HHS Public Access

Author manuscript

Mol Cell. Author manuscript; available in PMC 2019 July 26.

Published in final edited form as:

Mol Cell. 2018 March 15; 69(6): 1005–1016.e7. doi:10.1016/j.molcel.2018.02.012.

Systematic Discovery of RNA Binding Proteins that Regulate MicroRNA Levels

Julia K. Nussbacher[1,2,3] **Gene W. Yeo**[1,2,3,4,5,6,*]

[1]Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA 92037, USA

[2]Stem Cell Program, University of California, San Diego, La Jolla, CA 92037, USA

[3]Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA 92037, USA

[4]Molecular Engineering Laboratory, A★STAR, Singapore, Singapore

[5]Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[6]Lead Contact

SUMMARY

RNA binding proteins (RBPs) interact with primary, precursor, and mature microRNAs (miRs) to influence mature miR levels, which in turn affect critical aspects of human development and disease. To understand how RBPs contribute to miR biogenesis, we analyzed human enhanced UV crosslinking followed by immunoprecipitation (eCLIP) datasets for 126 RBPs to discover miR-encoding genomic loci that are statistically enriched for RBP binding. We find that 92% of RBPs interact directly with at least one miR locus, and that some interactions are cell line specific despite expression of the miR locus in both cell lines evaluated. We validated that ILF3 and BUD13 directly interact with and stabilize miR- 144 and that BUD13 suppresses mir-210 processing to the mature species. We also observed that DDX3X regulates primary miR-20a, while LARP4 stabilizes precursor mir-210. Our approach to identifying regulators of miR loci can be applied to any user-defined RNA annotation, thereby guiding the discovery of uncharacterized regulators of RNA processing.

*Correspondence: geneyeo@ucsd.edu.

AUTHOR CONTRIBUTIONS

Conceptualization, G.W.Y.; Methodology, J.K.N.; Formal Analysis, J.K.N.; Investigation, J.K.N.; Writing - Original Draft, J.K.N.; Writing - Review & Editing, J.K.N. and G.W.Y; Supervision, G.W.Y., Funding Acquisition, G.W.Y.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and one table and can be found with this article online at https://doi.org/10.1016Zj.molcel.2018.02.012.

DECLARATION OF INTERESTS

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

DATA AND SOFTWARE AVAILABILITY

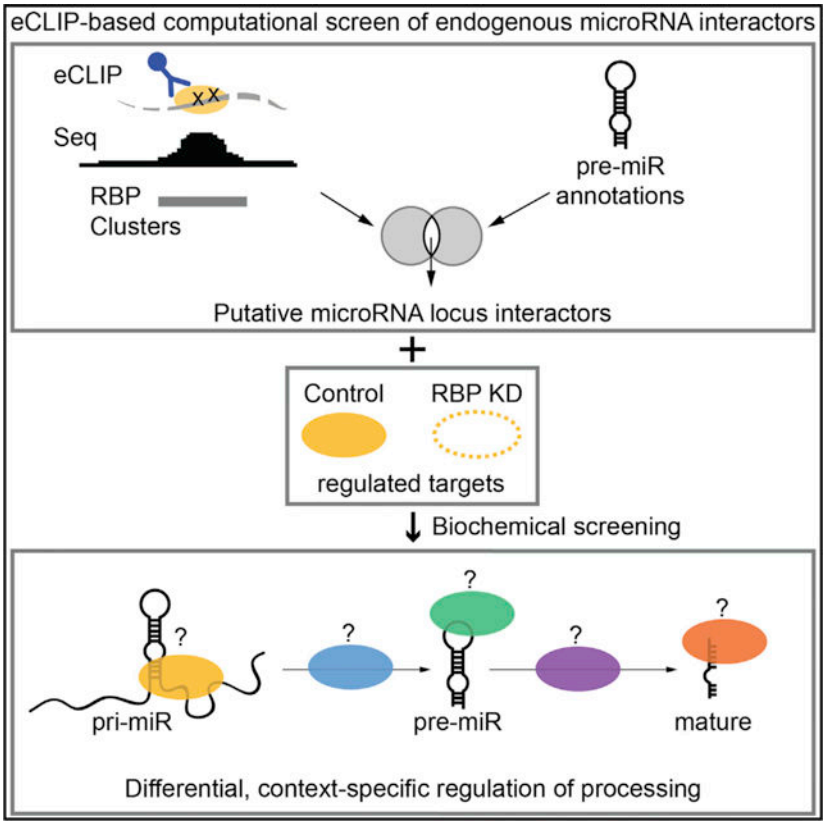
All downloaded eCLIP data and RNA-Seq used in this study is publically available from https://www.encodeproject.org, including BED files with the calculated fold change and p-adjusted values.

