

# LIN28A Is a Suppressor of ER-Associated Translation in Embryonic Stem Cells

Jun Cho,<sup>1,2,4</sup> Hyeshik Chang,<sup>1,2,4</sup> S. Chul Kwon,<sup>1,2</sup> Baekgyu Kim,<sup>1,2</sup> Yoosik Kim,<sup>1,2</sup> Junho Choe,<sup>3</sup> Minju Ha,<sup>1,2</sup> Yoon Ki Kim,<sup>3</sup> and V. Narry Kim<sup>1,2,\*</sup>

<sup>1</sup>Institute for Basic Science

<sup>2</sup>School of Biological Sciences

Seoul National University, Seoul 151-742, Korea

<sup>3</sup>School of Life Sciences and Biotechnology, Korea University, Seoul 136-701, Korea

<sup>4</sup>These authors contributed equally to this work

\*Correspondence: narrykim@snu.ac.kr

<http://dx.doi.org/10.1016/j.cell.2012.10.019>

## SUMMARY

LIN28 plays a critical role in developmental transition, glucose metabolism, and tumorigenesis. At the molecular level, LIN28 is known to repress maturation of let-7 microRNAs and enhance translation of certain mRNAs. In this study, we obtain a genome-wide view of the molecular function of LIN28A in mouse embryonic stem cells by carrying out RNA crosslinking-immunoprecipitation-sequencing (CLIP-seq) and ribosome footprinting. We find that, in addition to let-7 precursors, LIN28A binds to a large number of spliced mRNAs. LIN28A recognizes AAGNNG, AAGNG, and less frequently UGUG, which are located in the terminal loop of a small hairpin. LIN28A is localized to the periendoplasmic reticulum (ER) area and inhibits translation of mRNAs that are destined for the ER, reducing the synthesis of transmembrane proteins, ER or Golgi lumen proteins, and secretory proteins. Our study suggests a selective regulatory mechanism for ER-associated translation and reveals an unexpected role of LIN28A as a global suppressor of genes in the secretory pathway.

## INTRODUCTION

LIN28 is a conserved RNA binding protein whose homologs are found from worms to humans. It was originally identified as a regulator of developmental timing in *Caenorhabditis elegans* (Moss et al., 1997), and its expression is tightly regulated during animal development (Moss and Tang, 2003). Mammals have two homologs, *Lin28a* and *Lin28b*. *Lin28a* is highly expressed in embryonic stem cells (ESCs) and was shown as one of the four factors that convert fibroblasts into induced pluripotent stem cells (Yu et al., 2007). Perturbation of *Lin28* results in developmental defects and tumorigenesis. In mouse, for instance, *Lin28a* deficiency caused undergrowth and lethality in early stages of development, whereas its ectopic expression induced overgrowth and delayed the timing of puberty (Zhu et al., 2010).

Furthermore, *Lin28a/b* promote malignant transformation, and their expression is associated with advanced stages of many types of tumors, including hepatocarcinoma, nephroblastoma, ovarian carcinoma, and germ cell tumors (Thornton and Gregory, 2012; Viswanathan et al., 2009).

At the molecular level, LIN28 acts as a suppressor of let-7 microRNA biogenesis (Heo et al., 2008; Newman et al., 2008; Rybak et al., 2008; Viswanathan et al., 2008). In the nucleus, LIN28 binds to the primary transcript of let-7 (pri-let-7) and prevents its processing by RNase III DROSHA (Newman et al., 2008; Viswanathan et al., 2008). In the cytoplasm, it interacts with the precursor form of let-7 (pre-let-7) and interferes with pre-let-7 processing (Heo et al., 2008; Rybak et al., 2008). LIN28 recruits TUTase 4 (ZCCHC11) to induce oligo-uridylation of pre-let-7, which effectively blocks DICER processing and facilitates degradation of the RNA (Hagan et al., 2009; Heo et al., 2008; Heo et al., 2009). Although LIN28B is localized mainly in the nucleolus and interferes with nuclear processing, LIN28A is found mostly in the cytoplasmic compartment and acts in concert with TUTase 4 (Piskounova et al., 2011). LIN28 homologs commonly have two types of RNA binding domains: a cold shock domain and a cluster of two CCHC-type zinc finger motifs. We previously showed by biochemical analyses that the “GGAG” sequences in the terminal loop of let-7 precursors serve as the binding site for the zinc finger domains that are critical for let-7 regulation (Heo et al., 2009). More recent structural studies revealed the molecular basis of the interaction between LIN28 and the terminal loop of pre-let-7 (Loughlin et al., 2012; Nam et al., 2011).

Although let-7 is certainly a key target of LIN28, multiple lines of evidence support additional functions of LIN28. First, during retinoic-acid-induced neurogenesis, *Lin28a* overexpression altered the expression of several transcription factors involved in early embryonic cell fate decision before any increase in let-7 level was detected (Balzer et al., 2010). Second, impaired glucose tolerance and insulin resistance were observed in muscle-specific *Lin28a* knockout mice without significant changes in the let-7 level (Zhu et al., 2011). Third, LIN28A interacts with mRNAs and cosediments with polysome in sucrose gradient centrifugation (Balzer and Moss, 2007). Consistently, several studies reported that LIN28A can bind to and enhance translation of certain mRNAs such as *Igf2* in differentiating

myoblasts and Oct4 in ESCs (Polesskaya et al., 2007; Qiu et al., 2010; Xu and Huang, 2009; Xu et al., 2009).

In order to obtain a genome-wide view of LIN28A's function, we identified LIN28A-interacting RNAs by using a technique that combines UV crosslinking, immunoprecipitation, and high-throughput sequencing (CLIP-seq, also known as HITS-CLIP) (Chi et al., 2009; Licatalosi et al., 2008). This technique takes advantage of UV irradiation of live cells, which introduces covalent bonds between the bases and amino acids (such as Cys, Lys, Phe, Trp, and Tyr) in close proximity ( $\sim 1 \text{ \AA}$ ), capturing the physiological state of RNA-protein interaction (Licatalosi and Darnell, 2010; Wagenmakers et al., 1980). Because of the cross-linking, one can use harsh immunoprecipitation conditions to eliminate artificial RNA-protein interactions that commonly occur in cell lysates (Dreyfuss et al., 1984). In addition, sequencing of the coimmunopurified RNA fragments provides a global view of the interacting transcriptome and allows mapping of exact binding sites on target RNAs (Kishore et al., 2011; Licatalosi et al., 2008; Ule et al., 2005; Ule et al., 2003; Zhang and Darnell, 2011).

In this study, we find that LIN28A targets not only the let-7 precursors but also a large number of mRNAs. Analyses of the binding sites reveal general features of LIN28A binding motif. We further discover by ribosome footprinting that LIN28A acts as a suppressor of translation in the endoplasmic reticulum (ER) pathway.

## RESULTS

### LIN28A CLIP-seq from Mouse Embryonic Stem Cells

To purify the LIN28A-RNA complexes, we irradiated mouse embryonic stem cells (mESCs) with UV light, lysed the cells, treated the lysates with RNase A, and immunoprecipitated the complexes by using antibodies against LIN28A. RNA fragments from the immunoprecipitates were ligated to linkers, amplified by RT-PCR and sequenced with Illumina Genome Analyzer IIx (Figures 1A and S1A available online).

To ensure specific isolation of LIN28A-interacting RNAs, we obtained three different libraries by using three anti-LIN28A antibodies, one rabbit polyclonal antibody (Abcam) and two mouse monoclonal antibodies raised in-house (35L33G and 2J3). The amounts of RNA recovered from immunoprecipitates were significantly reduced when LIN28A was depleted, indicating that the antibodies specifically precipitated the LIN28A protein and its target RNAs (Figures 1B and S1B).

About 32 million reads were obtained from each CLIP cDNA library and 32% of them were confidently mapped to unique positions in the mouse genome (Table S1). To assess the reproducibility of the experiments, the correlation of transcript-level enrichments was evaluated across the three CLIP-seq libraries. The Pearson coefficients between any two of the three libraries were at least 0.93 (Figure S1C), indicating that most of the LIN28A CLIP tags had originated from the common pool of LIN28A-interacting RNAs.

### Confirmation of the Let-7 Family as a Target of LIN28A

Because the let-7 family miRNAs are the most thoroughly studied targets of LIN28A, they serve as important positive con-

trols for our experiment. As expected, we detected CLIP tags from let-7 precursors in the sequence alignment (Figures 1C and 1D for *Mirlet7g*; see also Figures S2A and S2B for *Mirlet7d* and *Mirlet7f-1*, respectively).

Because it remains unclear whether the let-7 family is the only miRNA target of LIN28A in mESCs, we analyzed our CLIP-seq library to find other LIN28A-interacting miRNAs. Because there is yet no reliable method to globally quantify the amounts of miRNA precursors, we had to use mature miRNA signals from microarrays for normalization. Despite the technical limitation, the let-7 family comprises the most enriched group in the CLIP-seq (Figure 1E). Although a few miRNA precursors such as pre-miR-677 and pre-miR-708 showed comparable enrichments, their mature miRNA levels did not change significantly upon *Lin28a* knockdown (Figure 1E and Table S2). Thus, this analysis indicates that, at least in mESCs, the let-7 family is likely to be the only functional miRNA target of LIN28A.

### LIN28A Binds to Let-7 Precursors through the GGAG Motif

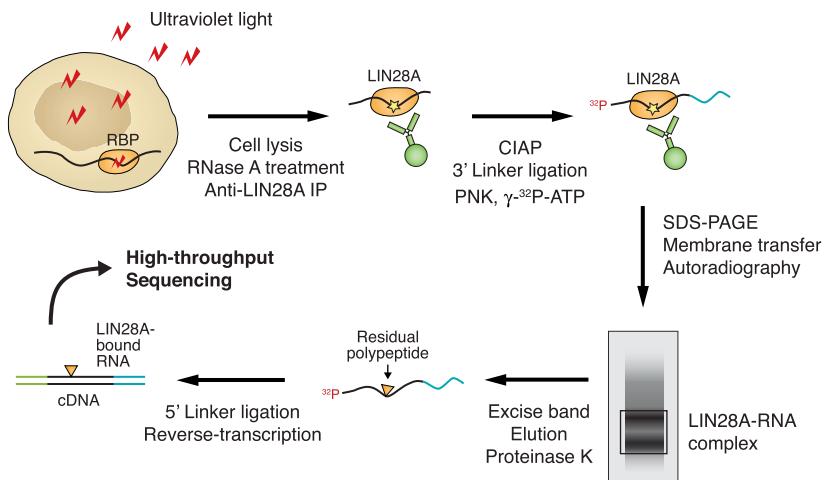
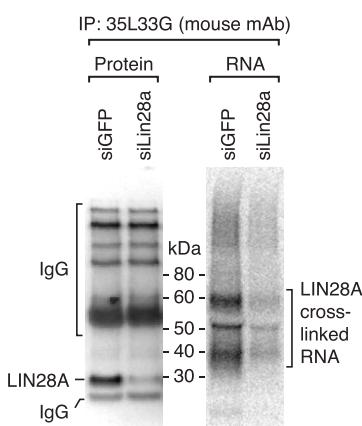
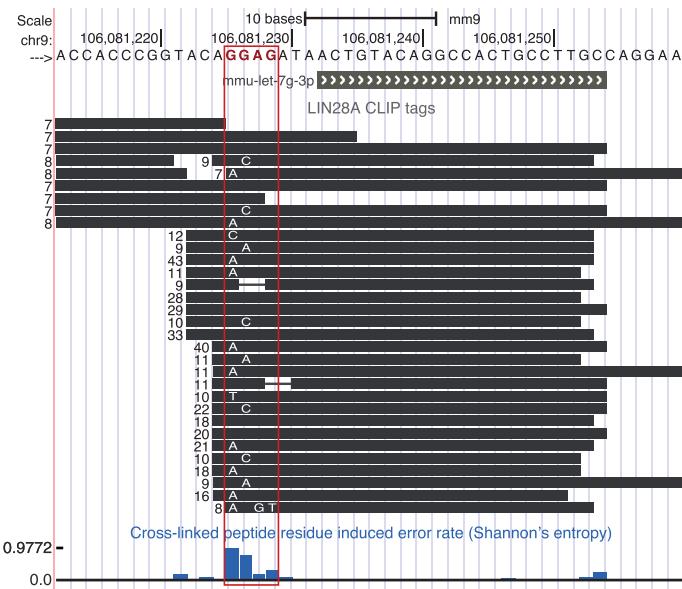
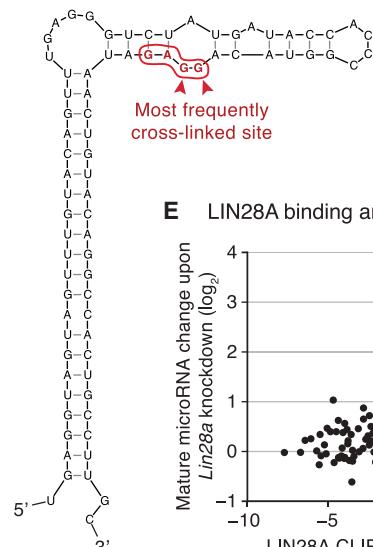
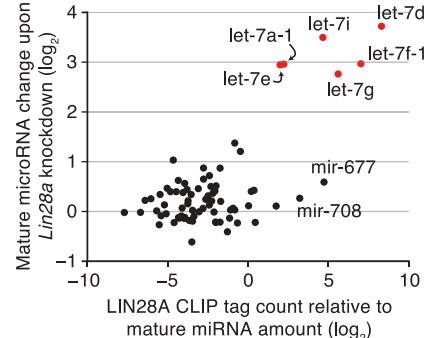
Our previous biochemical study showed that the GGAG motif in the terminal loop of the let-7 precursor serves as the binding center for the CCHC zinc finger domains and is required for specific recognition by LIN28A (Heo et al., 2009). Almost all let-7g CLIP tags contain the GGAG sequence, indicating that our CLIP libraries faithfully reflect physiological interactions between LIN28A and let-7g precursor.

Interestingly, we noticed frequent sequence alterations at the GGAG motif in our CLIP tags (Figures 1C and 1D). Particularly, the first guanosine of the motif is often substituted with other nucleotides. Similar substitution pattern was observed with other let-7 precursors as well (Figures S2A and S2B). Such alteration on a specific site is likely to be an outcome of UV crosslinking to LIN28A because a small piece of crosslinked peptide would remain even after proteinase treatment and interfere with the progression of reverse transcriptase as previously reported (Granneman et al., 2009; Kishore et al., 2011; Zhang and Darnell, 2011).

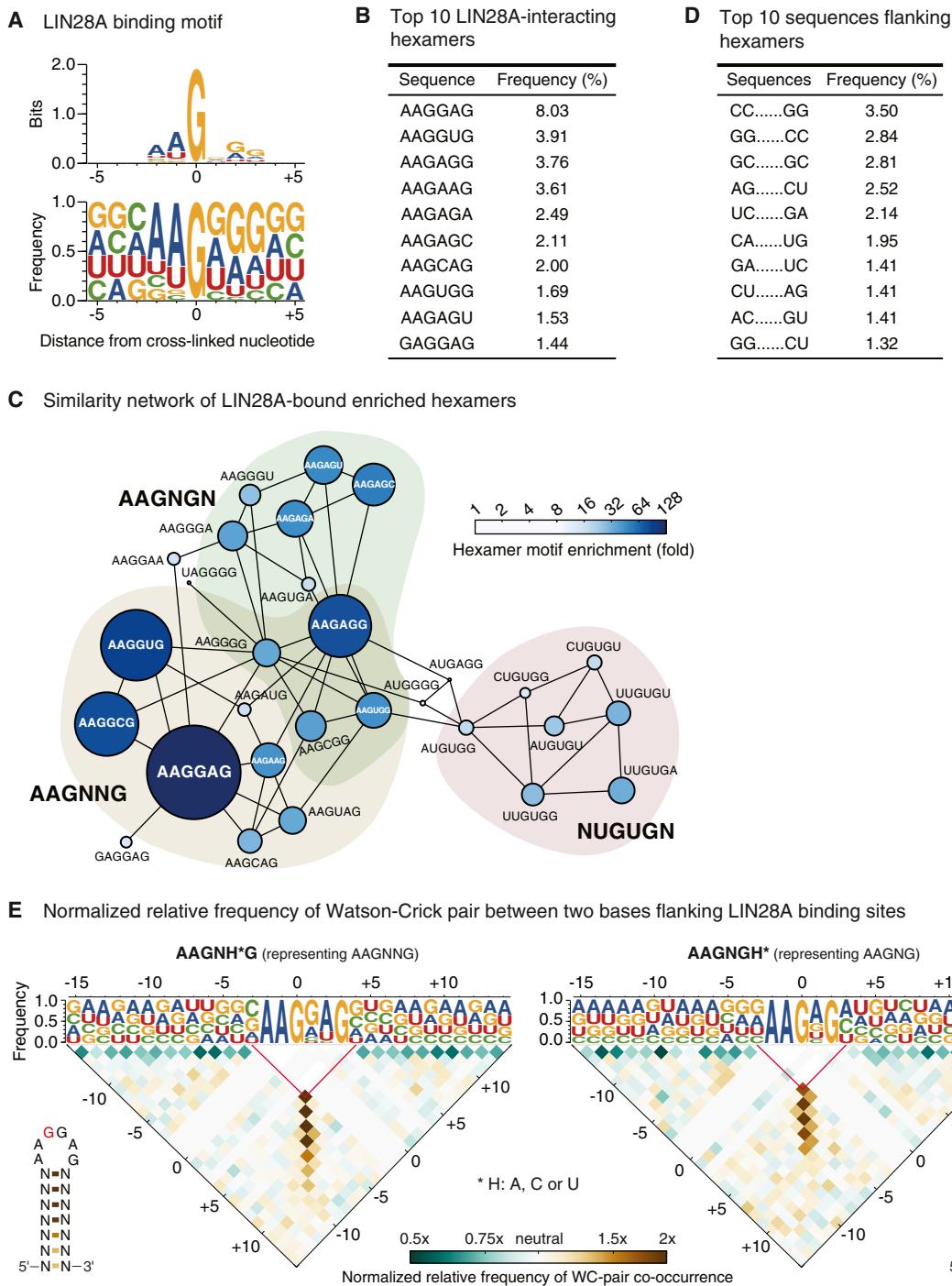
### Crosslinking-Induced Errors Allow Identification of LIN28A Binding Sites at Single-Nucleotide Resolution

Our libraries contain a large number of tags mapped to non-miRNA transcripts, suggesting that LIN28A may interact with other types of RNAs in addition to let-7 precursors. To precisely map the binding sites in such transcripts, we took advantage of the mutations. Substitution and deletion errors are significantly more prevalent in our CLIP-seq libraries (~1.5% for substitution and ~0.5% for deletion) than in the RNA-seq library (~0.25% for substitution and ~0.05% for deletion), whereas insertion errors occur at comparable rates in the two different types of libraries (Figure S2C). Like in the let-7 precursors, most substitutions and deletions in CLIP tags are found on G, whereas error frequencies on A, C, or U are similar to those in RNA-seq library (Figure S2D).

