, when all genes expressed in NIH 3T3 cells

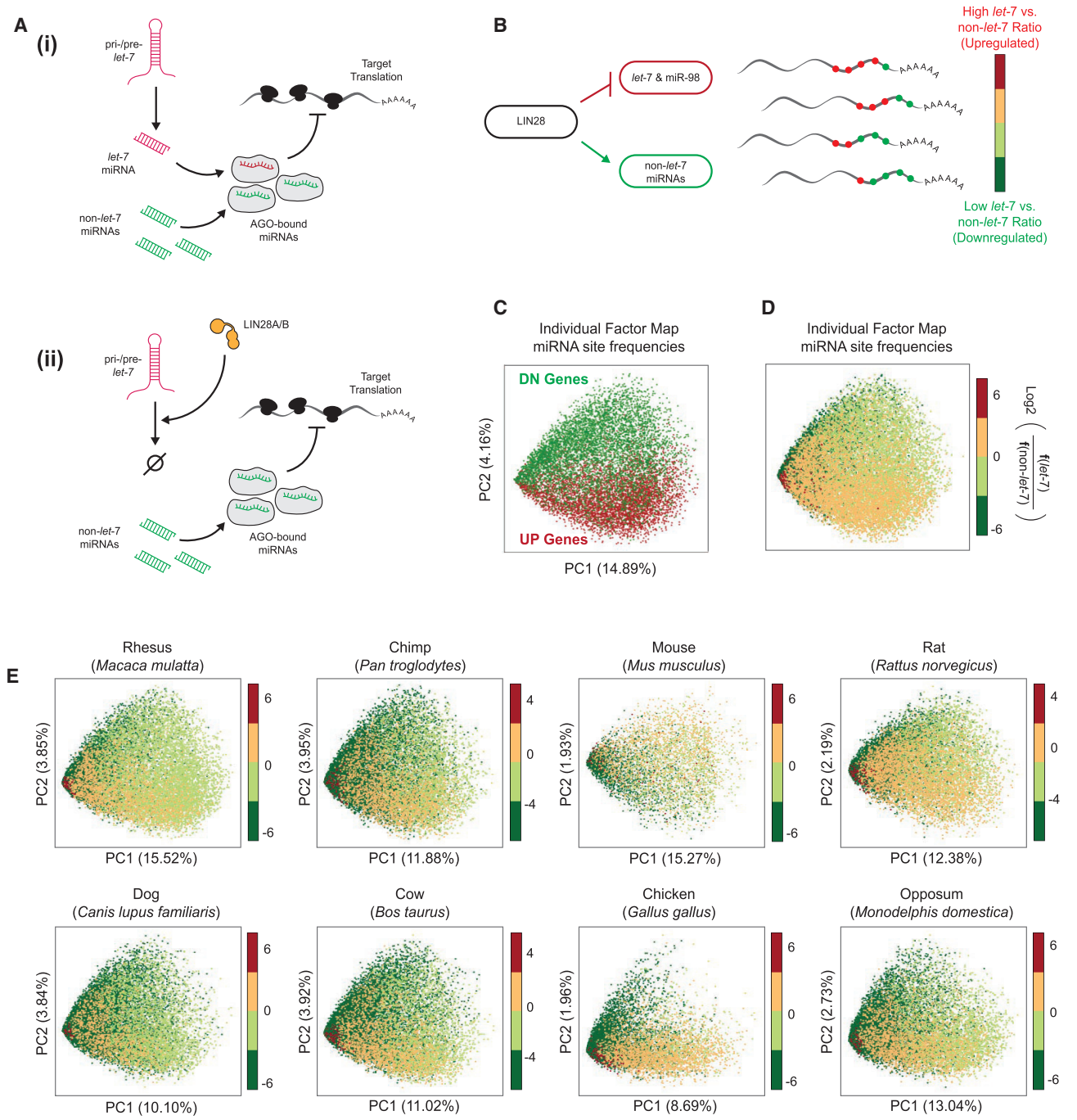



**Figure 3. A Model of LIN28B Regulation in Translational Control**

(A) LIN28B represses *let*-7/miR-98 biogenesis (i). However, this repression also allows AGO to bind more non-*let*-7 miRNAs (ii).

(B) LIN28B-dependent post-transcriptional regulation permits an increase and decrease in miRNA activity, effects that are resolved at the level of individual transcripts based on the composition of their 3′ UTR miRNA sites.

(C) UP and DN genes exhibit specific miRNA site enrichment patterns. UP and DN genes form two discrete territories when miRNA site frequencies per transcript are visualized using principal-component analysis.

(D) Principal-component analysis of variation in miRNA site frequencies per transcript reveals a correlation between PC2 and the enrichment of *let*-7 sites relative to non-*let*-7 miRNA sites per transcript.