The accession number for the small RNA-seq data reported in this paper is GEO: GSE102497.

In Brief

Nussbacher and Yeo perform a computational screen of eCLIP data to identify putative, endogenous regulators of microRNA biogenesis and show that this regulation can be cell type specific and both promote or inhibit processing. Their approach can be applied to any transcriptomic annotation (RNA modifications, etc.) to identify novel RBP:RNA interactions.

Graphical Abstract



INTRODUCTION

MicroRNAs (miRNAs or miRs) are ~21- to 24-nt RNA molecules that control the expression of a myriad of genes, with over 1,800 annotated miR loci (Kozomara and Griffiths-Jones, 2011) predicted to influence >18,000 mRNA targets (Helwak et al., 2013). The biogenesis of miRs is a complex process primarily mediated by RNA binding proteins (RBPs). Precursor miR (pre- miR) hairpins, which are embedded within exonic and intronic regions of host protein-coding and non-coding primary miR (pri-miR) transcripts, are excised by RBPs DROSHA and DGCR8 (Han et al., 2004; Gregory et al., 2004; Denli et al., 2004). The pre-miR hairpin is then exported from the nucleus and the loop region of the hairpin is bound and cleaved by the RBP DICER to produce the mature 3p/5p miR duplex. Either the 3p or 5p miR interacts with additional RBPs, including Argonautes (AGOs), to form the RNA-induced silencing complex (RISC). This complex is stably bound to target

mRNAs via imperfect sequence complementarity to the loaded miR and regulates the stability and translation of target transcripts (Bartel, 2009).

Similar to protein-coding genes, non-coding pri-miR expression is subject to post-transcriptional control. Yet, only a small fraction (1%−5%) of RBPs have been demonstrated to control the various stages of miR processing in an endogenous context, from pre-miR formation to nuclear export and mature miR stability (Hao et al., 2017; Loffreda et al., 2015). While the RBPs responsible for the general processing of miRs, namely, DROSHA and DGCR8, are very well characterized, other RBPs can play a more targeted role in miR biogenesis to affect a specific subset of miRs. An example of regulation of miR biogenesis, DKC1 promotes the excision of small nucleolar RNA (snoRNA)- derived miRs from snoRNA host genes (Scott et al., 2009). A regulator of miR turnover, LIN28B controls the expression of let-7 miRs by interacting with the pri- and pre-miR to promote pre-miR degradation through recruitment of the uridyltrans-ferases TUT4 and TUT7 (Heo et al., 2009). KSRP and hnRNPA1 have also been extensively characterized as processors of the let-7 family member let-7a (Michlewski and Cáceres, 2010). To gain a full understanding of miR processing and to identify potential therapeutic targets, there is an urgent need to identify and classify additional regulators of miR biogenesis.