The base substitution in CLIP-seq data is a useful feature that allows mapping of protein binding sites at single-nucleotide resolution. We searched for LIN28A-binding sites on a transcriptome-wide level by using Shannon's information entropy, which quantifies the randomness of nucleotide composition at a given

**A Overview of LIN28A CLIP****B Specificity of immunoprecipitation****C CLIP sequence alignment nearby "GGAG" motif within pre-let-7g****D Secondary structure of pre-let-7g****E LIN28A binding and miRNA regulation****Figure 1. LIN28A CLIP-seq on Mouse Embryonic Stem Cells and Confirmation of the Interaction between LIN28A and let-7 Precursors**

- (A) LIN28A CLIP-seq workflow. See [Extended Experimental Procedures](#) for details. Abbreviations: CIAP, calf intestinal alkaline phosphatase; PNK, polynucleotidyl kinase.
- (B) Monoclonal anti-LIN28A antibody, 35L33G, specifically precipitates the LIN28A-RNA complexes. Left: western blot of LIN28A immunoprecipitates from siGFP- or siLin28a-treated A3-1 mESC lysates. Right: autoradiography of 5'-<sup>32</sup>P-labeled RNAs crosslinked to LIN28A in siGFP- or siLin28a-treated A3-1 cells. Crosslinking-induced mobility shift of LIN28A was not observed in western blotting because only a small fraction of the protein was crosslinked to RNA in CLIP-seq condition.
- (C) Sequences from LIN28A CLIP libraries aligned to the let-7g locus. The previously known binding site of LIN28A, the GGAG motif in the terminal loop of precursor let-7g, is marked with a red box. Each unique sequence is represented by a black horizontal bar with the number of reads indicated on the left. Mismatched sequences are shown in white letters. Site mutation rate is quantified by using Shannon's entropy and is shown at the bottom with blue bars. Less frequent tags (<7 reads) are omitted to improve visibility. See [Figures S2A and S2B](#) for other let-7 loci.
- (D) Predicted secondary structure of pre-let-7g with the GGAG motif shown in red. The red arrowheads indicate the binding sites of LIN28A detected in our CLIP experiments.
- (E) A scatter plot of LIN28A CLIP-seq enrichment levels (x axis) and miRNA level changes after *Lin28a* knockdown (y axis). The CLIP tag counts for miRNA loci were normalized by using the sum of miRNA microarray signals from the 5'- and 3'-arms. See also [Figures S1 and S2](#) and [Table S1](#).



**Figure 2. Identification of LIN28A Binding Motifs**

(A) LIN28A binding motif identified by CLIP-seq. Sequences harboring a mutation were aligned with the mutated nucleotide centered at zero. The information content (top) and positional frequency (bottom) are visualized by using WebLogo (Crooks et al., 2004). See also Figure S3C for a wider view and individual biological replicates.

(B) List of ten most frequently observed LIN28A-interacting hexamers from the CLIP-seq.

(C) Clusters of LIN28A-interacting hexamers. Area and color of each node represent relative enrichment of the hexameric sequence compared to the background frequency from RefSeq transcripts. Any two connected nodes differ by a single nucleotide. With few exceptions, hexamers can be grouped into three clusters: AAGNNNG, AAGNG(N), and (N)UGUG(N) where N = A, C, G, or U.

(D) List of ten most frequently observed 2 nt sequence pairs flanking LIN28A-bound hexamers. Note that the flanking sequences can base pair with the nucleotides at the opposite side of the hexamer.

position. It reduces false positives derived from single-nucleotide polymorphism or paralogous genes, which generally cause only one or two types of substitutions rather than all three possible substitutions (i.e., G to A, C, or T). The mutation rate quantified by Shannon's entropy correlates with the enrichment level of the given site in CLIP-seq, albeit modestly (Figure S3A), which further supports our notion that the mutated sites are derived from genuine LIN28A binding sites.

Altogether, we identified 516,259 putative binding sites with the confidence level of 0.1% false discovery rate (FDR) (Figure S3B and Table S3A), which means that a confident LIN28A binding site occurs at every 21.7 guanosines. On average, 38.5 sites were detected per mRNA. This unexpectedly large number of binding sites suggests that LIN28A binding is not restricted to a few sites on a small group of RNAs. Rather, LIN28A interacts with a large proportion of transcriptome.

### LIN28A Favors Single-Stranded Purine-Rich Motifs

To better understand LIN28A-RNA interactions, we further analyzed patterns around the potential LIN28A binding sites (Figures 2A and S3C). The frequently mutated G is preceded by two bases with strong preference for A or U and is followed by three bases favoring G or A. We also observed clear depletion of C across the hexamer. The resulting consensus sequence in the WebLogo is "AAGNGG" (Figure 2A). Unexpectedly, this consensus is different from the most frequently observed hexamer, "AAGGAG" (Figure 2B and Table S3B). We reasoned that the consensus may actually be a mixture of multiple distinct motifs. Indeed, similarity network of the LIN28A-bound hexamers revealed three distinct clusters, AAGNNG, AAGNG(N), and (N)UGUG(N) (Figure 2C). The "AAGNNG"-type hexamers account for the majority of LIN28A binding sites. This is consistent with the recent structural study that showed that the two zinc finger motifs of LIN28A recognizes "AGNNG" (Loughlin et al., 2012). The "AAGNG"-type pentamers also appear, albeit with lower frequency. Note that the least prevalent "UGUG" motif has not been reported before and may reflect a different mode of LIN28A binding. We further validated these three motifs by carrying out electrophoretic mobility shift assay (EMSA) with recombinant LIN28A protein and chemically synthesized RNA segments (Figure S4). Three representative RNAs containing the AAGGAG, AAGAG, and UGUG elements interacted with LIN28A, with UGUG being the weakest binder. Mutations to the G residue (that is frequently substituted in CLIP-seq) strongly reduced the affinity toward LIN28A, confirming that the identified motifs provide genuine binding sites for LIN28A.

Because, in the case of pre-let-7, the GGAG motif is invariably located in the terminal loop near the stem (Heo et al., 2009), and our EMSA data show that the stem of pre-let-7 enhances the binding of LIN28A (Figure S4); we examined the secondary structure around the binding motif. Interestingly, the sequences upstream of the hexamers are often complementary to the other

side of the hexamer (Figure 2D). The bases surrounding AAG NNG-type binding sites have clear propensity to form Watson-Crick (WC) pairs compared to randomly permuted sequences (Figure 2E, left). Moderate but significant enrichment was observed between bases that are up to seven nucleotides from the LIN28A-interacting hexamers, indicating that the stem surrounding the hexamer is around 7 bp in length. Similarly, the AAGNG pentamer is located in the terminal loop of a stem-loop structure with a slightly shorter stem (~5 bp) (Figure 2E, right). It is noted that some binding sites are in single-stranded RNA regions of more complex structures such as branched hairpins.

The preference of LIN28A for a hairpin structure is further supported by folding energy analysis (Figure S3D). Stronger structures (lower free energy) tend to appear more frequently in the CLIP libraries than those with weak structure (Figure S3D). The hairpin structure may help present the LIN28A-recognition element in a more accessible manner. Together, our analyses indicate that LIN28A preferably binds to a "AAG(N)NG" motif located in the terminal loop of a hairpin with a stem of 5–7 bp.

When we looked at all AAGGAG instances within a stem-loop structure in abundantly expressed mRNAs, almost half of the instances (41.7% with a cutoff of free energy of -6 kcal/mol) were experimentally detected in our CLIP-seq (Figure S3E). Other frequently observed motifs were also detected at similar rates when combined with low free energy cutoff (Figure S3E). This suggests that the motifs identified from our analyses may indeed be sufficient for LIN28A binding as long as the binding site is accessible to LIN28A in the cell.

### Messenger RNAs Are the Major LIN28A Targets

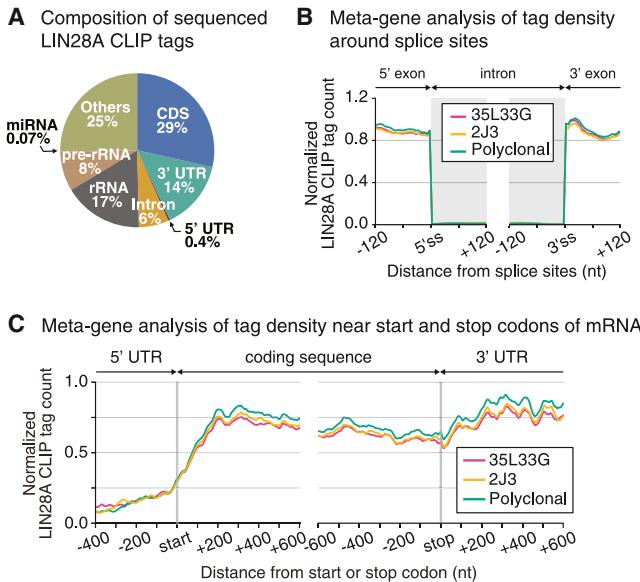
Next, we analyzed the types of RNAs that interact with LIN28A. Among the RNAs in the LIN28A CLIP-seq libraries, mRNAs constituted the most enriched class—over 42% of the tags were mapped to mRNAs, whereas miRNA loci made up only 0.07% of the sequenced reads (0.05% came from *let-7* loci) (Figure 3A and Table S4). The second most abundant class was ribosomal RNA, which occupied 17% of the CLIP tags. However, considering the overwhelming amount of rRNAs in cells (~80% of total RNA), our result indicates that mRNAs are the major interactors of LIN28A.

Within mRNAs, intronic regions are strongly depleted in CLIP tags (Figure 3B), indicating that LIN28A interacts with mature mRNAs after splicing is completed. We also analyzed the average density of the CLIP tags around start and stop codons (Figure 3C). The metagene analysis showed significant depletion of LIN28A binding in the 5' UTR compared to the coding sequences (CDS) and the 3' UTR.

### LIN28A Reduces Ribosome Occupancy without Affecting mRNA Abundance

To understand the functional significance of mRNA-LIN28A interaction, we first considered a possibility that LIN28A may

(E) Normalized relative frequency matrix of WC-pair co-occurrence around LIN28A-interacting sequences. Left and right panels represent the matrix for AAGNNG and AAGNG motifs, respectively. The actual motif patterns were reduced to AAGNHG and AAGNHH (where H = A, C, or U) to avoid interference between the two motifs. Sequences are aligned such that the crosslinked bases are centered at zero. Nucleotide frequency is presented above the triangular matrix. Colors in the triangular matrix, indicating WC-pair enrichment or depletion, are normalized to the background frequency estimated by permutation. See also Figures S3 and S4 and Tables S2 and S3.



**Figure 3. Messenger RNA as a Major Class of LIN28A Targets**

(A) Classification of LIN28A CLIP tags from the 35L33G (mAb) experiment. See also Table S4 for full description of classified annotations.  
 (B) CLIP tag density near splice sites (ss). The density was normalized per ss by mean alignment depth of exonic regions in the window (120 nucleotides from each of 5' ss and 3' ss) in RNA-seq.  
 (C) Density of LIN28A CLIP tags across different regions of mRNAs. Tag counts were normalized by RPKM in RNA-seq.

regulate the abundance of its targets. However, when we measured mRNA levels by RNA-seq after *Lin28a* knockdown, we could not find a significant correlation between CLIP tag enrichment (representing LIN28A binding; Figure 4A, x axis) and the changes in mRNA levels (Figure 4A, y axis). Therefore, LIN28A is unlikely to be involved in the stability control of its target mRNAs.

We next asked whether LIN28A regulates translation. LIN28A has previously been shown to be associated with polysomes in mouse teratocarcinoma P19 (Balzer and Moss, 2007) and in the differentiating mouse myoblast cell line, C2C12 (Polesskaya et al., 2007). Also, LIN28A is known to be a positive regulator of translation for mRNAs such as *Igf2*, cyclin A, cyclin B, histone 2a, and *Oct4* (Polesskaya et al., 2007; Qiu et al., 2010; Xu and Huang, 2009; Xu et al., 2009). Consistent with the previous observations, we found that LIN28A comigrates with polysome in mESCs (Figure 4B).

In order to monitor translational efficiency at the genomic level, we performed ribosome footprinting after *Lin28a* knockdown (Guo et al., 2010; Ingolia et al., 2009). Briefly, mESCs were transfected with control siRNA (siLuc) or siLin28a and were incubated for 48 hr prior to cycloheximide treatment and cell lysis (Figure 4C). The cell lysates were treated with RNase I to remove mRNAs that are not protected by ribosomes. The protected part of mRNA (the “footprints” of ribosomes) were retrieved by linker ligation and deep sequencing. Our experiment was successful judging from the characteristic enrichment of ribosome footprints in CDS (Figure S5A) and the three-nucleotide periodicity (Figure S5B).

Interestingly, there is a strong positive correlation between LIN28A interaction (Figure 4D, x axis) and ribosome density change (Figure 4D, y axis) ( $p = 1.01 \times 10^{-188}$ , Pearson’s correlation test). That is, ribosome occupancy of LIN28A-bound mRNAs tends to be higher in *Lin28a*-depleted cells compared to that in control siRNA-treated cells (Figures 4D, 4E, and S5C). This result indicates that LIN28A may negatively regulate the translation of its target mRNAs. This is surprising given that LIN28A is known as a positive regulator of translation (Polesskaya et al., 2007; Qiu et al., 2010; Xu and Huang, 2009; Xu et al., 2009). To validate the result, we carried out western blotting for genes whose ribosomal density increased or decreased after *Lin28a* knockdown without a change in the mRNA levels (Figure S5D and Table S5). The mRNAs coding LAMP1, EpCAM, and E-cadherin interact with LIN28A, whereas the other three mRNAs are not significantly enriched in CLIP. The changes in protein levels (monitored by western blotting, Figure S5D) are consistent with the ribosomal density changes (determined by ribosome footprinting, Table S5), indicating that LIN28A indeed targets mRNAs for translational repression.

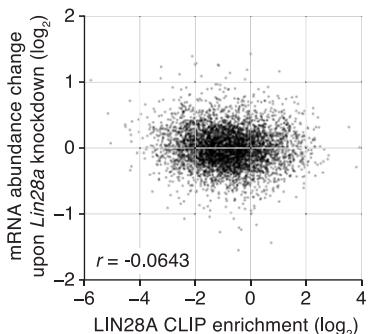
Of note, under our experimental conditions (knockdown for 48 hr), let-7 targets were not significantly affected in either mRNA abundance or ribosome occupancy (Figure S5E). Thus, LIN28A may act directly on translation of its target mRNAs independently of its effect through let-7.

### LIN28A Targets mRNAs Destined for Endoplasmic Reticulum

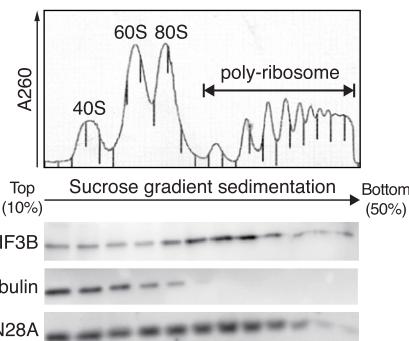
We next performed gene ontology (GO) analysis to understand functional consequences of LIN28A-mediated translational control (Figure 5A). This analysis revealed strong enrichment of several terms, with most biased GO terms related to cellular components. Interestingly, the majority of LIN28A targets are integral membrane proteins, secretory proteins, and ER or Golgi apparatus localized proteins (Figure 5A and Tables S5 and S6). These proteins are commonly translated in the ER-bound ribosomes and translocated into ER cotranslationally. In contrast, genes associated with nucleus and cytoplasm that are translated by free cytosolic ribosomes are underrepresented in the CLIP libraries (Figure 5A, x axis), and their translation is unaffected in LIN28A-depleted cells (Figure 5A, y axis). Considering the normalization issue caused by a shift of other groups, the modest changes associated with nuclear and cytoplasmic proteins are unlikely to be significant. Of note, CLIP enrichment of nucleosomal proteins is overestimated because histone mRNAs are depleted in RNA-seq data due to the lack of poly(A) tail. Taken together, LIN28A may preferentially bind to and control the mRNAs that are translated on ER.