(E) Principal-component analysis of variation in miRNA site frequencies per transcript for several species reveals broad conservation in miRNA site enrichment trends. PC2 is correlated with the enrichment of *let*-7 sites relative to non-*let*-7 miRNA sites per transcript.

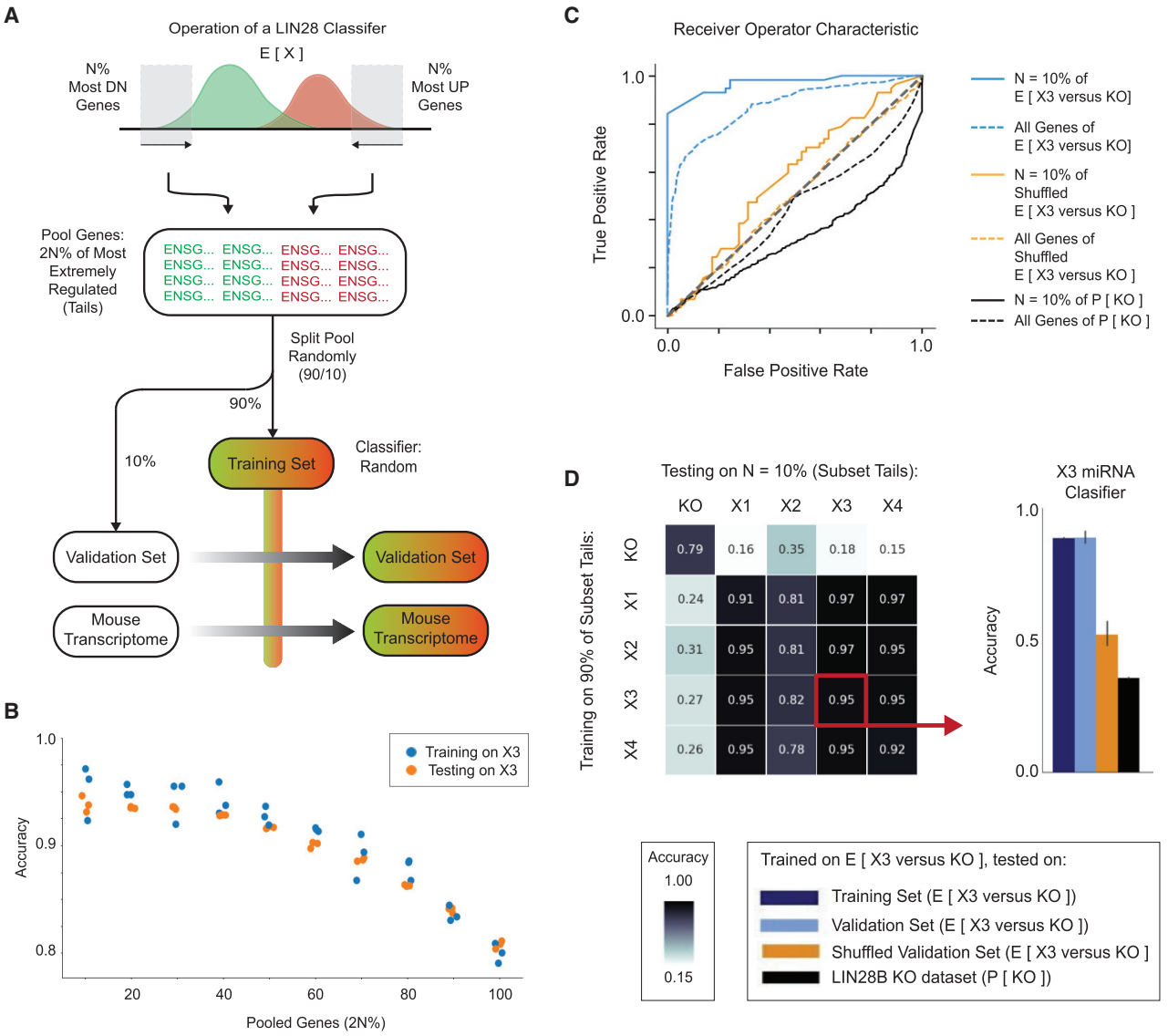



**Figure 4. Development of a Tree-Based Classifier for Predicting LIN28B-Dependent Translation Control**

(A) Cartoon outline of classifier implementation. (1) The 20% most extremely regulated genes are randomly split into a training set and a validation set. (2) Classifiers are trained on the features of the training set using random forest machine learning. (3) Classifier accuracy is assessed by performing predictions on a validation set. (4) Classifiers are used to predict LIN28B regulation in a test set.

(B) Training and testing accuracies as the percentage of genes used for training (2N%) is varied from 10% to 100%.

(C) Receiver operator characteristic showing the performance of a classifier that predicts based on miRNA site frequencies per gene.