Thus far, systematic and comprehensive approaches to identify RBP:pre-miR interactions have been rare and limited in their utility. To address this lack of understanding, we leveraged two previously published resources: 365 immunoprecipitation-grade RBP antibodies (Sundararaman et al., 2016) and enhanced UV crosslinking followed by immunoprecipitation (eCLIP) datasets for 126 RBPs (Van Nostrand et al., 2016) made available through ENCODE. RBPs enriched for binding at annotated pre-miR loci would represent candidate RBPs that influence miR processing. Using these datasets, we identified 116 RBPs with enriched binding within pre-miR annotations and near their junctions, suggesting that the complete repertoire of miR-regulating RBPs is much greater than was previously known. We also showed that these miR:RBP interactions may be context specific, where, for some miR loci expressed in both cell lines screened, an RBP interacted with the locus in one cell line but not the other, indicating that other cell-type-specific factors may be facilitating or inhibiting RBP binding to miRs. This hypothesis is further supported by our observation that, while BUD13 specifically interacted with a target miR locus (mir-144), ILF3 interacted non-specifically with hairpin-forming RNA *in vitro.*

We selected a subset (10) of these 116 RBPs and subjected them to knockdown followed by small RNA sequencing to determine differential regulation of miR levels. We then integrated these functional analyses with eCLIP data to identify bound and regulated miR loci. While the number of miR loci that were both bound and regulated was fewer than individually identified by eCLIP or small RNA sequencing (RNA-seq), 92% of our RBP candidates both bound and regulated at least one miR locus. We validated direct interaction of RBPs and targets by electrophoretic mobility shift assay and found that some interactions are sequence specific while others seem to be dependent on other factors such as RNA secondary structure. Finally, we explored the stage of miR biogenesis controlled by our RBP candidates by RNA-seq, qRT-PCR, and northern blot analyses.

RESULTS

Integration of eCLIP Data and Pre-MicroRNA Annotations Identifies Putative MicroRNA Processing Factors

To identify RBPs with a previously uncharacterized role in micro-RNA (miR) biogenesis, we implemented a statistical approach leveraging the recent availability of high-quality protein-RNA interaction datasets (Figure 1A). We used publicly available eCLIP datasets accessible through encode.org that were generated for 126 RBPs in HepG2 and K562 cell lines (Van Nostrand et al., 2016). These datasets were uniformly generated by adhering to a consistent set of experimental parameters and utilizing immunoprecipitation-grade antibodies that satisfied a rigorous set of criteria (Sundararaman et al., 2016). Together, these attributes minimize the technical variation inherent when analyzing CLIP datasets from different sources. Also, a size-matched input control is associated with each pair of immuno-precipitation (IP) experiments for each RBP, enabling the identification of enriched regions of interest. We next identified RBPs enriched at miR loci. To do this, we combined precursor stem- loop miR (pre-miR) annotations from miRBase (hg19 v.20), (Kozomara and Griffiths-Jones, 2011) with eCLIP-defined clusters identified using a published cluster-finding algorithm (Lovci et al., 2013). Conservatively, we only retained clusters that were identified in both eCLIP replicates (overlapped by at least 1 bp). We reasoned that analyzing only the mature miRs would potentially exclude identifying RBPs that bind to the hairpin loop or flanking region prior to DICER cleavage.

In all, we examined 126 RBPs for binding at 1,871 annotated human pre-miRs. For each of the 181 eCLIP experiments (1 RBP in 1 cell line), we filtered for clusters that overlapped an annotated pre-miR by at least 1 bp, and we termed these clusters “miR locus clusters.” We found that nearly all RBPs screened had at least one cluster at a miR locus. To confirm that eCLIP does not have bias for the most highly expressed miR loci, we plotted eCLIP $\log_2$(fold enrichment) versus mature guide strand miR expression (reads per million [RPM]) for global miR regulators DGCR8 and DROSHA and observed no correlation (Figure S1A). To identify RBPs with significant enrichment within miR loci, we generated volcano plots of the RBP cluster enrichment as fold enrichment (IP over size-matched Input or SMInput) against the adjusted p value (padj) (Figure 1B). To further resolve putative pre-miR and pri-miR binders, we categorized each cluster based on whether the cluster was exclusively within the boundaries of the miRBase annotated pre-miR or spanned the pre-miR/pri-miR junction (Figure 1A). Next, we determined how many RBPs had an eCLIP cluster within at least n = 1, 5, or 10 miR loci as we incrementally increased the stringency of the fold enrichment and padj cutoffs (Figures S1B and S1C). We found that for an IP/SMInput fold enrichment cut-off of 4 ($\log_2 = 2$) and a padj cutoff of 0.001 ($-\log_{10} = 3$), 92% of RBPs interacted with at least one unique miR locus, 51% of RBPs interacted with at least five unique miR loci, and 25% of RBPs interacted with at least 10 unique miR loci (Figure S1C). Using these cutoffs, we compared our DGCR8 eCLIP data to previously published DGCR8 high-throughput sequencing of RNA isolated by UV crosslinking followed by immunoprecipita-tion (HITS-CLIP) data (Macias et al., 2012). We found that the distribution of eCLIP clusters across coding regions and long non-coding RNAs (lncRNAs) were similar to Macias et al. both before and after applying our fold enrichment and padj