Translation of most ER-associated mRNAs begins in the cytosol but halts once the signal sequence is synthesized. Signal-recognition particle (SRP) binds to the signal sequence and guides the mRNA-ribosome complex to the surface of ER where translation resumes and the nascent polypeptide is simultaneously translocated into the ER (Deshaies et al., 1991). We found that mRNAs encoding integral membrane proteins, which comprise the majority of the ER-associated translation, interact with LIN28A about four- to six-fold more frequently than other mRNAs do (Figures 5B and S6A, x axes;  $p = 3.08 \times 10^{-33}$ ,

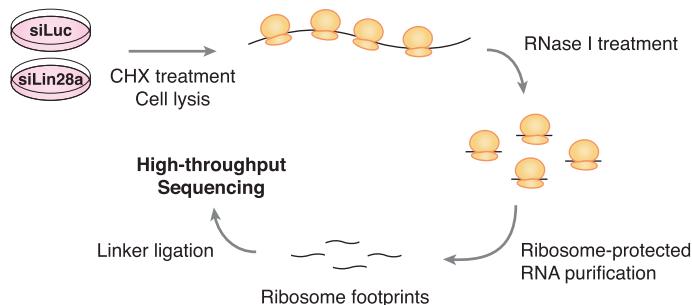
**A** Changes in RNA levels upon *Lin28a* knockdown



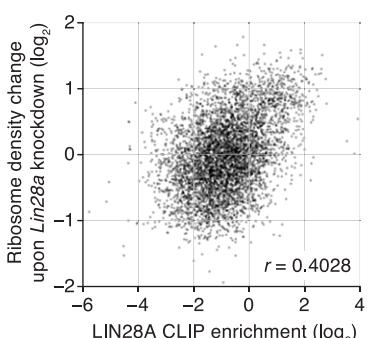
**B** Sucrose gradient fractionation and Western blot



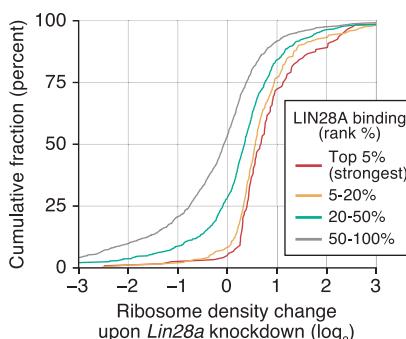
**C** Overview of ribosome footprinting workflow



**D** CLIP and ribosome footprinting upon *Lin28a* knockdown



**E** CLIP (binned) and ribosome footprinting



Mann-Whitney U test). Furthermore, their translation becomes more active compared to the other genes when *Lin28a* is knocked down, indicating that they are the major functional targets of LIN28A in translational control (Figures 5B and S6A, y axes;  $p = 6.97 \times 10^{-106}$ , Mann-Whitney U test). To validate the result, we carried out metabolic labeling after *Lin28a* knockdown and measured <sup>35</sup>S-methionine incorporation in rough ER (RER) microsome fraction (Hamilton et al., 1999) (Figure 5C). In LIN28A-depleted cells, protein synthesis rate was increased in RER fraction that is enriched with membrane proteins, indicating that LIN28A may indeed suppress ER-associated translation.

**Figure 4. Ribosome Occupancy Is Controlled by LIN28A-Binding**

(A) A scatter plot showing the correlation between CLIP tag enrichment and RPKM change upon *Lin28a* knockdown. Each point represents a single RefSeq transcript.

(B) Sucrose gradient fractionation of A3-1 mESC lysates. Absorbance at 260 nm was recorded while each fraction was collected. The locations of 40S subunit, 60S subunit, monosome, and poly-ribosome were determined by the characteristic pattern of UV absorbance. The protein levels of eIF3b, tubulin, and LIN28A in each fraction were determined by western blotting.

(C) Workflow of ribosome footprinting. CHX is an abbreviation for cycloheximide.

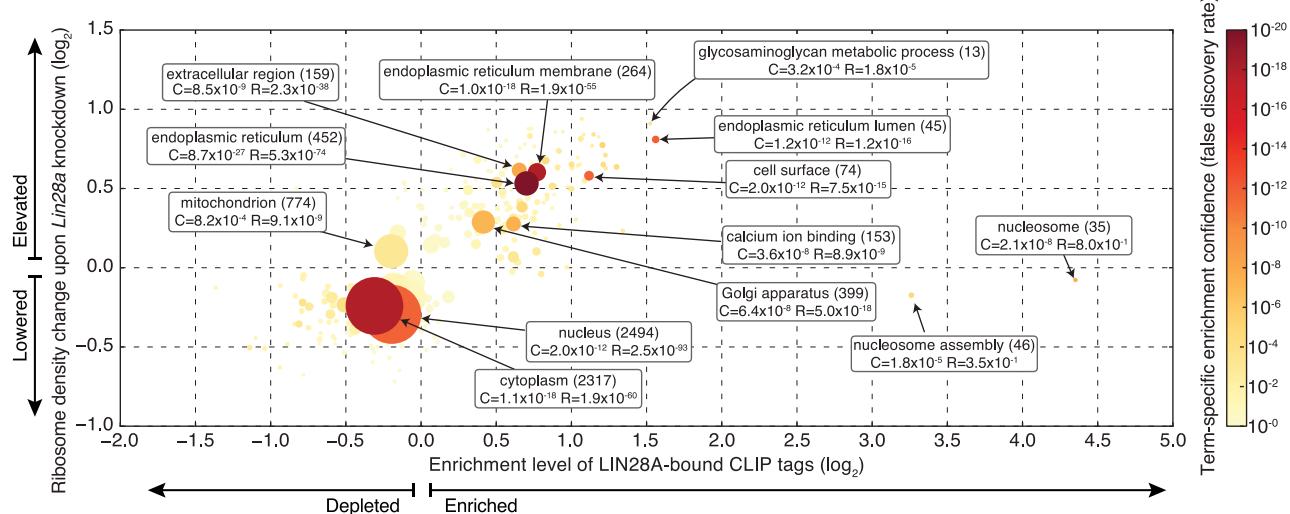
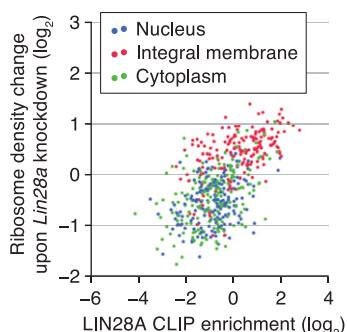
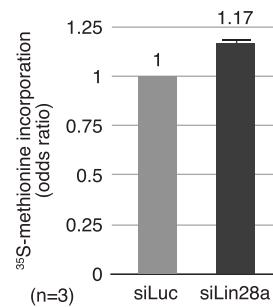
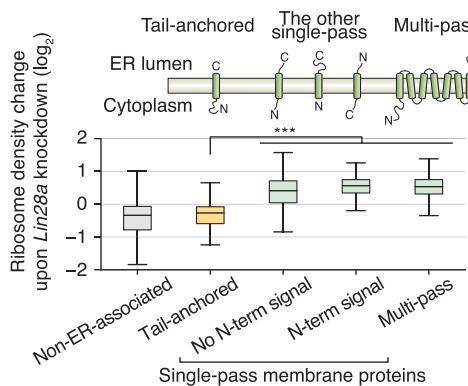
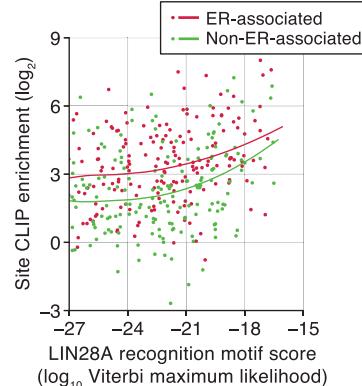
(D) A scatter plot showing the correlation between CLIP enrichment and the change of ribosome footprint density in LIN28A-depleted cells.

(E) Cumulative distributions of ribosome density changes (alternative representation of D). The patterns for strong binders are clearly different from those with lower CLIP tag enrichment ( $p = 2.04 \times 10^{-10}$  between top 5% and 20%-50% interval and  $p < 10^{-320}$  between top 5% and bottom 50%, Kolmogorov-Smirnov test). See Figure S5C for the results derived from different LIN28A antibodies. See also Table S4.

It is noted that mRNAs encoding tail-anchored transmembrane proteins behave differently from other ER-associated mRNAs. Unlike typical membrane proteins, these proteins are synthesized in the cytosol and translocated to ER posttranslationally (Kutay et al., 1993). Our data show that tail-anchored transmembrane proteins are neither enriched in LIN28A CLIP-seq nor increased in ribosome density in LIN28A-depleted cells (Figures 5D and S6B). This pattern is similar to other mRNAs that are translated by free cytosolic ribosomes. Thus, this result reinforces our conclusion that LIN28A differentially acts on ER-associated translation.

To understand the selectivity of the regulation, we asked whether ER-associated

mRNAs carry more LIN28A-recognition motifs than non-ER-associated mRNAs do. For this analysis, we first predicted the LIN28A-recognition sites on all mRNAs by using a hidden Markov model. To train the model, we used the binding sites experimentally identified from our CLIP-seq (see Figure S3F and Table S7 for the result and Extended Experimental Procedures for detail). Surprisingly, when we applied the model to predict potential LIN28A-interacting sites, non-ER-associated mRNAs harbor as many predicted sites as ER-associated mRNAs do (Figure S6C), indicating that the sequence and the structural features of ER-associated mRNAs are indistinguishable from those of non-ER-associated mRNAs. Nonetheless,

**A Gene ontology term-enrichment analysis for CLIP and ribosome profiling****B Protein localization****C Relative protein synthesis rate in RER microsomal fraction****D Protein topology****E Motif score vs. CLIP-seq enrichment****Figure 5. LIN28A CLIP-seq and Ribosome Footprinting Suggests ER-Associated Translation**

(A) GO enrichment analysis for CLIP-seq and ribosome footprinting data. Each circle represents a GO term. The color intensity indicates statistical significance of enrichment or depletion determined by Mann-Whitney U test. The size of the circle corresponds to the number of genes, whereas the coordinates indicate average CLIP tag enrichment (x axis) and ribosome density change (y axis). Balloons contain the number of genes (written in parentheses) and p values for CLIP-seq (C) and ribosome footprinting (R). Note that nucleosome-related terms are overestimated artificially in CLIP because of a normalization problem due to the lack of polyA tail in histone mRNAs.

(B) Specificity of LIN28A targeting. mRNAs encoding integral membrane proteins are enriched in CLIP libraries (x axis) and affected most significantly in ribosome density upon *Lin28a* knockdown (y axis). This plot shows 200 randomly chosen transcripts per term for better visibility, and the original plot is available in Figure S6A.

when we compared the sites with comparable motif scores, the sites on ER-associated mRNAs are detected by CLIP-seq more frequently than those on non-ER-associated mRNAs (Figures 5E and S6D;  $p = 3.46 \times 10^{-85}$ , two-way ANOVA). These results indicate that LIN28A-recognition element itself may not be sufficient to explain the preference for ER-associated mRNAs. In other words, there should be additional factor(s) that influences the interaction between LIN28A and its targets.

### LIN28A Is Localized in Peri-ER Area

We postulated that the preference for ER-associated mRNAs might be, at least in part, due to nonuniform subcellular distribution of LIN28A. Indeed, immunocytochemistry revealed that LIN28A is localized in the peri-ER area in mESCs. LIN28A signal is detected in the region surrounding ER that is marked by proteins containing ER-retention signal (KDEL) (Figure 6A, upper). Distribution of LIN28A is distinct from that of a cytosolic protein, GAPDH, which is more widely spread in the cytosol (Figure 6A, lower). For better visualization of the cytoplasmic compartment, we also examined ectopically expressed LIN28A in HeLa cells whose cytoplasm is larger than that of mESCs. Like the endogenous protein in mESCs, ectopically expressed LIN28A is enriched near ER in HeLa cells (Figure 6B).

We also carried out subcellular fractionation and found that LIN28A is present in the microsomal fraction derived from rough ER (RER) (Figure 6C) (Hamilton et al., 1999). Although cytosolic proteins such as tubulin and the cleaved fragment of ATF6 are not detected in the RER microsomal fraction, a significant amount of LIN28A is present. Calnexin and the uncleaved form of ATF6 were used as controls for ER membrane proteins. These results indicate that LIN28A may be bound to the cytosolic surface of RER on which ER-associated mRNAs are translated. Due to the colocalization, LIN28A may interact with ER-associated mRNAs more frequently than with other mRNAs.

### DISCUSSION

CLIP-seq technology allowed us to map the LIN28A binding sites on the genomic scale at single-nucleotide resolution. A recent study used a combination of coimmunoprecipitation (without crosslinking), oligo-dT enrichment, and deep sequencing, in order to find LIN28A targets in human ESCs (Peng et al., 2011). When we compared the published data with our results, we did not find significant overlap between the two data sets (Figure S7A). It is currently unclear why the data sets are so different. But given that the simple coimmunoprecipitation method is vulnerable to indirect interaction (Ule et al., 2005) and non-physiological binding of RNAs (Mili and Steitz, 2004) and that

genuine targets may be lost depending on washing conditions unless they are crosslinked to protein (Licalatosi and Darnell, 2010; Ule et al., 2005), our data may provide a more accurate list of LIN28A targets. In support of the notion, the LIN28A binding motif found in our CLIP-seq analysis is highly similar to that found in previous biochemical and structural studies on let-7 precursors. Moreover, we provide additional genome-wide functional evidence using ribosome footprinting, which demonstrates that the mRNAs identified in our study are indeed functional targets of LIN28A.

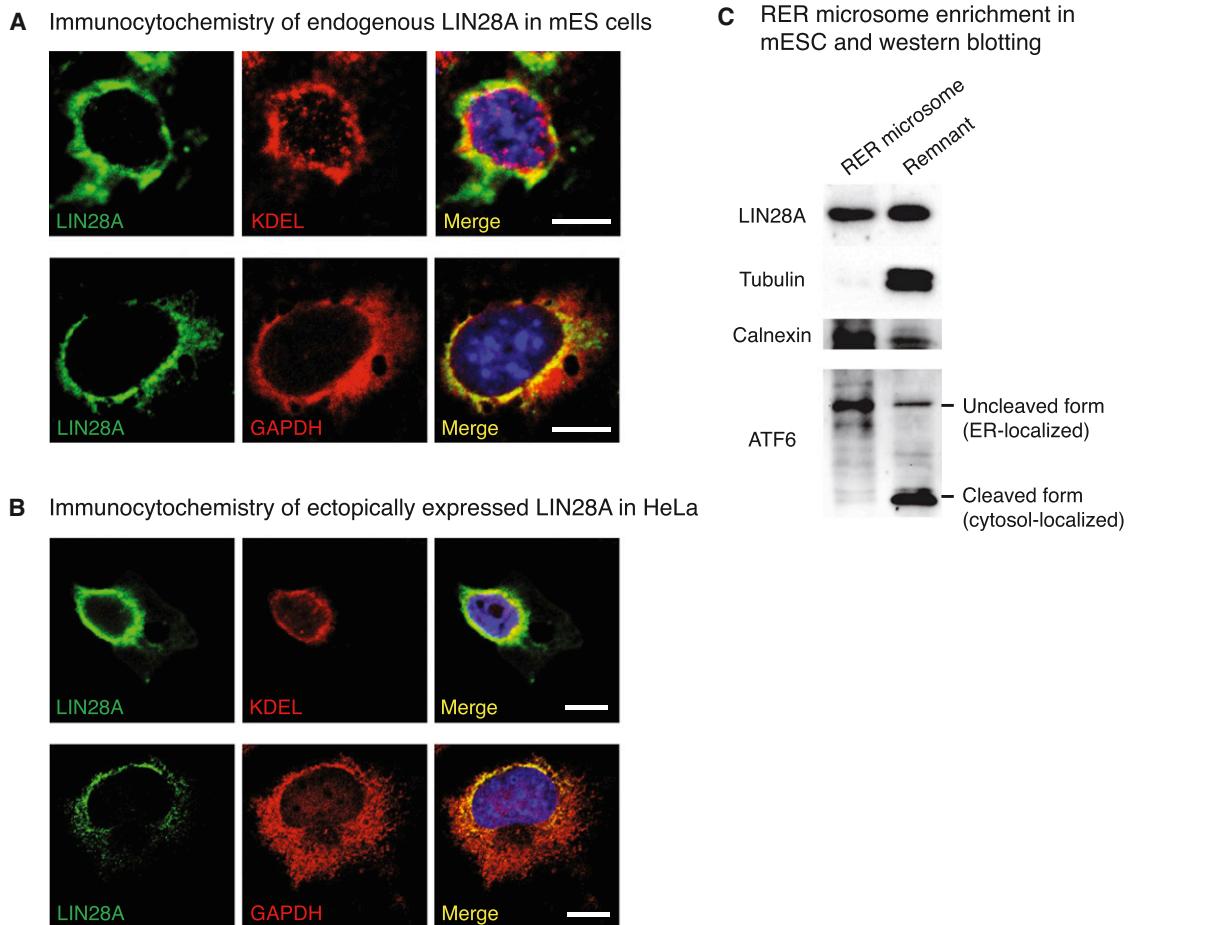
One of the key advantages of using the CLIP-seq approach is that one can map the binding sites by taking advantage of mutated sequences. Our study reveals the LIN28A-binding elements: AAGNNNG (with the most frequent sequence being AAG GAG) placed in the terminal loop of a small hairpin structure. According to the structural studies (Loughlin et al., 2012; Nam et al., 2011), the first G (that is most frequently mutated) forms a hydrogen bond with Lys160 in the zinc finger motif, which may be the actual crosslinking residue in our CLIP experiment. Another RNA binding domain, cold-shock domain (CSD), that is in the N-terminal part of LIN28A was reported to interact with a GNGAY motif located in the upstream of the GGAG motif of pre-let-7 (Nam et al., 2011). However, we did not find a significant enrichment of any motif upstream of the AAGNNNG element (data not shown). The CSD-RNA interaction may not have been captured by CLIP-seq partly because CSD interacts with RNA with very low sequence specificity (Heo et al., 2009; Nam et al., 2011). Due to the low complexity of the recognition element and the high abundance of LIN28A protein in ESCs, it is not surprising that we identify so many LIN28A binding sites in our CLIP-seq experiment (38.5 sites per mRNA). Our study suggests that LIN28A may bind to many mRNAs and dampen their usage rather than acting as an on-off switch for a small set of RNAs.

Our analyses indicate that the aforementioned motif may not be sufficient to predict which RNA is targeted by LIN28A because the actual interaction inside the cell is strongly influenced by the local concentration of the interactants (Figure 7A). Because of the localization of LIN28A in the peri-ER area, LIN28A may encounter ER-associated mRNAs more frequently than those distributed throughout the cytosol. Therefore, the selectivity at the transcript level may be dictated mainly by the intracellular localization of the transcript. The LIN28A-recognition element, on the other hand, determines which mRNA among the local pool interacts with LIN28A more frequently than other RNAs and which position on the given transcript LIN28A is loaded onto. Not mutually exclusively, peri-ER localization of LIN28A may be reinforced by its interaction with mRNAs

(C) Protein synthesis rate was measured by  $^{35}\text{S}$ -methionine incorporation in the fraction enriched with RER microsome (RER) and normalized against that in the remaining lysate. The result suggests that membrane protein synthesis was increased upon knockdown of *Lin28a* ( $n = 3$ ;  $p = 0.0032$ , one-tailed paired t test). The error bar represents SD of the normalized ratios from biological triplicates.