(D) Heatmap showing reassignment accuracies of classifiers trained on miRNA site frequencies from E[KO], E[X1 versus KO], E[X2 versus KO], E[X3 versus KO], or E[X4 versus KO] and tested on data from any LIN28B expression condition in HEK293A cells. The bar plot (adjacent to heatmap) shows the average performance of an E[X3 versus KO] miRNA classifier on the same training data (dark blue), on a validation set of genes that was not initially made available to the classifier during training (light blue), a validation set where identifiers were shuffled randomly (orange), and a dataset that had no LIN28B expression, E[KO] (black), from three classifier iterations.

were considered, prediction accuracies dropped only by ~20% for all classifier combinations (Figure 5C, right), which is still better than random assignment (Figure 5D). Given the less significant separation between UP and DN subpopulations in mouse NIH 3T3 cells, lower prediction accuracies were expected (Figure 5A). These data indicate that mouse genes were successfully classified and that this successful classification was due to the treatment of non-*let*-7 miRNA sites as a broad predictive class that distinguished DN genes from UP genes (Figure 5E).

These data affirm the significance of overall miRNA site enrichment, not the presence or absence of *let*-7 sites, as the determinant of miRNA-mediated LIN28 regulation. Using classifiers trained on human and mouse datasets, we can now provide predictions of regulation for other species (Table S1). We noted