cutoffs, with a majority of clusters occurring in introns (Figure S1D, top row pie charts). However, when peaks within non-coding RNAs were examined, we observed that >90% of clusters were assigned to miRs, compared to 30% in the HITS-CLIP data (Figure S1D). This emphasizes the importance of stringent removal of background signal by SMInput. Targets of DGCR8 were highly consistent between eCLIP cell-types compared to HITS-CLIP (Figure S1E). We also compared DROSHA eCLIP clusters to previously published formaldehyde crosslinking and immuno- precipitation (fCLIP) peaks (Kim et al., 2017) and found general agreement in the miR targets identified (Figure S1F), despite different crosslinking techniques and cellular contexts.

Satisfied with the fold enrichment and padj cutoffs, we determined the number of RBP-bound miR loci occurring in both HepG2 and K562. We found that several of the RBP:pre-miR interactions appeared cell type specific (Figure S1G), in agreement with a previous study (Treiber et al., 2017). To determine whether the cell-type-specific RBP:miR interactions were due to differential expression of the miR locus, we determined the expression of the uniquely bound loci in the cell line for which no interaction (G and H) Volcano plots of eCLIP clusters in HepG2 and K562 cell linesfor known regulators ofaspecific subset of miRs: (G) LIN28B (H) DKC1. Cutoffs ofeCLIP IP/SMInput $\log_2$(fold enrichment) = 2 and $-\log_{10}$(padj) = 3 are denoted by a dashed red line. Known binding targets of LIN28B (D) and DKC1 (E) are highlighted. eCLIP data in K562 were not available for DKC1. eCLIP IP/SMInputfold enrichment and padj were calculated as in Van Nostrand et al. (2016). padj = adjusted pvalue. Volcano plot clusters are colored based on the (non)-genic origin of the intersecting pre-miR based on ENSEMBL annotations and mirtron annotations from Ladewig et al. (2012). Volcano plot filled circles represent pre-miR exclusive clusters while open circles represent pre/pri-miR junction spanning clusters. was detected. We found that, of the loci bound only in HepG2 or only K562, 50% were expressed in the other cell line but not bound. We conclude that for certain RBPs, there are some cell-type-specific RBP:miR interactions in which the miR locus is expressed but differentially bound, possibly due to differentially expressed cofactors or competitors.

To evaluate whether our systematic approach recovers RBPs that are well characterized as regulators of miR biogenesis, we examined the results for the global miR processing RBPs DROSHA and DGCR8, RBPs that are known to interact with pre-miRs and affect cleavage from the pri-miR transcript. Using our stringent filtering criteria, we scanned for DROSHA and DGCR8 eCLIP clusters at miR loci for both RBPs and observed strong statistical enrichment of RBP binding at 160 and 184 (DGCR8) and 86 and 79 (DROSHA) miR loci in HepG2 and K562 cell lines, respectively (Figures 1C and 1D). The pre-miRs identified arose from a diversity of genic regions. We also found that of the eCLIP clusters within miR loci, a majority (87%−96%) of DGCR8 clusters were exclusively found in annotated pre- miRs while a majority (70%−80%) of DROSHA clusters were found at pre/pri-miR junctions, as illustrated by the mir-21 locus (Figures 1E and 1F). Our analysis clearly confirms pre/pri-miR junction binding preference by DROSHA and stem binding preference by DGCR8.