(D) Box plot of ribosome density changes of genes that are grouped according to protein topology. Tail-anchored membrane proteins remain unaffected, whereas other transmembrane proteins increase in ribosome density ( $p = 3.76 \times 10^{-23}$ , Mann-Whitney U test). Transmembrane proteins targeted to organelles other than ER were excluded from this analysis.

(E) Scatter plot of motif score of the binding sites (predicted by HMM using sequence and structure information) and their enrichment level in CLIP-seq relative to RNA-seq. Red dots represent CLIP-identified LIN28A binding sites on ER-associated mRNAs, whereas green dots represent those on non-ER mRNAs. To improve visibility, 200 sites with relatively high scores ( $\log_{10}$  maximum likelihood  $> -27$ ) were randomly chosen from each group. The complete representation is available in Figure S6D. See also Tables S5, S6 and S7.



**Figure 6. LIN28A Is Enriched in Peri-ER Area**

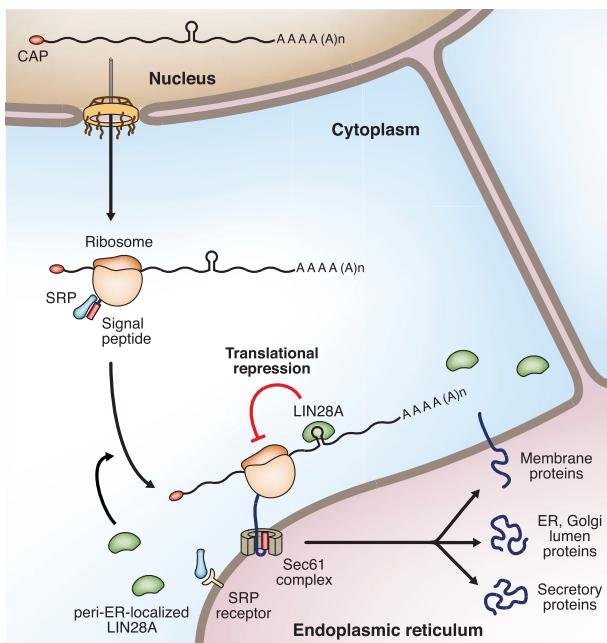
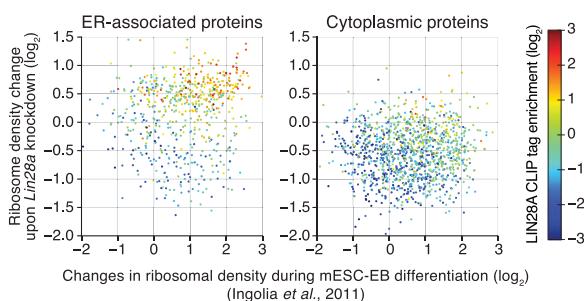
(A) Immunofluorescence of endogenous LIN28A and ER marker (KDEL peptide) in A3-1 mESCs. GAPDH was used as a control to show the distribution of a cytosolic protein. Scale bar, 10  $\mu$ m.  
 (B) Immunofluorescence of ectopically expressed LIN28A and endogenous ER marker (KDEL peptide) in HeLa. GAPDH was used as a control to show the distribution of a cytosolic protein. Scale bar, 10  $\mu$ m.  
 (C) Western blot analysis of the fraction enriched with microsomes derived from RER in A3-1 mESCs. The remaining fraction (remnant) was loaded as a control. Calnexin and uncleaved ATF6 are used as ER markers, whereas tubulin and cleaved ATF6 are representatives of free cytosolic proteins. LIN28A is significantly enriched in rough ER microsomal fraction compared to other cytosolic proteins.

concentrated near ER area. It remains to be investigated how LIN28A is transported and maintained in the peri-ER region. It will be interesting to identify and study the functions of the proteins that interact with LIN28A.

Ribosome footprinting provides evidence that LIN28A acts mainly as a suppressor rather than an enhancer of translation. Although we observe some genes that may be positively controlled by LIN28A, they constitute a minor fraction among LIN28A binders. It will be of interest to learn the mechanism by which LIN28A interaction leads to translational inhibition. Because interacting sites typically occur throughout coding sequences and 3' UTR at multiple locations, LIN28A may interfere with a step prior to or during elongation. To our knowledge, LIN28A is the first example of an mRNA-binding protein that represses ER-associated translation. Further investigation of the action mechanism of LIN28A may offer new insights into the

control of ER-associated translation and help us to uncover the regulatory machineries.

Our study reveals an unexpected role of a stem cell factor in the ER secretory pathway (Figure 7A). A recent report from the Weissman lab showed that translational efficiency of integral membrane proteins is significantly lower in undifferentiated ESCs compared to differentiating embryoid bodies (EBs) (Ingolia et al., 2011). Because *Lin28a* is downregulated in EBs, the data collectively suggest that LIN28A may be responsible at least partially for the global translational suppression of membrane proteins in ESCs. When we compared our results with the data from ESCs and EBs (Ingolia et al., 2011), we found significant overlap between the two data sets (Figures 7B, S7B and S7C). The genes that are translationally upregulated during EB differentiation tend to be enhanced translationally in LIN28A-depleted ESCs. The correlation was significant, especially for mRNAs that

**A** Schematic illustration for our suggested model**B** Translational control during mESC to EB differentiation

**Figure 7. Translational Suppression by LIN28A in Embryonic Stem Cells**

(A) Model for the regulation of ER-associated translation by LIN28A. ER-associated mRNAs are transported to ER via SRP pathway. Because LIN28A is localized in peri-ER region and it preferentially binds to the ER-associated mRNAs. The binding also depends on whether the mRNA contains a motif (AAGNNNG on a small hairpin) that is optimal for LIN28A binding.  
 (B) Comparison between the changes in ribosomal density upon *Lin28a* knockdown (y axis, our data) and those during mESC-EB differentiation (x axis; Ingolia et al., 2011). Each dot represents ribosome density change (position) and LIN28A CLIP enrichment level (color). Strong LIN28A binders (red dots) tend to be derepressed significantly in both conditions. See also Figure S7B for proteins targeted to nucleus and mitochondria.

were strongly bound to LIN28A (based on CLIP-seq data) (Figure 7B, red dots). The CLIP enrichment score correlates with translational changes accompanying the ES-EB differentiation ( $p = 6.85 \times 10^{-202}$ , Spearman's rank correlation test). Because none of the common pluripotency markers was decreased under our conditions (48 hr after transfection) (Figure S7D), the observed effects on translation in our experiments are likely to be

a consequence of LIN28A depletion rather than an indirect effect of differentiation.

Our data suggest that LIN28A may contribute to the global translational suppression of ER-associated mRNAs in undifferentiated stem cells. This type of regulation may reduce cell surface receptors and secretory proteins, which may influence cell signaling. One may also envision that global suppression of ER may be a way of redirecting cellular resources (such as energy and amino acids) to promote cell autonomous functions such as cell proliferation. It is currently unclear whether and how such regulation contributes to pluripotency and malignancy, but it will certainly be an intriguing arena to investigate.

## EXPERIMENTAL PROCEDURES

### CLIP-seq

LIN28A CLIP-seq was performed by using mouse embryonic stem cell A3-1 (Suzuki et al., 1997) as described in Extended Experimental Procedures. Two mouse monoclonal antibodies raised in-house (35L33G and 2J3) and one rabbit polyclonal antibody (Abcam, 46020) were used to construct three biological replicates of CLIP-seq libraries.

### Ribosome Footprinting

Ribosome footprinting libraries were generated from siLuc- or siLin28a-transfected A3-1 cells as described previously (Guo et al., 2010). Detailed process is described in Extended Experimental Procedures.

### RNA-seq

RNA-seq was performed with the RNA from untreated, siLuc, or siLin28a transfected A3-1 cells. mRNAs were poly-A enriched, fragmented to 35–55 bp, and sequenced.

### Sequence Analysis

The cDNA libraries described above were sequenced with Illumina Genome Analyzer II or Ilx. Sequence reads were aligned to mouse genome (*mm9* assembly) and RefSeq transcripts by using GSAP version 2012-01-11 (Wu and Nacu, 2010). Subsequent analyses were performed with in-house software as described in Extended Experimental Procedures. Complete source code and scripts written for this paper are available for downloading at <https://github.com/hyeshik/nrclip> under MIT-style license.

## ACCESSION NUMBERS

The NCBI Gene Expression Omnibus (GEO) accession number for the microarray and high-throughput sequencing data reported in this paper is accession number GSE37114.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and seven tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2012.10.019>.

## ACKNOWLEDGMENTS

We are grateful to Huili Guo for advice on ribosome profiling and Daehyun Baek for critical comments on the manuscript. We also thank Ahyoung Cho and Yun Cheng Chang for technical help and members of our laboratory for discussion. This work was supported by the Research Center Program (EM1202) of IBS (Institute for Basic Science); the BK21 Research Fellowships (J.C., H.C., S.C.K., and B.K.) from the Ministry of Education, Science and Technology of Korea; the TJ Park Postdoctoral Fellowship (Y.K.); and the National Honor Scientist Program (20100020415) through the National Research Foundation of Korea (NRF).

Received: July 17, 2012  
 Revised: September 8, 2012  
 Accepted: October 1, 2012  
 Published online: October 25, 2012

## REFERENCES

- Balzer, E., and Moss, E.G. (2007). Localization of the developmental timing regulator Lin28 to mRNP complexes, P-bodies and stress granules. *RNA Biol.* 4, 16–25.
- Balzer, E., Heine, C., Jiang, Q., Lee, V.M., and Moss, E.G. (2010). LIN28 alters cell fate succession and acts independently of the let-7 microRNA during neurogenesis in vitro. *Development* 137, 891–900.
- Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* 460, 479–486.
- Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190.
- Deshaines, R.J., Sanders, S.L., Feldheim, D.A., and Schekman, R. (1991). Assembly of yeast Sec proteins involved in translocation into the endoplasmic reticulum into a membrane-bound multisubunit complex. *Nature* 349, 806–808.
- Dreyfuss, G., Choi, Y.D., and Adam, S.A. (1984). Characterization of heterogeneous nuclear RNA-protein complexes in vivo with monoclonal antibodies. *Mol. Cell. Biol.* 4, 1104–1114.
- Granneman, S., Kudla, G., Petfalski, E., and Tollervey, D. (2009). Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc. Natl. Acad. Sci. USA* 106, 9613–9618.
- Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466, 835–840.
- Hagan, J.P., Piskounova, E., and Gregory, R.I. (2009). Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* 16, 1021–1025.
- Hamilton, R.L., Moorehouse, A., Lear, S.R., Wong, J.S., and Erickson, S.K. (1999). A rapid calcium precipitation method of recovering large amounts of highly pure hepatocyte rough endoplasmic reticulum. *J. Lipid Res.* 40, 1140–1147.
- Heo, I., Joo, C., Cho, J., Ha, M., Han, J., and Kim, V.N. (2008). Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol. Cell* 32, 276–284.
- 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.
- Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223.
- Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802.
- Kishore, S., Jaskiewicz, L., Burger, L., Hausser, J., Khorshid, M., and Zavolan, M. (2011). A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat. Methods* 8, 559–564.
- Kutay, U., Hartmann, E., and Rapoport, T.A. (1993). A class of membrane proteins with a C-terminal anchor. *Trends Cell Biol.* 3, 72–75.
- Licatalosi, D.D., and Darnell, R.B. (2010). RNA processing and its regulation: global insights into biological networks. *Nat. Rev. Genet.* 11, 75–87.
- Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., et al. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 456, 464–469.
- Loughlin, F.E., Gebert, L.F., Towbin, H., Brunschweiger, A., Hall, J., and Allain, F.H. (2012). Structural basis of pre-let-7 miRNA recognition by the zinc knuckles of pluripotency factor Lin28. *Nat. Struct. Mol. Biol.* 19, 84–89.
- Mili, S., and Steitz, J.A. (2004). Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* 10, 1692–1694.
- Moss, E.G., and Tang, L. (2003). Conservation of the heterochronic regulator Lin-28, its developmental expression and microRNA complementary sites. *Dev. Biol.* 258, 432–442.
- Moss, E.G., Lee, R.C., and Ambros, V. (1997). The cold shock domain protein LIN-28 controls developmental timing in *C. elegans* and is regulated by the lin-4 RNA. *Cell* 88, 637–646.
- 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.
- Newman, M.A., Thomson, J.M., and Hammond, S.M. (2008). Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA* 14, 1539–1549.
- Peng, S., Chen, L.L., Lei, X.X., Yang, L., Lin, H., Carmichael, G.G., and Huang, Y. (2011). Genome-wide studies reveal that Lin28 enhances the translation of genes important for growth and survival of human embryonic stem cells. *Stem Cells* 29, 496–504.
- Piskounova, E., Polytarchou, C., Thornton, J.E., LaPierre, R.J., Pothoulakis, C., Hagan, J.P., Iliopoulos, D., and Gregory, R.I. (2011). Lin28A and Lin28B inhibit let-7 microRNA biogenesis by distinct mechanisms. *Cell* 147, 1066–1079.
- Poleckskaya, A., Cuvelier, S., Naguibneva, I., Duquet, A., Moss, E.G., and Harel-Bellan, A. (2007). Lin-28 binds IGF-2 mRNA and participates in skeletal myogenesis by increasing translation efficiency. *Genes Dev.* 21, 1125–1138.
- Qiu, C., Ma, Y., Wang, J., Peng, S., and Huang, Y. (2010). Lin28-mediated post-transcriptional regulation of Oct4 expression in human embryonic stem cells. *Nucleic Acids Res.* 38, 1240–1248.
- Rybak, A., Fuchs, H., Smirnova, L., Brandt, C., Pohl, E.E., Nitsch, R., and Wulczyn, F.G. (2008). A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nat. Cell Biol.* 10, 987–993.
- Suzuki, H., Kamada, N., Ueda, O., Jishage, K., Kurihara, Y., Kurihara, H., Teruchi, Y., Azuma, S., Kadokawa, T., Kodama, T., et al. (1997). Germ-line contribution of embryonic stem cells in chimeric mice: influence of karyotype and in vitro differentiation ability. *Exp. Anim.* 46, 17–23.
- Thornton, J.E., and Gregory, R.I. (2012). How does Lin28 let-7 control development and disease? *Trends Cell Biol.* 22, 474–482.
- Ule, J., Jensen, K.B., Ruggi, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. *Science* 302, 1212–1215.
- Ule, J., Jensen, K., Mele, A., and Darnell, R.B. (2005). CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods* 37, 376–386.
- Viswanathan, S.R., Daley, G.Q., and Gregory, R.I. (2008). Selective blockade of microRNA processing by Lin28. *Science* 320, 97–100.
- Viswanathan, S.R., Powers, J.T., Einhorn, W., Hoshida, Y., Ng, T.L., Toffanin, S., O'Sullivan, M., Lu, J., Phillips, L.A., Lockhart, V.L., et al. (2009). Lin28 promotes transformation and is associated with advanced human malignancies. *Nat. Genet.* 41, 843–848.
- Wagenmakers, A.J., Reinders, R.J., and van Venrooij, W.J. (1980). Cross-linking of mRNA to proteins by irradiation of intact cells with ultraviolet light. *Eur. J. Biochem.* 112, 323–330.
- Wu, T.D., and Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26, 873–881.
- Xu, B., and Huang, Y. (2009). Histone H2a mRNA interacts with Lin28 and contains a Lin28-dependent posttranscriptional regulatory element. *Nucleic Acids Res.* 37, 4256–4263.
- Xu, B., Zhang, K., and Huang, Y. (2009). Lin28 modulates cell growth and associates with a subset of cell cycle regulator mRNAs in mouse embryonic stem cells. *RNA* 15, 357–361.

- 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.
- Zhang, C., and Darnell, R.B. (2011). Mapping *in vivo* protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat. Biotechnol.* 29, 607–614.
- Zhu, H., Shah, S., Shyh-Chang, N., Shinoda, G., Einhorn, W.S., Viswanathan, S.R., Takeuchi, A., Grasemann, C., Rinn, J.L., Lopez, M.F., et al. (2010). Lin28a transgenic mice manifest size and puberty phenotypes identified in human genetic association studies. *Nat. Genet.* 42, 626–630.
- 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.

# Supplemental Information

## EXTENDED EXPERIMENTAL PROCEDURES

### Cell Culture

Mouse embryonic stem cell line A3-1 was maintained on 0.1% gelatin-coated dishes with DMEM (WelGENE) containing 20% FBS (GIBCO), 1X nonessential amino acids (GIBCO), 1X nucleosides mix (SIGMA), 100 µM 2-mercaptoethanol (SIGMA), and 0.5% conditioned media from CHO cells secreting human leukemia inhibitory factor (LIF).

HeLa cells were cultured in RPMI (WelGENE) supplemented with 10% FBS (WelGENE).

### siRNA Transfection

A3-1 cells were reverse-transfected with siRNAs for 48 hr with Lipofectamine 2000 (Invitrogen). Briefly, 15 µl Lipofectamine and 20 µl of 20 µM siRNA were diluted individually in 500 µl Opti-MEM (GIBCO) and incubated for 15 min. Then, the two dilutes were combined and incubated for additional 15 min. The combined solution was then added to a 100 mm dish with complete growth medium containing ~2 million cells. The final concentration of siRNA was 50 nM. The sequences of the siRNAs are: siRNA sequences for ribosome footprinting and poly(A)<sup>+</sup> RNA-seq, siLuc 5'-UCGAAGUACUCAGCGUAAG dTdT-3'/5'-CUU retention ACGCUGAGUA CUUCGA dTdT-3'; siLin28a 5'-AUGUUUCUUCUUUUUGGCC dTdT-3'/5'-GGCCAAAGGGAAACAU dTdT-3' and siRNA sequences for microRNA microarray, siGFP 5'-UGAAUUAGAUGGCGAUGUU dTdT-3'/5'-GACAUCGCCAUCUAAUCA dTdT-3'; siLin28a 5'-GGGAGGAAGAGGAAGAGAU dTdT-3'/5'-AUCUCUUCCUCUCCUC dTdT-3'.