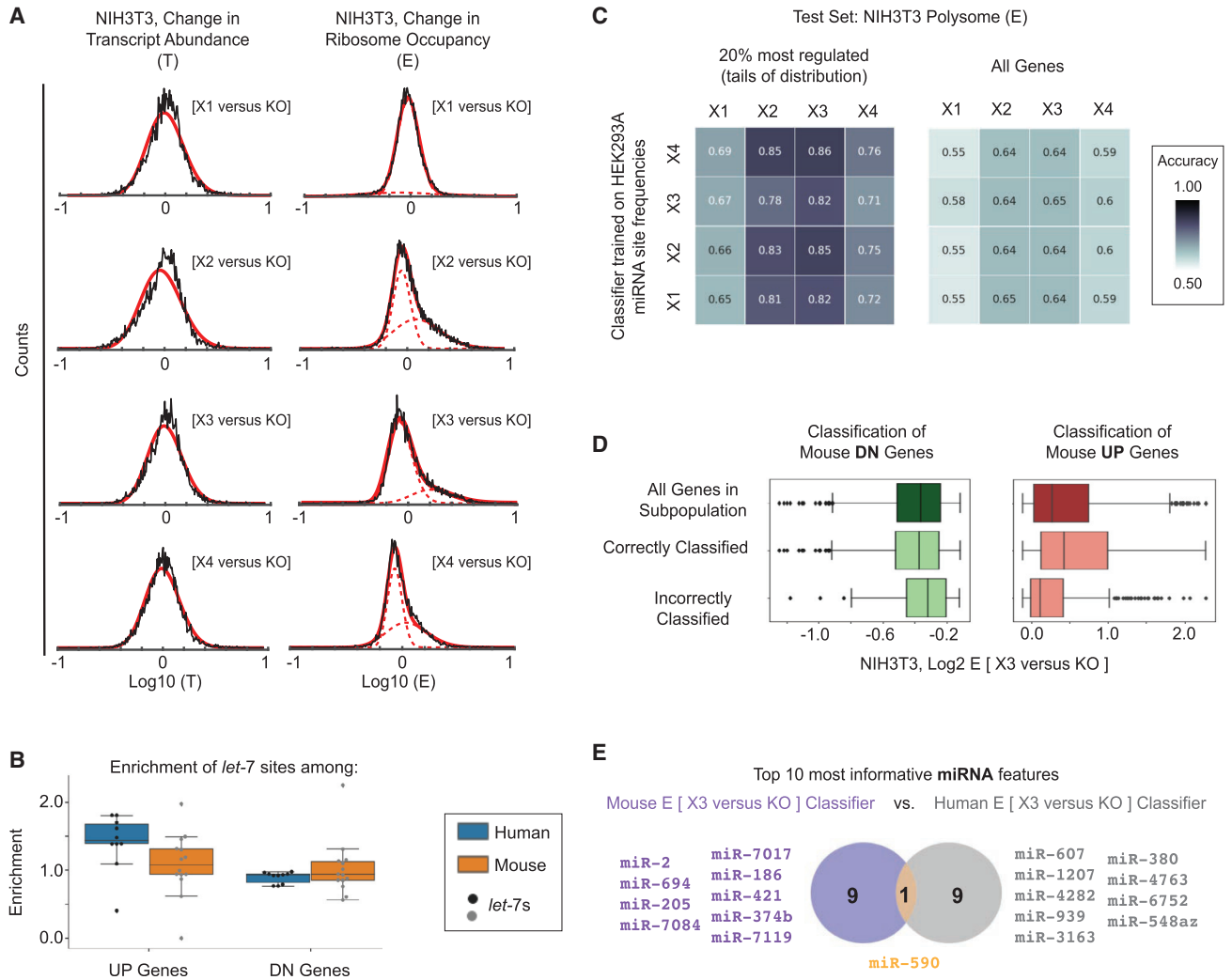



**Figure 5. Predicting LIN28B Regulation in Mouse NIH 3T3 Cells Using Human HEK293A Cell Data**

(A) LIN28B expression at level [X1], [X2], [X3], and [X4] in NIH 3T3 cells shows changes in transcript abundance and transcriptome-wide ribosome occupancy (E). Upregulated (UP) and downregulated (DN) subpopulations can be resolved. For each transcript abundance distribution, solid red lines represent a single gamma distribution fit to sample data. For each (E) distribution, dashed red lines approximate the shape of UP and DN subpopulations as single gamma distributions, and solid red lines are their sum, representing the best fit to sample data. These distributions show the changes in the regulation of any gene of more than 1 RPKM in NIH 3T3 cells (N = 9,314 genes).

(B) Boxplot showing enrichment and depletion of *let*-7 sites among UP and DN genes in human and mouse. Differences in enrichment may explain differences in the magnitude of post-transcriptional regulation between human and mouse.

(C) Heatmap showing prediction accuracies of classifiers trained on miRNA site frequencies (for 277 potential miRNAs overlapping human and mouse datasets) from E[X1 versus KO], E[X2 versus KO], E[X3 versus KO], or E[X4 versus KO] in HEK293A cells and tested on the mouse NIH 3T3 transcriptome. Prediction accuracy for the 20% most extremely regulated genes in NIH 3T3 cells (left heatmap) and for all NIH 3T3 genes (right heatmap) are presented for each applied LIN28B expression condition.

(D) Boxplots showing the correspondence between prediction outcome (correctly or incorrectly classified) using a human HEK293A miRNA classifier trained on E[X3 versus KO] and the measured fold change in ribosome occupancy resulting from LIN28B expression level [X3] in NIH 3T3 cells. For comparison, UP genes (E > E(median)) and DN genes (E < E(median)) in NIH 3T3 cells are presented separately.

(E) Venn diagram of the most informative miRNA features used by human HEK293A and mouse NIH 3T3 classifiers at LIN28B level [X3].

a high concordance in regulatory predictions between human and mouse classifiers. Together, these findings detail a conserved mechanism for translational control and one that exemplifies the central importance of RBP expression level in coordinating the regulation of multiple post-transcriptional pathways.

## DISCUSSION

RBPs are frequently outnumbered by their diverse and highly degenerate RNA targets, which often makes post-transcriptional control steps rate-limiting in regulating gene expression. Such properties may allow post-transcriptional regulation to occur

over a broader dynamic range of expression levels, which contrasts with behaviors at other levels of gene expression control (such as transcription or post-translational regulation). In this work, we show that differences in the functional readout of RNA regulation can be explained by differences in the availability of RBP and how it operates on predefined characteristics of a genome.

By using LIN28B to characterize the functional role of RBP expression level as a variable, we were able to revise its regulatory model (Figure 3). AGO and LIN28B proteins are bound competitively to determine global changes in miRNA activity. Using machine learning methods, we also discovered that the response to this post-transcriptional coordination in mammalian cells is evolutionarily conserved, encoded into the genome through changes in the number and composition of miRNA target sites. These biases also determine changes in mRNA abundance in response to gross changes in miRNA activity. Together, these correlations and anti-correlations in genomic features ultimately predict post-transcriptional control in human and mouse cells and indicate that almost every gene will be influenced by LIN28 expression. Interestingly, these data also suggest that cell-type-specific differences in AGO, LIN28, and endogenous *let*-7/miR-98 expression levels potentially alter the magnitude of regulation, which is an important consideration when interpreting downstream functional effects in different cellular or genomic contexts.

This work largely focuses on LIN28B-dependent regulation of *let*-7 miRNAs. However, LIN28 can play a role in the biogenesis of additional miRNA families in other cells (Peters et al., 2016; Warrander et al., 2016). LIN28 expression has also been shown to be involved in the regulation of co-activators or co-repressors of pri-miRNA transcription, which cause more global changes to the miRNA expression profile, especially during differentiation (Nowak et al., 2017). Such changes in absolute abundance should not be overlooked because they will also influence the final pool of available mature miRNAs in a cell. Other mediators of LIN28 regulation also remain to be explored, such as direct LIN28 binding to mRNAs. Previous work indicates that LIN28 binds mRNAs broadly to either promote (Graf et al., 2013; Hafner et al., 2013; Wilbert et al., 2012) or repress translation (Cho et al., 2012). How these and perhaps additional regulatory mechanisms respond to different levels of LIN28 expression are not known.