Based on the known biology of proteins such as LIN28B and DKC1 that are, unlike DROSHA and DGCR8, known to regulate a very specific subset of miRs, we expected the majority of RBPs would likely each affect only a small subset of miRs. Specifically,

LIN28B regulates 11 of the 12 let-7 family members (Triboulet et al., 2015) and DKC1 regulates at least 5 snoRNA- derived miRs (Scott et al., 2009), and as a control we examined our binding data for enrichment at these loci. Our LIN28B results showed binding to the expected targets, with clusters in let-7b/f/ d/i/g, and miR-98 (Figure 1G) as reported previously (Heo et al., 2008; Nam et al., 2011). DKC1 acts specifically on H/ACA snoRNA-derived miRs (Ge et al., 2010; Alawi and Lin, 2010). We observed statistical enrichment for DKC1 binding at snoRNA-derived miRs in HepG2 cells, detecting significant clusters in miR-664a, 664b, 1248, 1291, 3607, 3651, and 3653, all of which are snoRNA-derived miR loci (eCLIP data for DKC1 in K562 cells was not available) (Figure 1H). These findings confirmed that our fold enrichment and padj cutoffs are stringent enough to identify known, specific miR interactors such as LIN28B and DKC1 as well as general miR processors such as DROSHA and DGCR8. In conclusion, we validated parameters that allow successful identification of RBPs that bind known miR loci. Applying these parameters to the ENCODE eCLIP data revealed 116 RBPs that bind directly to miR loci.

LIN28B/DKC1 miR Targets Are Modulated as Expected upon Knockdown

Before we selected RBP candidates for validation of miR regulation, we further verified experimental conditions by which we can demonstrate miR regulation by LIN28B and DKC1 (Figure 2). As expected, LIN28B depletion in both cell lines significantly increased the expression of let-7 family members as measured by small RNA-seq (padj < 0.05) (Figures 2A and 2B), while loss of DKC1 resulted in a significant decrease in known snoRNA-derived miR-1291 by small RNA-seq (Figure 2C). We selected three examples of LIN28Band DKC1 targets to illustrate this validation (Figures 2D and 2F). Next, we compared both eCLIP and small RNA-seq data by simultaneously evaluating the binding enrichment of the RBP within the miR locus, and the miR expression level (Figures 2G and 2I), highlighting miRs that both changed significantly (small RNA-seq padj < 0.05) and harbored significant clusters (eCLIP) (padj < 0.05). As expected, LIN28B eCLIP clusters in the let-7 pre-miR loci were enriched in both cell lines (Figures 1G and 1H), and let-7s were significantly upre-gulated upon LIN28B depletion (Figures 2G and 2H). Likewise, we observed significant enrichment of DKC1 eCLIP clusters at miR-1291, which wasdownregulated upon DKC1 depletion (Figure 2I). Therefore, we feel confident that this approach is suitable to identify RBPs that bind and regulate specific miR loci.

Most RBPs Screened Bind Fewer Than 25 Unique miR Loci

In selecting RBP candidates for further validation of miR regulation, we aimed to capture a wide range of binding specificity; therefore, we determined the distribution of the number of unique miR loci bound by each RBP in our screen (Figure 3A). We found that the known global regulators of miR biogenesis, DGCR8 and DROSHA, were the RBPs with binding sites in the greatest number of miR loci (152 and 175,80 and 74 loci, respectively, in HepG2 and K562 cells). Our other positive controls LIN28B and DKC1 fell in the lower end of the distribution with 10 (LIN28B in HepG2), 19 (LIN28B in K562), and 11 (DKC1 in HepG2) miR loci with binding sites (Figure 3A). To validate RBPs not yet characterized as endogenous miR regulators, we selected 10 RBPs for further analysis (BUD13, DDX3X, ILF3, LARP4, LARP7, PRPF8, PTBP1, RBFOX2, SF3B4, and SLTM) for which the number of uniquely bound miR loci ranged from 3 loci (SLTM) to 58 loci (SF3B4). This

allowed us to study RBPs with both highly specific and broad putative (pre-)miR binding. While most eCLIP cluster-containing miR loci exist in a diversity of genic regions, some RBPs showed enriched for a particular region. For instance, splicing factors SF3B4 and PRPF8 were enriched at mirtrons (Figures 3B and S2).