### MiRNA Microarray Analysis

Two biological replicates for control (siGFP) and *Lin28a* knockdown (siLin28a) samples were prepared as described above. Total RNA was extracted with TRIzol (Invitrogen) according to the manufacturer's instructions. Cyanine-3 (Cy3)-labeled RNA was prepared from 100 ng RNA by using Agilent miRNA Complete Labeling and Hyb Kit following the manufacturer's instructions. The Cy3-labeled RNA was hybridized to an Agilent mouse miRNA microarray 1.0 G4472A slide for 20 hr at 55°C in 20 rpm set Agilent Hybridization Oven. After the hybridization, slides were washed with Agilent GE Wash Buffers and scanned immediately on the Agilent Microarray Scanner with one color scan setting for 8 × 15k array slides (Dye channel is set to Green and Green PMT is set to 100% High and 5% Low). The data were analyzed by using Agilent Feature Extraction Software 9.5.3.1 with default analysis settings of miRNA\_v1\_95\_May07. The difference between normalized means of log<sub>2</sub>-transformed signals of the two samples were used to determine the changes in miRNA levels (shown in Figure 1E). MiRNAs with low signal (<3 in log<sub>2</sub>-scale) were excluded and the remaining 75 miRNAs were used for the analysis.

### Constructing LIN28A CLIP Library—UV Crosslinking and RNase A-Treated Lysate Preparation

A3-1 mESCs (~20 million cells) grown on a 100 mm dish were rinsed with 5 ml of PBS and irradiated with 600 mJ/cm of UV-B (Ule et al., 2003) by using CL-1000 ultraviolet crosslinker (UVP). UV-irradiated cells were harvested and resuspended in 1 ml of CLIP Wash buffer (1X PBS, 0.1% SDS, 0.5% deoxycholate, and 0.5% NP-40) (Chi et al., 2009). After incubating for 10 min, the lysate was treated with 20 µl of DNase I (Takara) for 5 min at 37°C and with 10 µl of 100 ng/µl RNase A for 10 min at 37°C. The lysate was then centrifuged at 40,000G at 4°C for 20 min to remove cell debris. After centrifugation, the supernatant was moved to a new 1.7 ml microtube.

### Constructing LIN28A CLIP Library—Immunoprecipitation

LIN28A antibody was attached to protein A sepharose by incubation with 20 µl of protein A sepharose resin (GE healthcare, 17-5138-01) in 500 µl of Wash buffer for 2 hr at 4°C. Three different LIN28A antibodies were used for immunoprecipitation of LIN28A-RNA complex: 5 µg of a rabbit polyclonal antibody (Abcam, 46020) and 20 µg of two different mouse ascites fluid containing monoclonal antibodies (35L33G, 2J3). For IP with the mouse monoclonal antibody, 20 µl of rabbit anti-mouse IgG (AffiniPure, 315-005-008) was used as a linker to protein A sepharose resin due to weak affinity of mouse immunoglobulin to protein A. The antibody attached resin was washed twice with Wash Buffer and incubated with the cell lysate for 2 hr at 4°C. After the incubation, the immunoprecipitates were washed twice with Wash buffer, twice with high-salt Wash buffer (5X PBS, 0.1% SDS, 0.5% deoxycholate, and 0.5% NP-40), and twice with PNK buffer (50 mM Tris-HCl pH 7.4, 10 mM MgCl<sub>2</sub>, 0.5% NP-40) (Chi et al., 2009).

### Constructing LIN28A CLIP Library—CIAP Treatment, 3' Linker Ligation, and PNK Treatment

Because the 3' end of the RNA carries a monophosphate due to RNase A mediated cleavage, it was removed by using calf intestinal alkaline phosphatase (CIAP) so that the RNA can be ligated to the 5'-end of adenylated 3' linker. Each immunoprecipitate was incubated in 80 µl of 1X reaction mixture including 3 µl of CIAP (Takara, 2250A) for 20 min at 37°C. After the CIAP treatment, the immunoprecipitates were washed twice with PNK+EGTA buffer (50 mM Tris-HCl pH 7.4, 20 mM EGTA, and 0.5% NP-40), once with PNK buffer, and once with PNK-NP40 buffer (50 mM Tris-HCl pH 7.4, and MgCl<sub>2</sub>) (Chi et al., 2009). To ligate 3' linker, each of the washed immunoprecipitates was incubated in 20 µl of 1X ligation reaction mixture containing 100 mM MgCl<sub>2</sub>, 20% PEG #8000 (Sigma), 100 pmol 3' linker (IDT, 5' rApp-CTGTAGGCACCATCAAT-ddC 3'), and 3 µl truncated T4 RNA ligase 2 (NEB, M0242) for 2 hr at 22°C. After the reaction, immunoprecipitates were washed twice with Wash buffer, once with high-salt Wash buffer, and thrice with PNK buffer.

Because the 5' end of RNA lacks a phosphate group due to RNase A cleavage, the RNAs were phosphorylated by incubating the immunoprecipitates in 1X reaction mixture containing 2 µl of PNK (Takara, 2021A) and 1 µl of [ $\gamma$ -<sup>32</sup>P] ATP (Perkin Elmer) for 5 min at

37°C to label the RNA-protein complexes. After the labeling, 10 µl of 1 mM ATP was added and incubated for 10 min at 37°C. After PNK treatment, immunoprecipitates were washed 3 times with PNK buffer and mixed with 2X LDS loading dye (Invitrogen, NP0007) containing 50 mM DTT. The samples were incubated at 70°C for 10 min to detach RNA-protein complexes from the beads.

#### **Constructing LIN28A CLIP Library—NuPAGE, Transfer, and Proteinase K Treatment**

NuPAGE is required to avoid RNA degradation via alkaline hydrolysis during size-dependent separation of RNA-protein complexes (Chi et al., 2009). Supernatant was removed from the beads and run on Novex 4%–12% NuPAGE Bis-Tris gels (Invitrogen) in MES running buffer (Invitrogen) for 40 min at 200 V and transferred to Hybond C extra-western membrane (Amersham) in NuPAGE transfer buffer (Invitrogen) for 2 hr 30 min at 80 V. After the transfer, the membrane was rinsed in RNase-free water and exposed to phosphor-imaging plate (Fujifilm) and read with the BAS-2500 system (Fujifilm) to monitor the distribution of the RNA-LIN28A complexes. Since the RNA-LIN28A complexes were located between 30 kDa and 80 kDa, the corresponding region of the membrane was carefully excised with a scalpel. The excised membrane piece was treated with 10 mg/ml of proteinase K (Macherey-Nagel) in 200 µl of PK buffer (100 mM Tris-HCl pH 7.4, 50 mM NaCl, 10 mM EDTA) for 20 min at 37°C and incubated for 20 min at 50°C in 200 µl of PK+Urea buffer (100 mM Tris-HCl pH 7.4, 50 mM NaCl, 10 mM EDTA, and 7M urea).

#### **Constructing LIN28A CLIP Library—RNA Purification and 5' Linker Ligation**

To purify RNA, 400 µl of acidic phenol:chloroform (Ambion) was added and incubated for 1 hr at 37°C and 800 rpm in Thermomixer Comfort (Eppendorf). After the incubation, the sample was vigorously vortexed and centrifuged at 12,000 rpm for 8 min at 4°C. Then, 0.75 µl glycogen (Ambion), 50 µl 3M NaOAc (Ambion), and 1 ml of 1:1 ethanol:isopropanol were added to the aqueous phase of the sample, and RNAs were precipitated overnight at –80°C. RNA pellets were washed twice with 75% ethanol and resuspended in 5 µl of RNase-free water. To ligate the 5' linker, 1 µl of T4 RNA ligase (Takara, 2050A), 1.3 µl of 10X buffer, 1 µl of 10 mM ATP, 3 µl 50% PEG #8000 (Sigma), 5 pmol of 5' linker (5'-GUUCAGAGUUCUACAGUCCGACGAUC-3'), and 1 µl of SUPERase-In (Ambion) were added to each RNA sample and incubated for 16 hr at 16°C. The RNAs were then purified by phenol extraction, precipitated, and resuspended in 5 µl of RNase-free water.

#### **Constructing LIN28A CLIP Library—RT-PCR and High-Throughput Sequencing**

Each of the ligated RNA sample was mixed with 5 µl of 2.5 mM dNTP mix (Takara) and 2 µl of 10 pmol/µl RT primer (5'-ATT GATGGTGCCTACAG-3') and incubated for 10 min at 65°C. The mixtures were then incubated for 2 min at 4°C. 1 µl of SuperScript II RT enzyme (Invitrogen, 18064), 4 µl 5X buffer, 2 µl 0.1M DTT, 1 µl RNasin (Promega) were added to each mixture and incubated for 1 hr at 42°C. After RT reaction, the samples were incubated for 15 min at 70°C to inactivate the enzyme. To amplify the cDNA for Illumina DNA sequencing, 4 µl from a total of 20 µl RT sample was mixed with 0.25 µl of Phusion HF polymerase (NEB, F-530L), 5 µl 5X buffer (NEB), 2 µl 2.5 mM dNTP (Takara), 0.3 µl of 10 µM 5' primer (5'-AATGATAACGGCGACCACCGACAGGTTACAGTCGA-3'), 0.3 µl of 10 µM 3' primer (5'-CAAGCAGAACGGCATACGAATTGATGGTGCTACAG-3'), and 13 µl of distilled water. This PCR mixture was run in T1 Thermocycler (Biometra) for 25 cycles with a program of 98°C for 10 s, 60°C for 30 s, and 72°C for 30 s. Amplified cDNA libraries were purified by phenol extraction and sequenced by using Illumina Genome Analyzer IIx for 78 cycles.

#### **Sequence Processing and Alignment of CLIP Libraries**

The first few steps in sequence analysis were done by using Assaf Gordon's FASTX-Toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)). First, the 3' adaptor sequences were removed from reads by using fastx\_clipper. The rest was trimmed from the 3'-end so that the remaining reads have Phred quality of 25 or higher. After clipping and trimming, reads of 20 nt or longer were collapsed to generate a set of unique sequences. The sequences were aligned to abundant contaminant sequences (Illumina adaptor/primer sequences and mouse ribosomal DNA complete repeating unit, GenBank accession BK000964.1) with GSNAp version 2012-01-11 (Wu and Nacu, 2010) with 10% mismatch rate. Filtered reads that do not match to any contaminant and have sufficient sequence complexity (Shannon's entropy, at least 0.7 for mononucleotide, 1.5 for dinucleotide) were aligned to UCSC (the University of California, Santa Cruz) mm9 genome assembly with GSNAp version 2012-01-11 with options of 10% mismatch rate, no terminal clipping, and splice site annotations from RefSeq (downloaded from UCSC Genome Browser on Aug 19, 2011). The alignment results were filtered to leave only the single best hit with minimum edit distance (up to two edits) to obtain a set of single-hit reads. Those with multiple best hits were ignored as repetitive sequences.

#### **Sequence Annotation and Classification of CLIP Reads**

The alignments were annotated with RefSeq, RepeatMasker (retrieved from UCSC Genome Browser on Aug 19, 2011), miRBase release 18 (retrieved from [miRBase.org](http://miRBase.org) on Dec 7, 2011), Rfam (retrieved from [rfam.sanger.ac.uk](http://rfam.sanger.ac.uk) on Aug 25, 2011), and GtRNAdb (retrieved from [gtrnadb.ucsc.edu](http://gtrnadb.ucsc.edu) on Aug 25, 2011) by using intersectBed of BEDTools (Quinlan and Hall, 2010). A representative class for a given read was determined as the first matching class from all annotations for all alignments for the read in the following priority: miRNA, rRNA, tRNA, Mt-tRNA, snoRNA, scRNA, srpRNA, snRNA, RNA, ncRNA, misc\_RNA, Cis-reg, ribozyme, RC, IRES, frameshift\_element, LINE, SINE, Simple\_repeat, Low\_complexity, Satellite, DNA, LTR, CDS, 3'UTR, 5'UTR, intron, Other, Unknown. The annotated representative classes were combined with read counts of previously removed sequences in the first contaminant

filtration, and used for CLIP tag classification statistics (Figure 3A). For subsequent analyses, the reads classified as rRNA or tRNA were excluded and the rest was used. The alignments for filtered reads were converted to bam format and visualized with the UCSC Genome Browser (Figures 1C, S2A and S2B). These resource files are available at <http://www.narrykim.org/s/cho-lin28a-2012/>.

### Visualization of Predicted Secondary Structures

Predicted secondary structure of pre-let-7g (Figure 1D) was drawn by using UNAFold (<http://mfold.rna.albany.edu/?q=DINAMelt/software>). Expected DROSHA and DICER cleavage sites of let-7 (Figures 1C, 1D, S2A and S2B) were marked according to miRBase release 18 (Kozomara and Griffiths-Jones, 2011).

### Crosslinking-Induced Reverse-Transcription Error Analysis

Mismatches in alignment can arise from the differences between the mm9 reference genome and the A3-1 cell genome as well as from reverse transcription errors that often occur near splice sites. These mismatches can substantially increase the number of false positives as well as false negatives. To overcome these problems, RNA-seq was performed with A3-1 cells without siRNA treatment, and the genome-aligned reads were probed for single-nucleotide polymorphisms (SNPs). The data were used to correct for the SNPs ( $\geq 10$  reads at a position,  $\geq 90\%$  single-type mutation—such as C-to-T or A-to-G) detected in the nonredundant reduced set of RefSeq sequences that retain the longest isoforms among overlapping transcripts. The filtered reads (no rRNA/tRNA/adapters) were aligned to the SNP-corrected nonredundant RefSeq sequences with GSAP version 2012-01-11 (Wu and Nacu, 2010) with options allowing 5% mismatches and no terminal clipping. Multi-hits were resolved by the same criterion applied in the alignment to the genome.

Statistics for error profiles (Figures S2C and S2D) were collected from all positions of nonredundant RefSeq without weighting by positional read depth. Crosslinking-induced reverse-transcription error score (CRES) was determined by using Shannon's entropy:

$$C = - \sum_n p(n) \log_2 p(n)$$

where  $C$  denotes CRES and  $n$  is a nucleotide or a deletion. Insertions were ignored from the calculation of CRES because they were rare compared to substitutions or deletions and not enriched in CLIP libraries compared to the RNA-seq library. Significance of CRES was determined by using permutation-based FDR estimation as (Zhang and Darnell, 2011) developed.

Since FDR also depends on alignment depth, it is difficult to estimate the exact FDR of a particular data set. Hence, we initially used CRES for low depth position ( $\leq 130$  reads), applied LOESS smoothing ( $\alpha = 0.5$ ) for medium depth positions ( $< 250$  reads), and set FDRs for high depth positions ( $\geq 250$  reads). Our analyses portraying distribution of binding sites (e.g., Figures S3A, S3E, and S6D) used FDR cut-off of 0.001. However, application of different CRES according to alignment depths introduces bias to highly expressed transcripts in confident binding sites pool, so we used constant CRES cut-offs (e.g., 0.8) together with depth cut-offs (e.g., 50 reads) for other analyses. For example, in the 35L33G CLIP library, FDR for positions of CRES cut-off 0.8 and depth cut-off 50 is estimated to be 0.001.

### Binding Site Motif Analysis

To analyze enriched sequence motif near LIN28A binding sites, we first extracted neighboring sequences with cut-offs of 0.8 for CRES and 50 for read depth. The 20 nt long flanking sequences were used as inputs to WebLogo version 3.2 (Crooks et al., 2004) with the default options for RNA sequence analysis.

Using the same threshold, hexamer sequences around binding sites ( $-2$  to  $+3$ ) were analyzed for binding site pattern for each sample. Background frequency of a given hexamer was calculated from the nonredundant subset of RefSeq defined above. Hexamer enrichment level was defined as  $\log_2$  ratio between the frequency of the binding site and that of background hexamer, and Fisher's exact test was used to determine the statistical significance of the enrichment.

Levenshtein distances between any two hexamers that are enriched ( $\log_2$  enrichment  $> 0$ ) in LIN28A binding sites were used for similarity network analysis. To visualize network topology and enrichments, NetworkX (<http://networkx.lanl.gov/>) was invoked with hexamers as nodes and any two connected nodes differ by a single nucleotide. Both area and color of nodes reflect the enrichment level where area is proportional to linear enrichment level and color gradient is proportional to  $\log_2$  enrichment level in dynamic range of  $2^3$  -  $2^7$ .

### Binding Site Secondary Structure Preference Analysis

Secondary structure prediction software was initially used to check for the existence of a specific structure nearby LIN28A binding sites. However, the predictions were too sensitive to window width, resulting in high false positive rates for the occurrence of a hairpin structure for small window width. To determine unbiased likelihood of base pairing near LIN28A-interacting sequences, Watson-Crick (WC) pair co-occurrence frequencies of two positions around LIN28A-binding sites were compared against randomly permuted sequences generated with positional base probabilities. Binding sequences were permuted for 1,000 times by shuffling all bases in the same position of different sequences and pooled the WC-pair co-occurrence between two positions on each permutation. The

enrichment level was calculated as  $\log_2$  ratio between observed WC-pair co-occurrence frequency near LIN28A-binding sequences and the background determined by permutation.

#### **Binding Site-Level Quantification of Enrichment for CLIP Libraries**

To determine the enrichment of a given binding site, we compared alignment depths of the binding site from one of each CLIP libraries with that from RNA-seq library. Enrichment levels were calculated as  $\log_2$  ratio between reads in the CLIP libraries (plus a pseudocount of one) and those in the RNA-seq library (plus a pseudocount of one). In the analysis (Figure S3A), binding sites with reads less than 20 in RNA-seq were excluded to reduce noise. When multiple binding sites exist within 70 nt window, binding sites with low CRES were also excluded from the analysis to focus on major interaction spots with single LIN28A-interacting local structure. Local folding energy in the analysis of correlations among crosslinking susceptibility (mutation rate), enrichment level, and secondary structure (Figures S3D and S3E) was predicted by using 18-mer sequences from -8 to +9 positions by RNAfold in the ViennaRNA package (Hofacker et al., 1994) with -noCloseGU option at 37°C.