The mechanisms described above relate to LIN28 specifically, but the insights are relevant to any RBP involved in miRNA biogenesis. Previous work identified numerous RBPs that associate with miRNA sequences (Nussbacher and Yeo, 2018; Treiber et al., 2017). When characterizing the miRNA-regulatory function of these additional RBPs, an important consideration is to avoid experiments that depend on the miRNA pathway as a means to control gene expression. The introduction of saturating amounts of any exogenous short hairpin RNA (shRNA), small interfering RNA (siRNA), or complementary miRNA (antagomir) may influence the regulatory readout. Previous work studying LIN28 regulation using miRNA pathway-mediated perturbations should also be carefully reconsidered.

Our decomposition of LIN28 regulation highlights the coordination that may exist between post-transcriptional regulatory pathways and expands the scope of regulation beyond *let*-7 and its specific targets. Sub-saturating levels of LIN28 and AGO redistributes miRNA activity from *let*-7 to non-*let*-7 miRNAs, which enable widespread changes in gene regulation. The biological and pathological effects of LIN28 expression could therefore result from both a loss of function in *let*-7 regulation and a gain of function in non-*let*-7 regulation. This mechanism appears to be conserved and may have allowed LIN28 to play a significant role in the evolution of miRNA sites and their genome-wide enrichment patterns.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Poly-ribosome Fractionation and Library Preparation
  - Small RNA Extraction and Library Preparation
  - Quantitative PCR
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Change in Total Transcript Abundance (T) & Change in Polysome Enrichment (E)
  - Regulatory Feature Enrichment Analysis
  - Random Forest Classifiers
- DATA AND SOFTWARE AVAILABILITY

### SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and one table and can be found with this article online at https://doi.org/10.1016/j.molcel.2018.10.041.

### ACKNOWLEDGMENTS

We would like to thank Dr. Amy Pasquinelli for constructive advice and feedback on the manuscript. F.E.T. is supported by a postdoctoral research fellowship (grant 129547-PF-16-060-01-RMC) from the American Cancer Society. E.C.W. is supported by grants from the University of California, San Diego Genetics Training Program (T32 and GM008666) and the NSF Graduate Research Fellowship Program. J.K.N. is an Achievement Rewards for College Scientists Foundation Fellow and was partially supported by a Genentech Foundation fellowship and the University of California, San Diego, Cancer Training Program through an institutional training grant from the National Institute of General Medical Sciences (T32CA067754). G.W.Y was partially supported by grants from the NIH (HG007005, HG009889, HL137223, and HG004659).

### AUTHOR CONTRIBUTIONS

Conceptualization, F.E.T. and G.W.Y.; Methodology, F.E.T., S.S., J.K.N., E.C.W., and S.P.; Formal Analysis, F.E.T., S.S., and E.C.W.; Investigation, F.E.T.; Writing – Original Draft, F.E.T.; Writing – Review & Editing, F.E.T., E.C.W., S.S., and G.W.Y.; Supervision, G.W.Y., Funding Acquisition, G.W.Y.

### DECLARATION OF INTERESTS

G.W.Y. is a co-founder of Locana and Eclipse Bioinnovations and a member of the scientific advisory boards of Locana, Eclipse Bioinnovations, and Aquinnah Pharmaceuticals. The terms of this arrangement have been reviewed and approved by the University of California, San Diego, in accordance with its conflict of interest policies.

Received: February 1, 2018
Revised: August 21, 2018
Accepted: October 27, 2018
Published: December 6, 2018

## REFERENCES

Agarwal, V., Bell, G.W., Nam, J.-W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. eLife *4*, e05005.

Breiman, L. (2001). Random forests. Mach. Learn. *45*, 5–32.

Cho, J., Chang, H., Kwon, S.C., Kim, B., Kim, Y., Choe, J., Ha, M., Kim, Y.K., and Kim, V.N. (2012). LIN28A is a suppressor of ER-associated translation in embryonic stem cells. Cell *151*, 765–777.

Copley, M.R., Babovic, S., Benz, C., Knapp, D.J.H.F., Beer, P.A., Kent, D.G., Wohrer, S., Treloar, D.Q., Day, C., Rowe, K., et al. (2013). The Lin28b-let-7-Hmga2 axis determines the higher self-renewal potential of fetal haematopoietic stem cells. Nat. Cell Biol. *15*, 916–925.

Graf, R., Munschauer, M., Mastrobuoni, G., Mayr, F., Heinemann, U., Kempa, S., Rajewsky, N., and Landthaler, M. (2013). Identification of LIN28B-bound mRNAs reveals features of target recognition and regulation. RNA Biol. *10*, 1146–1159.