Depletion of miR-Binding RBPs Affects miR Levels

To evaluate whether candidate RBPs enriched for binding pri-,pre-, or mature miRs also affects the steady-state levels of the mature species, we depleted each of our 10 candidate RBPs (BUD13, DDX3X, ILF3, LARP4, LARP7, PRPF8, PTBP1, RBFOX2, SF3B4, and SLTM) in both HepG2 and K562 cells as we had our controls (Figures S3A-S3C) and then performed small RNA-seq to evaluate changes in miR levels over non-targeting control. To demonstrate the reproducibility of our small RNA-seq data, we calculated the distribution of the correlation coefficients (R^2 values) of RPM data for quadruplicate control experiments and duplicate RBP knockdown experiments (Figures S3D-S3G). We quantified differentially expressed miRs upon RBP depletion and generated volcano plots representing the enrichment of the miR levels in knockdown conditions versus control $\log_2$(fold change) against the -$\log_{10}$(padj) (Figure 3C). Of the 10 RBPs selected from our screen for further analysis, all showed alteration of mature miR levels upon depletion of the RBP. Notably, the miR $\log_2$(fold change) revealed approximately equal numbers of mature miRs significantly increasing and decreasing in expression upon RBP depletion. The range of the $\log_2$(fold change) was also approximately equal for increasing and decreasing miRs, such that RBPs causing large increases in expression caused equally large decreases in expression of mature miRs upon depletion. Some notable exceptions include the upregulation of five or fewer miRs in PRPF8 (K562), PTBP1 (K562), ILF3 (HepG2/K562), and SF3B4 (HepG2), as well as downregulation of five or fewer miRs in LARP4 (HepG2), PRPF8 (HepG2), and DDX3X (K562). We confirmed the alteration of several miRs identified in our initial screen, and we highlighted four examples: BUD13 (mir-144 and mir-210), DDX3X (mir-20a), ILF3 (mir-144), and LARP4 (mir-210) using the University of California, Santa Cruz (UCSC) genome browser (Figure 3D).

Candidates Directly Interact with Bound and Regulated miR Loci

We generated an integrated analysis of RBP binding enrichment and miR expression for our selected subset of 10 RBPs (Figures 4A and S4). Of the 10 candidate RBPs tested, only SLTM failed to show any concordance between bound miR loci and miRs altered by RBP knockdown in either cell line. For the remaining 9 candidates, we identified several RBP:miR loci interactions that also displayed altered expression upon RBP depletion (upper-right quadrants), suggesting that these RBPs either suppress miR processing or enhance degradation. In contrast, we observed RBPs that bound miRs, but their expression decreased upon RBP knockdown (lower-right quadrants), indicating that these RBPs promote miR biogenesis or enhance stability. While some RBPs seem to have a unidirectional effect on bound miR loci (DDX3X, PRPF8, and SF3B4), others appear to have differential effects (BUD13, ILF3, and RBFOX2). Further-more, while many loci were bound in both HepG2 and K562, fewer were both bound and also regulated in both cell lines.