#### **Simple Sequence and Local Structure Scanning of mRNAs for LIN28A Binding Sites**

To understand the extent to which the highly enriched hexamer sequences (e.g., AAGGAG and AAGGUG) contribute to LIN28A binding, all instances of a specific hexamer in all expressed transcripts were probed (>30 reads of local alignment depth at the third base "G" position in RNA-seq). They were compared to the binding sites identified from our LIN28A CLIP (FDR < 0.01), and the overlapping fractions were calculated to denote specificity (Figure S3E). The analysis of binding site predictability was performed for reduced nonredundant data sets. Different free energy limits (from -6.0 to 0.0 kcal/mol as the upper limit) was used to restrict the sites to those with secondary structures. The free energy was predicted with RNAfold in the ViennaRNA package (Hofacker et al., 1994) with -noCloseGU option at 37°C for 18 nt windows (-8 to +9 positions from the third base "G").

#### **Hidden Markov Model Training for Recognition Motif and Scanning mRNA Sequences**

The above analysis showed that sequence and secondary structure may provide sufficient features for predicting LIN28A binding sites. To use the contextual information as well as sequence information, hidden Markov model (HMM) was trained to learn LIN28A binding elements by using the General Hidden Markov Model (GHMM) library (<http://www.ghmm.org>). Topology of the model was designed with one start marker state (S) and eighteen match states (M01 to M18). The possible observations were given as a set including S for start marker, A, C, G, U for unpaired matches, (" and ') for matches to paired bases. The initial probability matrices are shown in Table S7. Training set was prepared from 18-mer sequences that are adjacent to high-confidence binding sites (-8 to +9 from the frequently mutated "G"; CLIP depth  $\geq$  50, CRES > 1.2). Secondary structures of the adjacent sequences were annotated with RNAfold from the ViennaRNA package (Hofacker et al., 1994) with -noCloseGU option at 37°C. The training set was put to Baum-Welch algorithm from the GHMM library and optimized for 1,000 iterations. The resulting probability matrices are shown in Table S7. The HMM likelihood of a given 18-mer sequence to contain at least one LIN28A binding site was calculated with Viterbi algorithm implementation from the GHMM library (Figures S3F, S6C and S6D).

#### **Metagene CLIP Tag Density Analysis**

CLIP tag density within mRNAs (Figure 3C) was calculated for nonhistone mRNAs expressed more than 15 RPKM in RNA-seq. We excluded genes whose names are starting with "Hist" to avoid normalization artifact by poly(A) enrichment. Local tag density in a single transcript was computed from tag alignment depths in one of CLIP libraries or RNA-seq and divided by RPKM from RNA-seq. Local densities from one of the CLIP libraries were normalized by tag densities from RNA-seq, which was divided by transcript-level RPKM to correct coverage biases in 5' UTR and sequencing preference biases in 3' UTR. The normalized tag densities were normalized again by total read count of a respective CLIP library and smoothed by using moving average with 13 nt windows.

CLIP tag density around splice sites was calculated by using splice sites annotated in RefSeq (downloaded from the UCSC Genome Browser on Aug 19, 2011). Non-histone mRNAs with at least four RNA-seq reads with 5' ends that lie within a window of -120 to +120 nt around the splice sites are used. Number of the 5'-end of the reads in a specific position (instead of alignment read depths) was used because sequence bias near splice sites was strong, and the bias influenced the read length due to sequence specificity of RNase A. The 5'-end counts around the 5'- and 3'-splice sites were normalized by average 5'-end densities in the window in RNA-seq. The normalized densities were grouped into 5 nt bins and the average densities of the given bins are presented in Figure 3B.

#### **Sucrose Gradient Fractionation of Polyribosome**

Eleven ml of 10%–50% continuous sucrose gradient in 50 mM MOPS-NaOH at pH 7.4, 150 mM NaCl, 15 mM MgCl<sub>2</sub>, 100 µg/ml cycloheximide was prepared one day prior to cell harvest. A3-1 cells (200 million cells per dish) were washed with PBS and harvested and transferred to a 1.7 ml tube. Cells were resuspended in 1 ml of 50 mM MOPS-NaOH at pH 7.4, 150 mM NaCl, 15 mM MgCl<sub>2</sub>, 0.25% Triton X-100, 1 mg/ml Heparin, 100 µg/ml cycloheximide with 5 µl protease inhibitor cocktail (Calbiochem) and 50 µl RNasin (Promega). Cells were incubated in the buffer for 10 min at 4°C, and the cell lysate was centrifuged at 12,000 rpm for 10 min and 4°C to remove debris and nuclei. The supernatant was loaded to the top of the continuous sucrose gradient and centrifuged at 36,000 rpm for 2 hr 30 min at 4°C with SW41ti rotor (Beckman). After centrifugation, 1 ml fractions were collected from top to bottom of the

gradient by using BioLogic LP system and fraction collector (BioRad) with UV absorbance at 260 nm. Each of the fractions was mixed with 250  $\mu$ L of 5X SDS loading dye and boiled for 5 min. 100  $\mu$ L of each fraction was used for western blotting.

### Western Blot Analysis

Samples were resolved on 10% SDS-polyacrylamide gels or 4%–12% NuPAGE gels (Invitrogen) and transferred to Immobilon-P transfer membrane (Millipore) or Hybond C extra-western membrane (Amersham). Primary antibodies used in this study were rabbit anti-LIN28A (Abcam; 1:1,000), rabbit anti-eIF3B (Abcam; 1:1,000), rabbit anti-alpha tubulin (Abcam; 1:3,000), mouse anti-calnexin (BD Biosciences; 1:1,000), mouse anti-ATF6 (Abcam; 1:500), mouse anti-MCM7 (Santa Cruz; 1:300), rat anti-LAMP1 (DSHB; 1:1,000), rat anti-EpCAM (DSHB; 1:1,000), and rat anti-E-cadherin (SIGMA; 1:1,1000).

### Electrophoretic Mobility Shift Assay

The 5'-end  $^{32}$ P-labeled synthetic RNA oligo (10 fmol, 0.5  $\mu$ L) of  $1 \times 10^4$  –  $1 \times 10^5$  cpm is mixed with 2  $\mu$ L H<sub>2</sub>O and 2.5  $\mu$ L 6X loading dye. It is mixed with another mixture composed of 1  $\mu$ L 10X EMSA buffer, 0.5  $\mu$ L 100 mM DTT, 0.2  $\mu$ L RNase inhibitor (40 U/ $\mu$ L, Takara), rLIN28A, and 2.8  $\mu$ L H<sub>2</sub>O. After 10 min of the incubation at room temperature, the 11  $\mu$ L solution is loaded into native gel. The gel is run at 150 V for 4 hr at 4°C. The gel is directly exposed to Phosphor Imaging Plate (Fujifilm) and is read with the BAS-2500 system (Fujifilm) for quantification. The 10X EMSA buffer is composed of 150 mM HEPES (pH 7.4), 1 M KOAc, 20 mM Mg(OAc)<sub>2</sub>, and 2% BSA. The native gel is composed of 5% polyacrylamide (40:1), 40% glycerol and 1× TBE. The 6× loading dye is composed of 0.1% bromophenol blue, 0.1% xylene cyanol and 50% glycerol. The synthetic RNA oligo were labeled at the 5'-end with T4 polynucleotide kinase (Takara, 2021A).

### Construction of Ribosome Footprinting Library

Preparation of ribosome footprinting libraries of siLuc and siLin28a treated A3-1 cells was carried out as previously described (Guo et al., 2010) with the following modifications. Two 100 mm dishes and three 100 mm dishes with A3-1 cells were respectively treated with siLuc or siLin28a for 48 hr. For Illumina high-throughput sequencing, 5' linker (IDT, 5'- GUUCAGAGUUCUACAGUCCGACGAUC-3') and 3' linker (IDT, 5' rApp-CTGTAGGCACCATCAAT-ddC 3') were used for 5' and 3' ligation, respectively. The linker ligated RNA was reverse-transcribed and PCR-amplified with the same manner as described in CLIP-seq library preparation, but with a different RT primer (5'-CAAGCAGAAGACGGCATAGCA-3') and a PCR 3' primer (5'-CAAGCAGAAGACGGCATACGA-3'). Sequencing of the ribosome footprinting libraries were performed by using Illumina Genome Analyzer II with 36 base reads per ribosome footprint tag.

### Construction of RNA-seq Library

RNA-seq libraries were prepared as previously described (Guo et al., 2010) with the following modifications. 300 ng of oligo-dT enriched RNAs were fragmented by using NEBNext Fragmentation buffer (NEB), and 35–55 bp of RNA fragments were purified from 10% urea-polyacrylamide gel. We used the same 5' and 3' linker, RT-primer, and PCR primers that were used for RPF library preparation to minimize potential biases. These linkers and primers were also used for CLIP-seq library preparation. The linker-ligated RNA was reverse-transcribed and PCR-amplified following the same protocol described for CLIP-seq and RPF library preparations, except that cDNAs were amplified for 12 cycles in PCR. Sequencing of the RNA-seq libraries were performed by using Illumina Genome Analyzer II with 54 base reads per RNA-seq tag.

### Sequence Processing and Alignment for Ribosome Footprinting and RNA-seq Libraries

For the control RNA-seq sample, we followed the procedure that we used for the CLIP libraries. Processing of the other libraries was also performed with FASTX-Toolkit. First, the sequences were trimmed from the 3'-end so that all final remaining read bases have Phred quality of 25 or higher. After clipping and trimming, reads that are 20 nt or longer were collapsed into a set of unique sequences. The sequences were aligned to abundant contaminant sequences (Illumina adaptor/primer sequences and mouse ribosomal DNA complete repeating unit, GenBank accession BK000964.1) with GSNAP version 2012-01-11 (Wu and Nacu, 2010) with 5% mismatch rate. Sequences that do not match to any contaminant and have sufficient sequence complexity (Shannon's entropy, at least 0.7 for mononucleotide, 1.5 for dinucleotide) were aligned to mm9 genome assembly with GSNAP version 2012-01-11 with 5% mismatch rate and no terminal clipping. Then, brief annotations of aligned positions were retrieved in the same way used for annotation of the CLIP libraries. Except for the reads annotated as rRNA and tRNA, the sequences were trimmed to 27 nucleotides to leave 5' parts and aligned to nonredundant RefSeq transcript sequences, which were prepared in CLIP data processing, with GSNAP with options of 5% mismatch rate and no terminal clipping. The alignment results were filtered to leave only the single best hits with minimum edit distance (up to one edit). If multiple best hits occur at the same edit distance, all alignments for the read were discarded as repetitive sequence. The position of ribosome footprint read relative to start or stop codons (Figure S5A) was determined based on CDS region in RefSeq annotation.

### Transcript-Level Quantification and Normalization for All High-Throughput Sequencing Libraries

CLIP tags were first aligned to mm9 genome. Using the nonredundant set of RefSeq to define the boundary of transcript, the number of reads corresponding to a given region annotated in RefSeq were counted. When a single read aligned across the boundary

between different regions (e.g., 5' UTR versus CDS), the reads were divided proportional to the aligned length in the given region (Figure 3A).

To calculate the CLIP tag density at the transcript level, the number of total CLIP reads from a given gene (defined by the nonredundant set of RefSeq) was normalized by dividing it with the number of filtered reads of the library. The CLIP tag enrichment at the transcript level was estimated as  $\log_2$  ratio between the normalized tag counts from the CLIP library and the normalized tag counts from the control RNA-seq library. We estimated FDR of enrichment level of a transcript by using  $\chi^2$  test of ranks in multiple CLIP libraries and Benjamini-Hochberg FDR estimation as performed by (Darnell et al., 2011).

For quantification of mRNA abundance (from RNA-seq) and ribosome occupancy (from ribosome footprinting), the tags were aligned to nonredundant RefSeq sequences. We excluded transcripts with low ribosome occupancy (<80 raw reads in siLuc) to reduce noise. Ribosome density of a transcript was calculated as the  $\log_2$  ratio between the read counts of ribosome footprints and the RNA-seq tag counts that were mapped to coding sequences (–15 from start and stop codon positions for 5'-end of sequence reads). Then, the ratios were normalized by the total read counts and were used to calculate density changes, defined as arithmetic differences in  $\log_2$  ribosome densities between control and *Lin28a* knockdown samples.

### Correlation Analysis of CLIP Enrichment Against Ribosome Footprinting or RNA-seq

Correlation analyses between LIN28A CLIP enrichment and mRNA abundance change (Figure 4A) or ribosome density change (Figure 4D) were done with  $\log_2$  enrichment level of the 35L33G CLIP library and the control RNA-seq library or  $\log_2$  fold change of ribosome density from ribosome footprinting experiment, respectively. Ribosome densities were processed as described in the previous section. We excluded transcripts with low read counts in RNA-seq (<30 raw reads) or low ribosome footprints (<80 raw footprint tags in siLuc library). Pearson's correlation coefficient was calculated by using *stats* package in SciPy (<http://www.scipy.org>).

As an alternative approach, we examined cumulative distributions of ribosome density according to LIN28A binding (Figures 4E and S5C). Transcripts were divided into four groups (top 5%, 5%–20%, 20%–50%, and 50%–100%) based on LIN28A CLIP enrichment. Non-coding RNAs and histone mRNAs (gene symbols starting with “Hist”) were excluded from the analysis. Differences between the test group and the weakest binder group (50%–100%) were tested by using Kolmogorov-Smirnov test in the R *stats* package. Note that the p values were smaller than the smallest supported nonzero number in IEEE 754 double precision (approx.  $4.94 \times 10^{-324}$ ).

### Analysis of Let-7 Targets

A possibility of indirect effect via *let-7* regulation was tested by assessing the effects on *let-7* targets in the mRNA level and ribosome density (Figure S5E). The predicted targets of *let-7* were retrieved from TargetScan mouse release 6.1 (<http://www.targetscan.org/>) and the validated targets were retrieved from the validated targets module of the miRWALK database (Dweep et al., 2011). Non-redundant RefSeq transcripts with greater than 15 RPKM in RNA-seq were divided into nontargets, targets with 7-mer site(s), and targets with 8-mer site(s). Targets that contain both 7-mer and 8-mer sites were included in the 8-mer group. The “validated targets” group was defined with allowed overlap to predicted 7-mer and 8-mer sites, however the transcripts in the group were excluded from the “nontargets” group. K-S test was performed to examine the significance in the difference between any two groups. None of the predicted targets showed significant response at the time of analysis (48 hr posttransfection).

### Gene Ontology Analysis

Gene ontology (GO) associations between GO term and UniProt entries were retrieved from UniProt-GOA database (Dimmer et al., 2012) on December 16<sup>th</sup>, 2011. UniProt accession numbers were mapped to RefSeq mRNA accession with UniProt ID mapping table of the latest version on the same day. Discrimination of the ribosome density changes among a specific GO term membership compared to the remaining genes was examined by using Mann-Whitney U test by comparing  $\log_2$  fold change of ribosome density. The p values were adjusted to false discovery rates (FDRs) by Benjamini-Hochberg method (Benjamini and Hochberg, 1995). For easier visualization, the list of GO terms was reduced to show only those with FDR < 0.05 for associated transcripts and the subset terms were omitted (Figure 5A).

Categorial terms used in analyses were defined as follows: nucleus (GO:0005634 – GO:0031966), integral membrane (GO:0016021 – GO:0031966), cytoplasm (GO:0005737 – GO:0031966), ER associated (GO:0016021 U GO:0005576 U GO:0009986 U GO:0005794 U GO:0005783 – GO:0031966), non-ER associated (all proteins – ER associated – GO:0031966) where U and – are set operators for union and subtraction, respectively. For transcript-level analysis (Figures 5B and S6A), transcripts with more than 15 RPKM in the siLuc-treated RNA-seq sample were separated based on protein product's gene ontology association as described above. For binding site-level analysis (Figure 5E), positions in RefSeq transcripts having alignment depth of at least 150 reads in CLIP library were collected. Then, binding sites with low CRES were eliminated if there exist one with higher CRES within 70 nt.

### Rough Endoplasmic Reticulum Microsome Purification

RER microsome were isolated from A3-1 mESCs with Endoplasmic Reticulum Isolation Kit (Sigma). A3-1 cultured in 100 mm dish were homogenized by passing the cells through a 26 gauge needle. Subsequent purification steps were performed according to

the manufacturer's protocol. The precipitates containing RER microsome were solubilized with the buffer containing 1X PBS, 0.1% SDS, 0.5% deoxycholate, and 0.5% NP-40.

### **<sup>35</sup>S-Methionine Metabolic Labeling**

A3-1 mESCs transfected with siLuc or siLin28a for 48 hr were incubated with media lacking methionine for 40 min. <sup>35</sup>S-labeled methionine was added to the media and incubated for additional 20 min. Cells were harvested and RER microsome fraction was isolated as described above. The protein concentration of RER microsome fraction and the remaining fraction in the RER microsome isolation were determined by BCA assay in triplicates. Equal amount of protein were separated by 10% SDS-PAGE and exposed to an imaging plate (Fujifilm). Radioisotope signal were read with the BAS-2500 system (Fujifilm).

### **Analysis for Selective Regulation of ER-Associated mRNAs**

To examine whether proteins of different topology are affected differentially by *Lin28a* knockdown, we took the annotations from the UniProt database (Magrane and Consortium, 2011). First, the list of mouse single-pass (SL-9904) and multi-pass (SL-9909) transmembrane proteins was obtained from the latest version of UniProt as of April 19<sup>th</sup>, 2012. Since many entries in UniProt lack detailed topology of mouse transmembrane proteins, they were reclassified based on regional annotations. Presence of N-terminal signal peptide in a single-pass transmembrane protein was determined from a "SIGNAL" feature starting from the first ten amino acids. Single-pass transmembrane proteins without N-terminal signal peptide were further subdivided into "tail-anchored" and "others" based on the relative position of "TRANSMEM" feature. Tailed-anchored transmembrane proteins were classified as those with transmembrane regions located adjacent to C-terminal (N-terminal-most position < 50 amino acids). These criteria were more generous than those of an ordinary prediction method such as in (Kalbfleisch et al., 2007), but the loosened threshold was a conservative take for our hypothesis.