Hafner, M., Max, K.E.A., Bandaru, P., Morozov, P., Gerstberger, S., Brown, M., Molina, H., and Tuschl, T. (2013). Identification of mRNAs bound and regulated by human LIN28 proteins and molecular requirements for RNA recognition. RNA *19*, 613–626.

Heo, I., Joo, C., Kim, Y.-K., Ha, M., Yoon, M.-J., Cho, J., Yeom, K.-H., Han, J., and Kim, V.N. (2009). TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. Cell *138*, 696–708.

Loughlin, F.E., Gebert, L.F.R., Towbin, H., Brunschweiger, A., Hall, J., and Allain, F.H.-T. (2011). Structural basis of pre-let-7 miRNA recognition by the zinc knuckles of pluripotency factor Lin28. Nat. Struct. Mol. Biol. *19*, 84–89.

Madison, B.B., Liu, Q., Zhong, X., Hahn, C.M., Lin, N., Emmett, M.J., Stanger, B.Z., Lee, J.-S., and Rustgi, A.K. (2013). LIN28B promotes growth and tumorigenesis of the intestinal epithelium via Let-7. Genes Dev. *27*, 2233–2245.

Nam, Y., Chen, C., Gregory, R.I., Chou, J.J., and Sliz, P. (2011). Molecular basis for interaction of let-7 microRNAs with Lin28. Cell *147*, 1080–1091.

Nowak, J.S., Hobor, F., Downie Ruiz Velasco, A., Choudhury, N.R., Heikel, G., Kerr, A., Ramos, A., and Michlewski, G. (2017). Lin28a uses distinct mechanisms of binding to RNA and affects miRNA levels positively and negatively. RNA *23*, 317–332.

Nussbacher, J.K., and Yeo, G.W. (2018). Systematic discovery of RNA binding proteins that regulate microRNA levels. Mol. Cell *69*, 1005–1016.e7.

Peters, D.T., Fung, H.K.H., Levdikov, V.M., Irmscher, T., Warrander, F.C., Greive, S.J., Kovalevskiy, O., Isaacs, H.V., Coles, M., and Antson, A.A. (2016). Human Lin28 forms a high-affinity 1:1 complex with the 106∼363 cluster miRNA miR-363. Biochemistry *55*, 5021–5027.

Roush, S., and Slack, F.J. (2008). The let-7 family of microRNAs. Trends Cell Biol. *18*, 505–516.

Shinoda, G., Shyh-Chang, N., Soysa, T.Y., Zhu, H., Seligson, M.T., Shah, S.P., Abo-Sido, N., Yabuuchi, A., Hagan, J.P., Gregory, R.I., et al. (2013). Fetal deficiency of lin28 programs life-long aberrations in growth and glucose metabolism. Stem Cells *31*, 1563–1573.

Shyh-Chang, N., and Daley, G.Q. (2013). Lin28: primal regulator of growth and metabolism in stem cells. Cell Stem Cell *12*, 395–406.

Shyh-Chang, N., Zhu, H., Yvanka de Soysa, T., Shinoda, G., Seligson, M.T., Tsanov, K.M., Nguyen, L., Asara, J.M., Cantley, L.C., and Daley, G.Q. (2013). Lin28 enhances tissue repair by reprogramming cellular metabolism. Cell *155*, 778–792.

Singh, R., and Valcárcel, J. (2005). Building specificity with nonspecific RNA-binding proteins. Nat. Struct. Mol. Biol. *12*, 645–653.

Treiber, T., Treiber, N., Plessmann, U., Harlander, S., Daiß, J.-L., Eichner, N., Lehmann, G., Schall, K., Urlaub, H., and Meister, G. (2017). A compendium of RNA-binding proteins that regulate microRNA biogenesis. Mol. Cell *66*, 270–284.e13.

Van Wynsberghe, P.M., Kai, Z.S., Massirer, K.B., Burton, V.H., Yeo, G.W., and Pasquinelli, A.E. (2011). LIN-28 co-transcriptionally binds primary let-7 to regulate miRNA maturation in Caenorhabditis elegans. Nat. Struct. Mol. Biol. *18*, 302–308.

Viswanathan, S.R., Daley, G.Q., and Gregory, R.I. (2008). Selective blockade of microRNA processing by Lin28. Science *320*, 97–100.

Warrander, F., Faas, L., Kovalevskiy, O., Peters, D., Coles, M., Antson, A.A., Genever, P., and Isaacs, H.V. (2016). lin28 proteins promote expression of 17∼92 family miRNAs during amphibian development. Dev. Dyn. *245*, 34–46.

Wilbert, M.L., Huelga, S.C., Kapeli, K., Stark, T.J., Liang, T.Y., Chen, S.X., Yan, B.Y., Nathanson, J.L., Hutt, K.R., Lovci, M.T., et al. (2012). LIN28 binds messenger RNAs at GGAGA motifs and regulates splicing factor abundance. Mol. Cell *48*, 195–206.