Of the 10 candidate RBPs, 9 affected at least 1 mature miR concordant with significant enrichment of RBP binding (Figures 4A and S4). Interestingly, the RBP-interacting miR loci that were also altered by RBP depletion were not uniformly up- or downregulated upon RBP depletion. For example, BUD13 had significant enrichment at both the miR-210 and miR-144 loci; however, BUD13 depletion resulted in increased expression of miR-210 and decreased expression of miR-144. This differential effect was also observed for ILF3, LARP4, and RBFOX2. Other RBPs, specifically PRPF8 and SF3B4, were enriched at multiple miR loci, and the depletion of these RBPs resulted in uniformly suppressed expression of the mature miRs at these loci. Because these RBPs are known splicing regulators, and the affected loci originate from mirtrons, depletion of these RBPs likely resulted in reduced splicing of these mirtrons causing diminished expression of the mature miRs. We conclude that in addition to having effects upon specific, bound miR loci, depletion of our RBP candidates have concomitant but indirect effects on pathways that cause suppression and upregulation of non-target miR loci. We further conclude that RBPs may bind and regulate miR loci to cause both increases and decreases in miR expression in a locus-specific, and often cell-type-specific manner, suggesting multiple mechanisms of regulation.

To confirm that these regulated loci could interact with the RBP independent of *in vivo* co-factor associations, we performed electrophoretic mobility shift assays (EMSAs). Using purified recombinant ILF3, we observed *in vitro* interactions with the segment of the miR-144 locus with eCLIP enrichment (Figure 4B). However, this interaction appeared non-specific to other miR targets *in vitro* (miR-20a locus has no eCLIP enrichment), consistent with ILF3 being a double-stranded RBP. This result suggests that ILF3 may have an *in vitro* sequence or structural dependence for miR locus interaction. This result also demonstrates that purely *in vitro* approaches to discovering RBP-target interactions are insufficient to reveal the *in vivo* specificity of endogenous interactions. In contrast, purified BUD13–2xFLAG (Figure 4C) interacted specifically with the miR-144 locus, but not to non-target (miR-20a locus), recapitulating target-specific binding *in vitro*, independent of *in vivo* co-factor associations (Figure 4D).

Candidate RBPs Control the Biogenesis and Stability of Target miRs in Multiple Ways

To evaluate which step of miR processing is affected by our candidate RBPs, we first identified changes in miR host-gene (pri-miR) expression upon RBP depletion (Figure 5A). Using encode.org RNA-seq data for shControl versus shRNA knockdown of an RBP (shRBP), we looked at differential expression of annotated host genes and found that ~50% of host genes had no significant change upon knockdown, suggesting regulation downstream of the pri-miR, while ~50% showed a slight but statistically significant ($p < 0.05$) change, indicating potential regulation at the point of pri-miR production or stability. For a subset of these bound and regulated miR loci (BUD13, DDX3X, ILF3, and LARP4 in K562 cells), we evaluated pre- and mature miR expression upon RBP depletion by northern blot analysis (Figure 5B). BUD13 and ILF3 were shown to regulate the miR- 144 locus, and, although there is no annotated host gene, we detected downregulation of both precursor and mature miR-144 (Figures 5B-5D). These data combined with the observed eCLIP peak locations (Figure 3D) suggest that BUD13 stabilizes the primary transcript while ILF3 may stabilize

the pre-miR. DDX3X was observed to bind and regulate the miR-20a locus, and a reduction was observed for pri/pre/mature species, suggesting a role for DDX3X in pri-miR production or stability. Finally, both BUD13 and LARP4 bound and regulated the miR-210 locus. Neither showed significant effects on the primary species (Figure 5A) and both showed a reduction in precursor (Figures 5B and 5C). Interestingly, BUD13 suppression caused upregulation of mature miR-210 (Figure 5D), implicating this RBP in suppression of miR-210 precursor processing. Conversely, LARP4 suppression caused downregulation of mature miR-210 (Figure 5D), suggesting that this RBP may stabilize the precursor. These experiments confirmed that modulation of RBP expression alters miR expression and provide insights into the stage of biogenesis affected.