### **Immunocytochemistry for A3-1 mESCs**

Cultured A3-1 cells were fixed in 4% formaldehyde and washed with deionized water. The fixed cells were smeared on a gelatin-coated slide and incubated for 30 min at 37°C. The fixed A3-1 on the slide was visualized with standard immunofluorescence technique and a fluorescence microscope (LSM700, Zeiss). The following primary antibodies were used in this study: LIN28A rabbit polyclonal antibody (Abcam, 1:500), KDEL mouse monoclonal antibody (EnzoLifeSciences; 1:100), GAPDH mouse monoclonal antibody (Santa Cruz; 1:100). The Alexa Fluor-conjugated secondary antibodies (Invitrogen) were used for visualization (1:1000). DAPI (Vector Laboratories) was used to visualize nuclei.

### **Immunocytochemistry for HeLa Cells**

To stain KDEL, HeLa cells were fixed in 4% formaldehyde for 10 min and permeabilized with 0.1% Triton X-100 for 10 min. In the case of GAPDH staining, HeLa cells were fixed and permeabilized in 100% methanol for 10 min at -20°C. The permeabilized HeLa cells were visualized with standard immunofluorescence technique and a fluorescence microscope (LSM700, Zeiss). The identical antibodies and DAPI used in A3-1 were also used for HeLa cells.

### **Comparison with the Data from ESC-EB Differentiation**

We retrieved the data from Ingolia et al. (2011), which reported the alterations in ribosome density during mESC to embryoid body (EB, 8 day) differentiation. Alignments to the UCSC known gene sequences are as provided in Ingolia et al. (2011). Ribosome density for each gene was quantified in the same way as the one used in our ribosome footprinting libraries. For comparisons, these genes were mapped to our set of nonredundant RefSeq transcripts by using *knownToRefSeq* table of the UCSC genome browser.

The most significant overlap of translationally regulated genes in two data sets (*Lin28a*-knockdown and mESC-EB differentiation) (Figure S7C) was searched by global scanning of minimum hypergeometric p value. Hypergeometric tests were performed for all possible pairs of rank cut-offs for two ranked lists of transcripts from the experiments. For faster computation, number of overlapping transcripts in specific cut-off pair was calculated with the following recurrence relation:

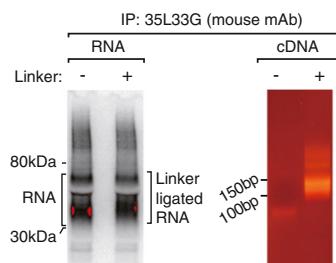
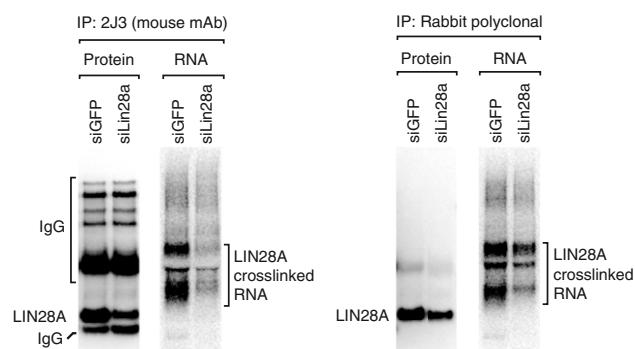
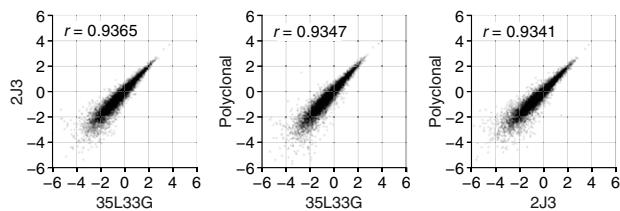
$$V_{i,j} = V_{i-1,j} + V_{i,j-1} - V_{i-1,j-1} + M_{i,j}$$

where  $V_{i,j}$  is the number of overlapping transcripts between top  $i$  translationally upregulated transcripts in *Lin28a*-knockdown and top  $j$  translationally upregulated transcripts in mESC-EB differentiation, and  $M_{i,j}$  is 1 if  $i$ -th transcript in *Lin28a*-knockdown and  $j$ -th transcript in mESC-EB differentiation were identical, and otherwise 0. Log-transformed p values from hypergeometric tests were calculated with *gsl\_cdf\_hypergeometric\_Q* function in GNU Scientific Library (GSL; <http://www.gnu.org/software/gsl/>).

### **SUPPLEMENTAL REFERENCES**

- Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Series B Stat. Methodol.* 57, 289–300.  
 Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* 460, 479–486.

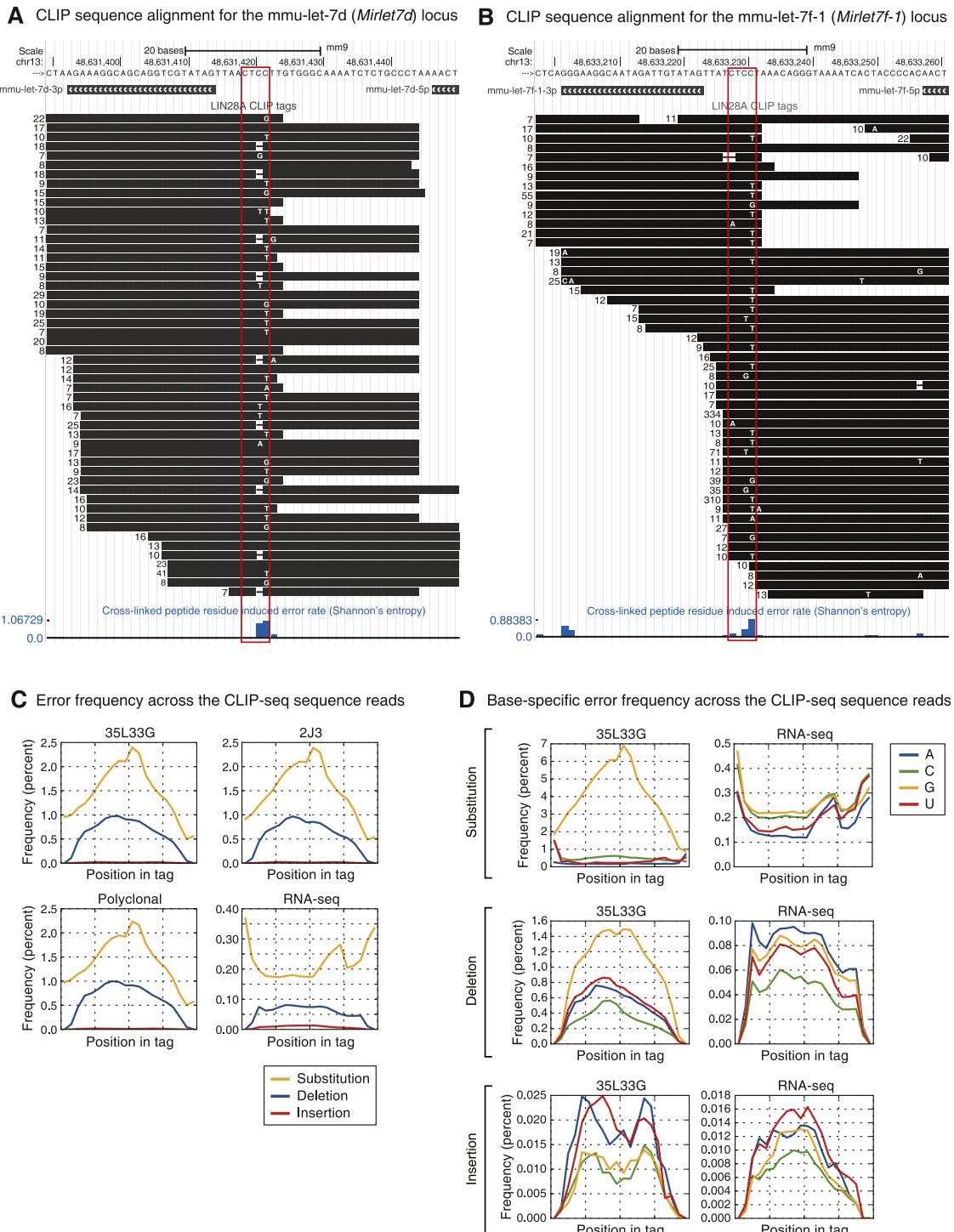
- Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190.
- Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W., et al. (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* 146, 247–261.
- Dimmer, E.C., Huntley, R.P., Alam-Faruque, Y., Sawford, T., O'Donovan, C., Martin, M.J., Bely, B., Browne, P., Mun Chan, W., Eberhardt, R., et al. (2012). The UniProt-GO Annotation database in 2011. *Nucleic Acids Res.* 40 (Database issue), D565–D570.
- Dweep, H., Sticht, C., Pandey, P., and Gretz, N. (2011). miRWALK—database: prediction of possible miRNA binding sites by “walking” the genes of three genomes. *J. Biomed. Inform.* 44, 839–847.
- Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466, 835–840.
- Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, S., Tacker, M., and Schuster, P. (1994). Fast Folding and Comparison of RNA Secondary Structures. *Monatsh. Chem.* 125, 167–188.
- Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802.
- Kalbfleisch, T., Cambon, A., and Wattenberg, B.W. (2007). A bioinformatics approach to identifying tail-anchored proteins in the human genome. *Traffic* 8, 1687–1694.
- Kozomara, A., and Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* 39 (Database issue), D152–D157.
- Magrane, M., and Consortium, U. (2011). UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* 2011, bar009.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. *Science* 302, 1212–1215.
- Wu, T.D., and Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26, 873–881.
- Zhang, C., and Darnell, R.B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat. Biotechnol.* 29, 607–614.

**A LIN28A CLIP-seq library preparation****B Specificity of immunoprecipitation****C Correlation of CLIP tag enrichment between biological replicates****Figure S1. Preparation of LIN28A CLIP libraries, Related to Figure 1**

(A) Preparation of LIN28A CLIP-seq library. Left: Autoradiography of the LIN28A-RNA conjugates that were end-labeled with  $^{32}\text{P}$  and ligated to the 3' linker. Because of the small size of the linker, mobility shift of the ligated product was not obvious on this gel. Right: RT-PCR product of the linker-ligated RNAs. Reverse transcription was performed after RNA purification and 5' linker ligation (see [Extended Experimental Procedure](#)). Samples incubated without linkers (-) were used as controls.

(B) Specificity of immunoprecipitation was tested by measuring the amounts of immunoprecipitated protein by western blotting (left) and the coimmunoprecipitated RNA by autoradiography (right). Compared to the control (siGFP treated mESCs), significantly less RNA was pulled-down with anti-LIN28A antibodies in *Lin28a*-depleted cells (siLin28a-treated mESCs).

(C) Reproducible results from the three biological replicates obtained with different LIN28A antibodies. Each dot in the scatter plots represents a log<sub>2</sub> enrichment relative to the background abundance measured by poly(A)<sup>+</sup> RNA-seq for a single nonredundant RefSeq transcript.

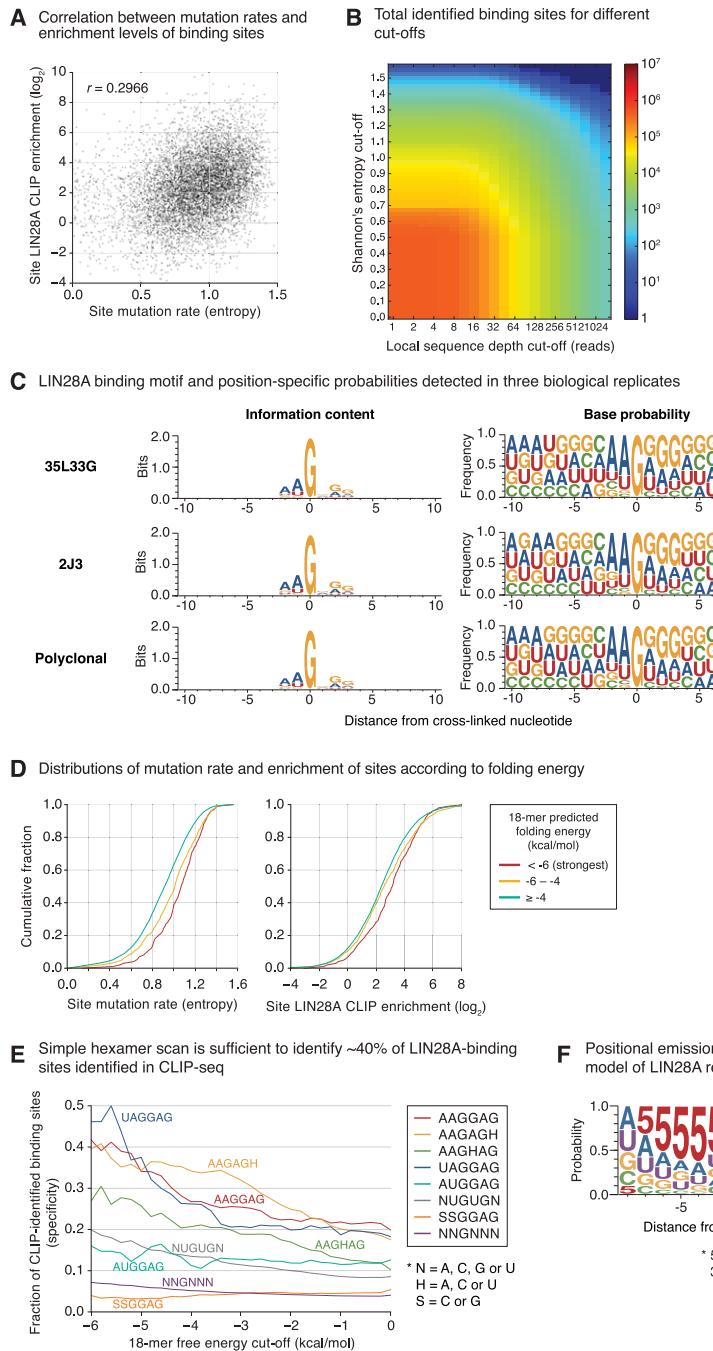


**Figure S2. Mutations at LIN28A Binding Sites, Related to Figure 1**

(A and B) LIN28A CLIP tags aligned to the *Mirlet7d* locus (A) and the *Mirlet7f-1* locus (B) presented in the same way as in Figure 1C. The genes are transcribed from minus strand.

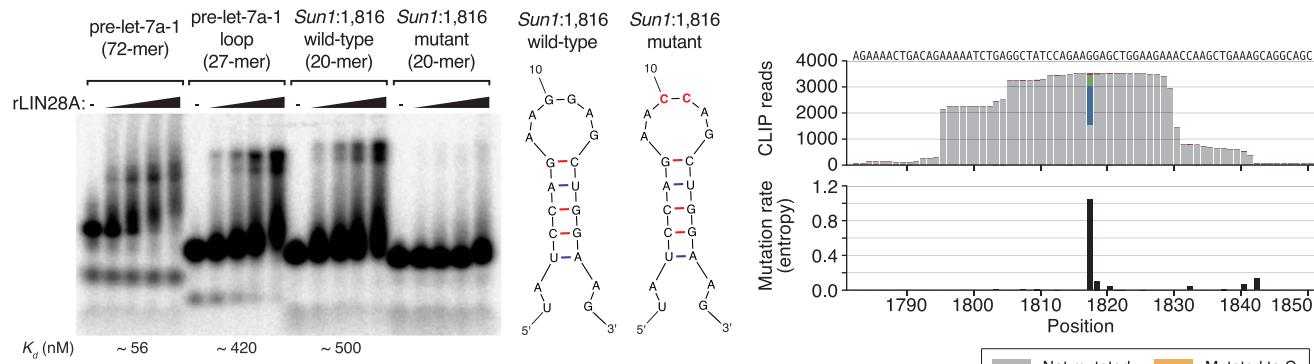
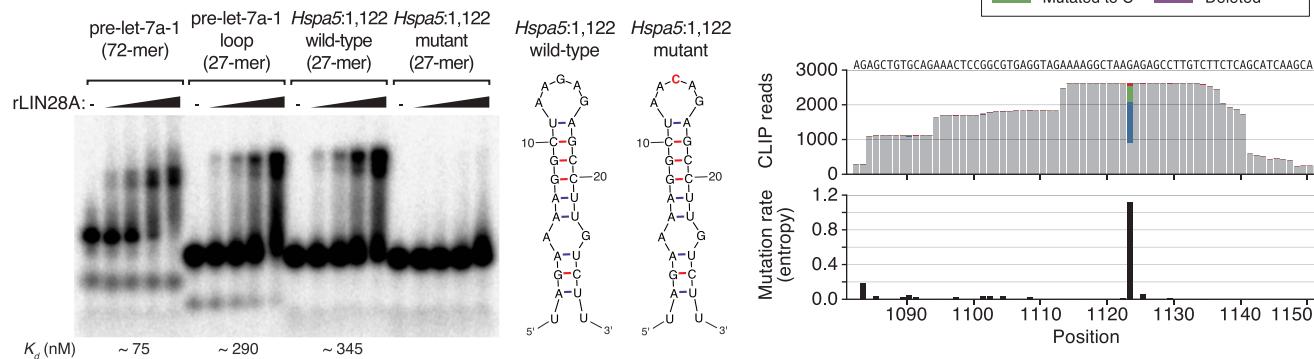
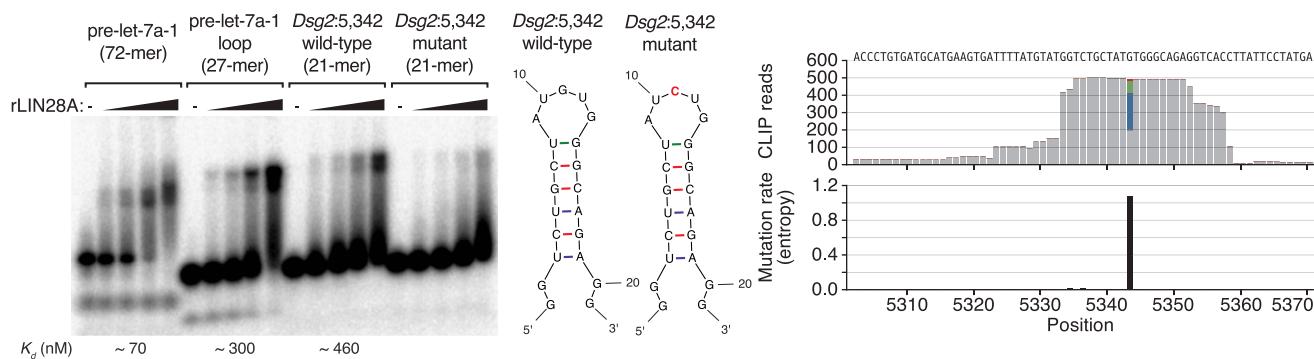
(C) Error frequency profiles as a function of position along the CLIP tags. Position within the tag was partitioned into 20 bins with the 5'-end of the reads as the leftmost bin (x axis). To avoid underestimation of errors at both ends, we replaced the sequences removed by terminal soft clippings with the original sequences obtained from sequencing. In the case of insertion errors, we assumed that the errors occur only at the left-side of a given base.