Yang, D.-H., and Moss, E.G. (2003). Temporally regulated expression of Lin-28 in diverse tissues of the developing mouse. Gene Expr. Patterns *3*, 719–726.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917–1920.

Yuan, J., Nguyen, C.K., Liu, X., Kanellopoulou, C., and Muljo, S.A. (2012). Lin28b reprograms adult bone marrow hematopoietic progenitors to mediate fetal-like lymphopoiesis. Science *335*, 1195–1200.

Zhang, J., Ratanasirintrawoot, S., Chandrasekaran, S., Wu, Z., Ficarro, S.B., Yu, C., Ross, C.A., Cacchiarelli, D., Xia, Q., Seligson, M., et al. (2016). LIN28 regulates stem cell metabolism and conversion to primed pluripotency. Cell Stem Cell *19*, 66–80.

Zhou, J., Ng, S.-B., and Chng, W.-J. (2013). LIN28/LIN28B: an emerging oncogenic driver in cancer stem cells. Int. J. Biochem. Cell Biol. *45*, 973–978.

Zhu, H., Shyh-Chang, N., Segrè, A.V., Shinoda, G., Shah, S.P., Einhorn, W.S., Takeuchi, A., Engreitz, J.M., Hagan, J.P., Kharas, M.G., et al.; DIAGRAM Consortium; MAGIC Investigators (2011). The Lin28/let-7 axis regulates glucose metabolism. Cell *147*, 81–94.

# STAR★METHODS

## KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Antibodies | | |
| Anti-V5 | MBL | Cat# M167-3; RRID: AB_1953024 |
| Anti-LIN28B (D4H1) | Cell Signaling Technology | Cat# 11965; RRID: AB_2750978 |
| Anit-LIN28A (6D1F9) | Cell Signaling Technology | Cat# 5930; RRID: AB_1903976 |
| Anti-AGO2 (9E8.2) | Millipore | Cat# 04-642; RRID: AB_1586875 |
| Chemicals, Peptides, and Recombinant Proteins | | |
| Doxycycline hydrochloride | Sigma | D3447 |
| Critical Commercial Assays | | |
| Truseq RNA Sequencing kit v2 | Illumina | RS-122-2001 |
| Truseq Stranded RNA Sequencing kit | Illumina | 20020594 |
| Trusseq Small RNA kit | Illumina | RS-200-0012 |
| Deposited Data | | |
| RNA sequencing data | This Paper | GEO: GSE109423 |
| Mendeley: Uncropped Western Blots | Mendeley | https://data.mendeley.com/datasets/m5czwjj3t5/1 |
| Experimental Models: Cell Lines | | |
| LIN28B knockout (human HEK293A) | This Paper | N/A |
| Human HEK293A | | Related to CRL-1573 |
| Mouse NIH 3T3 | | Related to CRL-1658 |
| Oligonucleotides | | |
| Let-7a mirVana miRNA mimic | Thermo Fisher Scientific | AM17100 |
| mirVana Negative Control | Thermo Fisher Scientific | 4464058 |
| Recombinant DNA | | |
| LIN28B transgene PiggyBAC | This Paper | N/A |
| AGO2 transgene PiggyBAC | This Paper | N/A |
| Software and Algorithms | | |
| Random Forest Classifiers | Breiman, 2001 | N/A |

## CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to and will be fulfilled by the Lead Contact Gene Yeo (geneyeo@ucsd.edu).

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

LIN28B double knockout human HEK293A cells (KO) were generated using CRISPR-Cas9 targeting of the LIN28B start codon (spacer sequence: 5′-gtgagggcccgtggggcaaca-3′). LIN28B expression was reintroduced into KO cells using inducible constructs housed on PiggyBAC shuttle vectors (see Figure 1), which were stably integrated at variable copy number via transient transfection. Four discrete windows of transgene copy number were chosen (as gauged by YFP expression; named [X1], [X2], [X3] and [X4]), and cells falling within these windows were sorted and cultured separately for bulk experimentation or expression profiling. The same protocol was used to maintain stable LIN28B expression mouse NIH 3T3 cells. Transgene expression was induced using 10ng/ml Doxycycline. All cell cultures were maintained in DMEM supplemented with 10% Fetal Bovine Serum and grown in a humidified incubator at 37°C and 5% $CO_2$.

## METHOD DETAILS

### Poly-ribosome Fractionation and Library Preparation

Cultures of approximately 2 × $10^7$ HEK293A cells per LIN28B expression condition were harvested and lysed in polysome buffer (20mM Tris-HCl pH 7.4, 150mM NaCl, 5mM MgCl2, 0.5% NP-40 + Murine RNase inhibitor (20U/ml) + 1mM DTT + 1x Protease inhibitors). Lysates were incubated on ice for 30 min, and briefly centrifuged to separate nuclei and cellular debris. Collected supernatants were split, with a volume set aside for Total RNA extraction, and an equal volume loaded onto a 10%–50% sucrose cushion. Cellular transcripts were separated on sucrose cushions based on density (110,000 x g for 3 hr at 4°C) and were dispensed via fractionation as twenty-two individual aliquots (0.5mL/each) spanning the 10%–50% sucrose range. A poly-ribosome enriched RNA sample was prepared by pooling 50ul from each poly-ribosome fraction (approximately fractions #9 through #20). Total RNA extraction and poly-ribosome enriched RNA extraction was then performed with Trizol LS. Total RNA and poly-ribosome enriched RNA sequencing libraries were prepared using a TruSeq Library Prep Kit v2 (Illumina).

### Small RNA Extraction and Library Preparation

Approximately 2 × $10^7$ cells per experimental condition were lysed in polysome buffer on ice for 30 min, and briefly centrifuged to separate nuclei and cellular debris. Collected supernatants were then split, with a volume set aside for Total Small RNA extraction, and an equal volume for AGO2 immunoprecipitation (Millipore, Anti-Ago2 Clone 9E8.2). Small RNAs derived from native lysates or from immunoprecipitations were isolated using miRvana kit reagents (Ambion). Isolated small RNAs were then processed to generate sequencing libraries using a Small RNA Library Prep Kit (Illumina).

### Quantitative PCR

Changes in transcript abundance were assessed by RT-qPCR using High-Capacity cDNA Reverse Transcription Kit (Thermo) and Power SYBR Green PCR Master Mix (Thermo) on BioRad's CFX Real-time PCR Detection Systems.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Change in Total Transcript Abundance (T) & Change in Polysome Enrichment (E)

Only transcripts with RPKM > 1 were considered. All statistics were computed using length and depth normalized measurements (RPKM) for each LIN28B expression condition, [ i ].

$$[i] = [X1], [X2], [X3] \text{or} [X4].$$

The change in total transcript abundance T[ i versus KO ], is the ratio of Total RNA transcript abundance in LIN28B expression condition [ i ], over Total RNA transcript abundance in knockout cells KO.

$$T[i \text{ versus } KO] = \frac{\text{Total RNA}[i]}{\text{Total RNA}[KO]}$$

The relative polysome enrichment per condition, P [ i ], is the ratio of transcript abundance in polysome fraction derived RNA in LIN28B expression condition [ i ], to Total RNA transcript abundance in LIN28B expression condition [ i ]. P [ KO ] is the relative polysome enrichment in KO cells.

$$P[i\,] = \frac{\text{Polysome RNA}[i\,]}{\text{Total RNA}[i\,]},\ P[KO] = \frac{\text{Polysome RNA}[KO]}{\text{Total RNA}[KO]}$$

The change in poly-ribosome enrichment, E[ i versus KO ], is the ratio P [ i ] and P [ KO ].

### Regulatory Feature Enrichment Analysis

TargetScan (v7.1) miRNA site predictions were used to compute site enrichment. All binding sites, conserved and non-conserved with 6-8nt seeds, were considered. The enrichment of every miRNA was computed, on a per gene basis. Enrichment was scored as the frequency of a specific miRNA among all sites found on a gene, normalized by its frequency across all genes.

### Random Forest Classifiers

We developed Random Forest Classifiers to predict LIN28B regulation in non-human cell types (Breiman, 2001). More specifically, these Classifiers were trained and tested on the co-variance between miRNA site enrichment scores the change in ribosome occupancy of individual transcripts. Co-variance was profiled for five different poly-ribosome distributions (P[KO], E[X1 versus KO], E[X2 versus KO], E[X3 versus KO] or E[X4 versus KO]) across two different species (human and mouse), to generate 10 Random Forest Classifier models.

We trained and validated Classifiers on the miRNA enrichment statistics of the most extremely regulated genes (Figure 4). Importantly, training and validation was conducted on separate gene batches, derived from the same Gene Subset. 90% of the Gene Subset was slated for training (Training Set), with the remaining genes used for validation (Validation Set). Classifier parameters were

optimized to achieve high re-assignment accuracy (> 80%) on validation data, as gauged by three individual iterations of training and validation. Each model was then tested on data from all other LIN28B levels, to gauge inter-Classifier prediction accuracies. Finally, the models were also tested on data from mouse NIH 3T3 at different LIN28B expression levels (Test Set). For graphical diagram, see Figure 4C.

## DATA AND SOFTWARE AVAILABILITY

The accession number for the RNA sequencing reported in this paper is in GEO: GSE109423

Computer code used for bioinformatics processing and analysis of RNA sequencing datasets will be made available upon request.

Original uncropped western blots available on Mendeley. https://data.mendeley.com/datasets/m5czwjj3t5/1.