(D) Error frequency profiles of CLIP tags by error types, samples, and base types. Data from other CLIP libraries (2J3 and polyclonal antibody) show virtually identical patterns (not presented here due to space limitation).

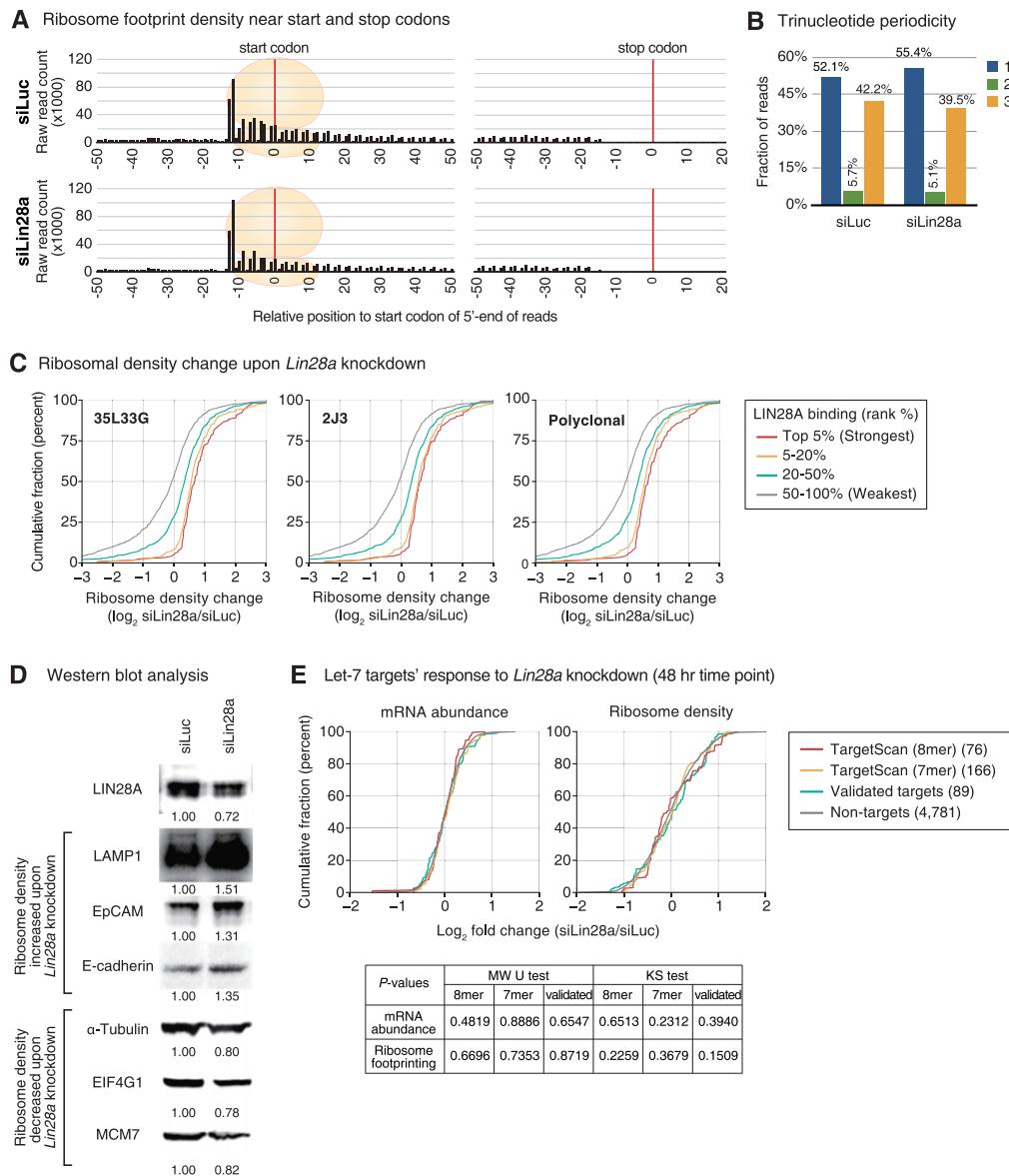


**Figure S3. Identification of LIN28A Binding Sites, Related to Figure 2**

- (A) Scatter plot for site mutation rate (Shannon's entropy, x axis) and site enrichment level in the CLIP-seq (y axis). Each dot represents an identified binding site (FDR < 0.001).
- (B) Total number of identified binding sites (indicated in rainbow color) depending on different thresholds of mutation rate (Shannon's entropy; y axis) and sequence alignment depth (reads; x axis).
- (C) Consensus of confident binding sites from individual CLIP-seq libraries (related to Figure 2A).
- (D) Distributions of site mutation rate (left) and CLIP enrichment (right) according to different intervals of predicted folding energy for 18-mer flanking sequences.
- (E) Simple search for hexameric motifs (for instance, AAGGAG) can predict ~40% of the experimentally identified LIN28A binding sites (FDR < 0.001) when combined with low folding energy (x axis). Folding energy prediction was done in 18-mer blocks with the hexamer located at the center. See [Extended Experimental Procedures](#) for the detailed analysis.
- (F) Emission probability matrix of the hidden Markov model trained for LIN28A-recognition motif. A, C, G and U events stand for respective unpaired nucleotide, whereas 5 and 3 are for any paired nucleotides. See also [Table S7](#) for the detailed training parameters and results.

**A** Electrophoretic mobility shift assay (EMSA) for the AAGNNG group**B** EMSA for the AAGNG group**C** EMSA for the UGUG group**Figure S4. Validation of Identified LIN28A-Recognition Motifs, Related to Figure 2**

Electrophoretic mobility shift assay (EMSA) was carried out with 5'-end  $^{32}$ P-labeled synthetic pre-let-7a-1, the loop region of pre-let-7a-1, and three snippets from LIN28A binding sites detected from LIN28A CLIP-seq and their mutants. Each selected target represents the AAGNNG, AAGNG or UGUG group (A, *Sun1*, NM\_024451:1,816; B, *Hspa5*, NM\_022310:1,122; C, *Dsg2*, NM\_007883:5,342). The phosphorimages from EMSA experiments are shown on the left side, predicted secondary structures of target RNAs and mutants are in the middle. Distribution of CLIP tags and mutations are displayed on the right side. The complex formation between the protein and RNAs were done with 0, 18.5, 55.6, 167 or 500 nM of recombinant LIN28A protein, respectively.



**Figure S5. Measuring Translational Efficiency after Depletion of *Lin28a*, Related to Figure 4**

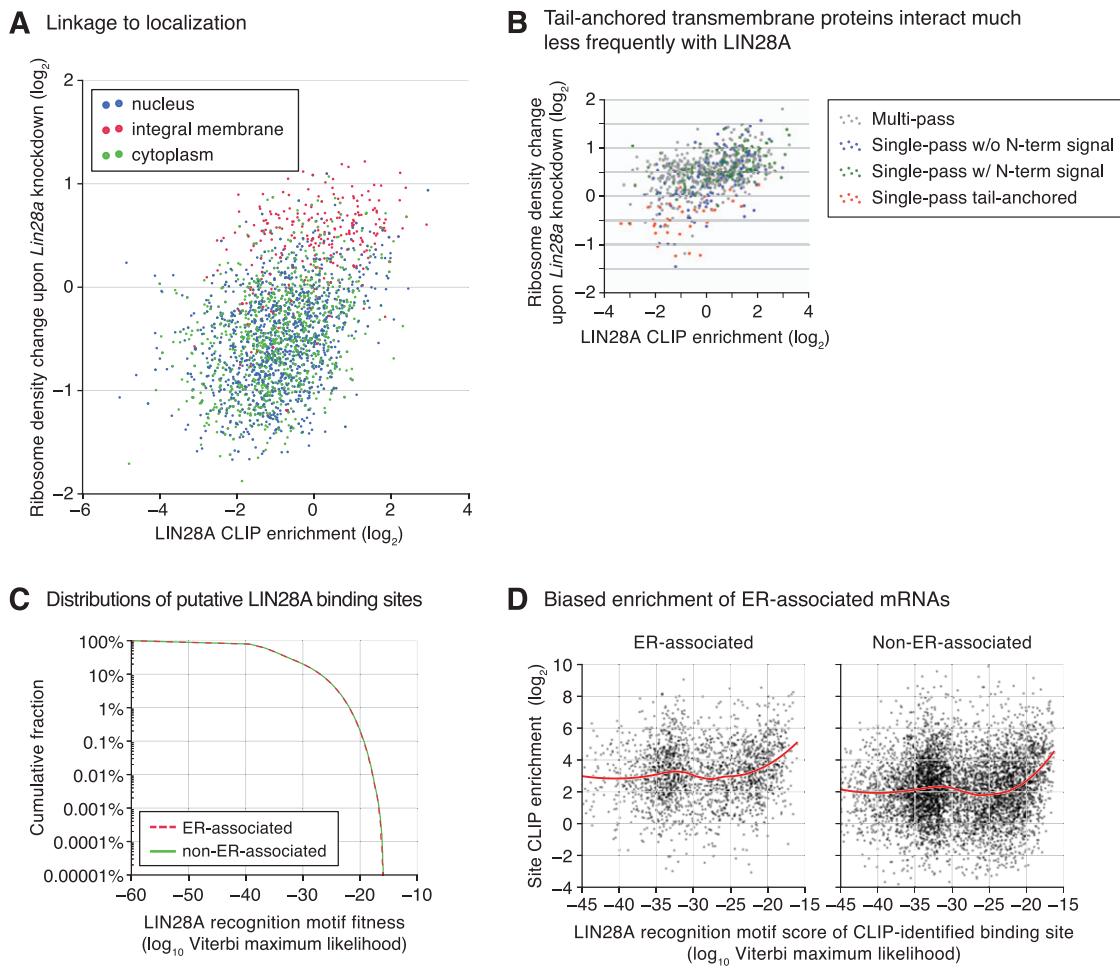
(A) Density of ribosome footprints near start and stop codons. All reads aligned to RefSeq transcripts (a nonredundant reduced set with only the longest isoforms) were merged. “0” represents the position of start (left) or stop (right) codons and indicated by red vertical lines. Because a ribosome spans 27–30 nucleotides on mRNA, ribosome footprints typically begin at 12–15 nt upstream of the start codon.

(B) Fraction of footprints mapping to each of the three nucleotides in codons in (A). Three nucleotide periodicity reflects the movement of ribosomes along mRNAs by three nucleotides at a time.

(C) Cumulative distributions of ribosome density change upon *Lin28a*-depletion. RefSeq transcripts were divided into four groups according to the degree of their enrichment in each CLIP-seq replicate. Strong binders tend to be de-repressed more significantly in *Lin28a*-depleted cells compared to weak binders. Note that the upper left (35L33G) is the same as Figure 4E.

(D) Western blot analysis of A3-1 mESCs upon *Lin28a* depletion. The same amount of total proteins was loaded according to the protein concentration determined by BCA assay. Six proteins were examined: LAMP1, EpCAM, E-cadherin, α-tubulin, EIF4G1, and MCM7. Based on RNA-seq, none of the six tested genes showed significant change at the mRNA levels upon *Lin28a* knockdown. LAMP1, EpCAM, and E-cadherin interact with LIN28A (based on CLIP-seq) and are translationally repressed by LIN28A (based on ribosome profiling), whereas the other three genes are not predicted to be LIN28A targets. The densitometric values normalized by control (siLuc) were indicated below respective lanes.

(E) Cumulative distributions of the changes in mRNA abundance (left) and those in ribosome density (right) in response to *Lin28a*-depletion. Let-7 target mRNAs were grouped into those with at least one 8-mer target site (red line), those with at least one 7-mer site (either 7-mer-A1 or 7-mer-m8 site) (yellow line), or nontarget groups (gray line) based on predictions in TargetScan mouse 6.1. The validated targets (greenish blue line) were retrieved from the miRWalk database (Dwee et al., 2011), and they were also used to define the “nontarget” group here. Both MW U test and KS test indicate that the let-7 targets are not influenced by LIN28A depletion at this early time point (48 hr posttransfection).



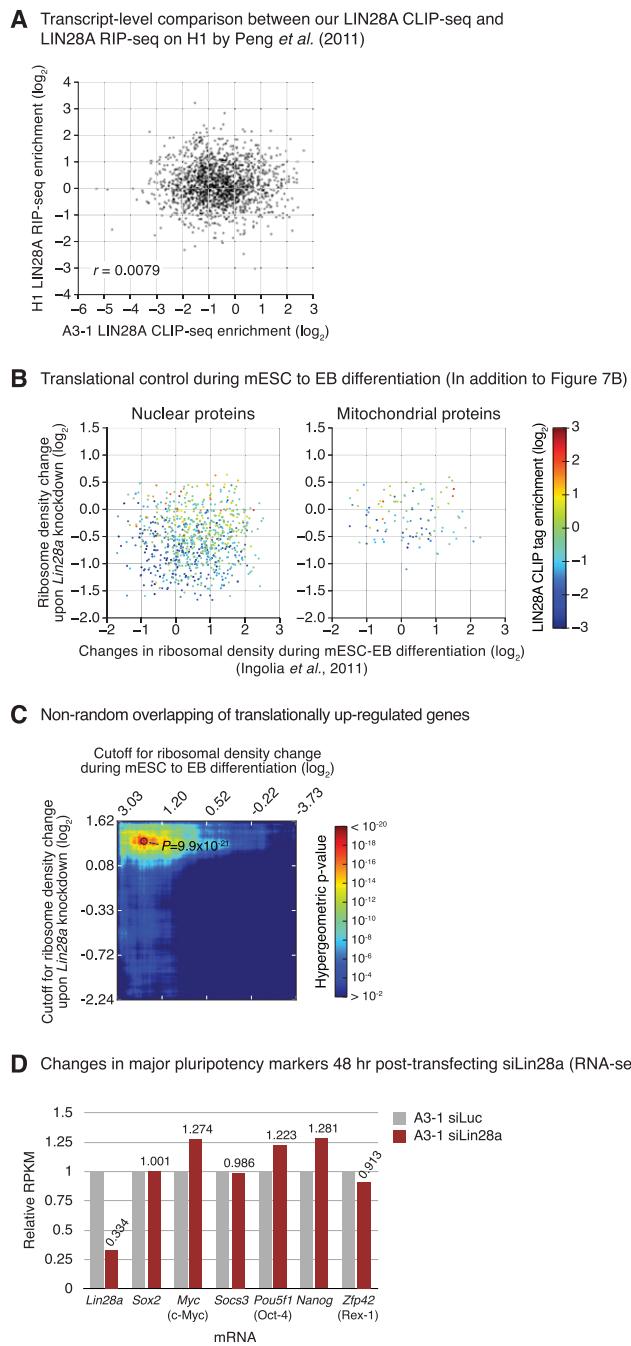
**Figure S6. LIN28A Targets ER-Associated mRNAs, Related to Figure 5**

(A) Complete version of Figure 5B.

(B) Degree of CLIP enrichment (x axis, reflecting LIN28A interaction) and changes in ribosome density (y axis) upon *Lin28a* knockdown. Compared to other transmembrane proteins, single-pass tail-anchored proteins are bound to LIN28A less frequently and their translation is affected less by LIN28A.

(C) Cumulative distribution of Viterbi maximum likelihood of hidden Markov model of LIN28A-recognition motif. Every possible 18-mer windows in RefSeq transcripts (a nonredundant reduced set) with > 15 RPKM expression were examined. Sequences that resemble the model (shown in Figure S3G) received higher likelihood. The likelihood distribution of ER-associated mRNAs are indistinguishable from that of non-ER mRNAs, suggesting that ER-associated mRNAs do not carry more LIN28A-recognition elements than other mRNAs do.

(D) Complete version of Figure 5E. The major group in lower motif score ( $\sim 10^{-33}$ ) represent for G/A-rich sequences that do not form stem-loop structures in 18-mer window. However, most of them lies on single-stranded regions in the middle of relatively strong secondary structures when predicted with bigger window.



**Figure S7. LIN28A may account for the global translational suppression of membrane proteins in pluripotent stem cells, Related to Figure 7**

(A) Transcript-level enrichments from our CLIP-seq on mESCs (x axis) and the RNA IP-seq on H1 by Peng *et al.* (2011) (y axis). There is no significant correlation between two data sets.

(B) Additional subcellular component terms to Figure 7B.

(C) The upregulated transcripts in *Lin28a* knockdown overlaps significantly with those that are increased during ESC-EB differentiation. Color in the matrix indicates p value from a hypergeometric test for significance of overlap between the two groups of genes with specific cut-offs. The most significant overlap was observed for the transcripts with increased ribosomes by more than 1.35-fold (0.43 in log<sub>2</sub> scale) by *Lin28a* knockdown and 2.91-fold (1.54 in log<sub>2</sub> scale) during ESC to EB differentiation ( $p = 9.9 \times 10^{-21}$ , Hypergeometric test). Importantly, we could not find such overlap for downregulated targets, indicating that LIN28A is involved in translational suppression rather than translational activation.

(D) The mRNA levels of pluripotency markers do not change significantly until this early time point (48 hr posttransfection). mRNA levels were quantified by RNA-seq.