DISCUSSION

While the RBP components of the global miR processing machinery (DROSHA, DGCR8, TUTases) and a select few regulators of specific miR subsets including LIN28B, DKC1, and hnRNPA1 have been well characterized, there has yet to be an exhaustive approach to identifying RBPs that endogenously bind and regulate miRs during all stages of biogenesis and turnover. A recent study implemented an *in vitro* mass-spectrometry approach but was limited to only 72 pre-miR baits of ~1,900 annotated pre-miRs (Treiber et al., 2017). Here, we expand upon these findings by exploring the putative role of 126 of ~1,800 annotated RBPs in an endogenous context. In this study, we leveraged well-curated ENCODE eCLIP datasets to identify RBP binding sites that are enriched either exclusively within pre-miR annotations or spanning pri/pre-miR junctions. We identified 116 RBPs that interacted with at least 1 unique miR locus, with 32 of these RBPs interacting with at least 10 unique miR loci, and 431 unique miR loci that interact with at least one of the 116 RBPs. In total, we identified 1,740 unique RBP-miR locus interactions. As expected, the miR processing enzymes DROSHA and DGCR8 interacted with the greatest number of miRs. Nevertheless, we also validated known regulators of a subset of miRs such as LIN28B (the let-7 family members) and DKC1 (snoRNA-derived miRs) to gain confidence in our approach.

To identify previously uncharacterized RBP-miR interactions, we selected 10 RBPs that represent a broad range of miR loci binding. Included in this list of candidates is RBFOX2, and while there was a previous study characterizing *in vitro* RBFOX2 regulation of miR-20b and miR-107 (Chen et al., 2016), we observed no eCLIP reads at these loci and no change in mature miR expression upon RBFOX2 knockdown, leading us to believe these may be non-endogenous interactions. We observed that many miR loci expressed in both HepG2 and K562 cells were differentially bound by RBPs. Such cell type specificity was also observed in Treiberet al. and suggests there are additional factors regulating binding specificity, such as cofactors or RNA secondary structure. We also observed that the double-stranded RNA (dsRNA) binding protein ILF3 is able to bind target hairpins *in vitro,* but that this was non-specific to another dsRNA, further suggesting that secondary structure alone is insufficient to confer specificity of target interaction *in vivo.* In the case of BUD13, we observed direct and specific binding of a predicted miR target. Together these data suggest that *in vitro* experiments are useful but not always able to identify endogenous RBP:miR interactions.

To better understand how these RBPs may be regulating biogenesis of their target miR loci, we performed several biochemical assays to quantify various stages of processing. Upon RBP knockdown, we observed significant alteration of mature miR expression. For example, upon suppression of BUD13 and ILF3, we observed a downregulation of target miR-144. BUD13is an RBP with putative roles in splicing (Dziem- bowski et al., 2004) that has been associated with metabolic syndrome (Lin et al., 2016), while ILF3 has a putative role in cardiovascular disease (Yoshida et al., 2011a, 2011b) and was previously shown to interact with DGCR8 (Shiohama et al., 2007). We found that these RBPs affect the miR-144 locus, mostly likely by promoting stability of the precursor (ILF3, which has a binding site within the pre-miR) or the primary (BUD13, which has a pre/ pri-junction spanning cluster) miRNA transcript. miR-144 has been shown to act as both a tumor suppressor (Ren et al., 2017) and a potential suppressor of SCA1 through downregula-tion of ATXN1 (Persengiev et al., 2011), suggesting that BUD13 and/or ILF3 may have protective affects against cancer and neurodegeneration. DDX3X is an RNA helicase, a family of proteins involved in the remodeling of the transcriptome that has been associated with tumorigenesis and cancer (Valentin-Vega et al., 2016). It was previously shown that DDX3X interacts with DROSHA/ DGCR8 (Zhao et al., 2016), but whether DDX3X interacts with other miR processing RBPs remains unclear. Our results suggest that DDX3X directly interacts with miR loci, for example, miR-20a, to promote pri-miR expression, potentially through influencing transcription. Interestingly, the DDX3X peak at miR-20a drops off at the boundaries of the mature guide strand, indicating possible interaction with the guide:passenger dsRNA species after dicer cleavage. Our results also showed that LARP4 stabilizes the miR-210 locus, mostly likely the pre-miR, as pri-miR was unaffected, but both pre-miR and mature were downregulated upon loss of LARP4. LARP4 is a ribonucleo-protein known to regulate mRNA translation through binding to the poly(A) tract (Yang et al., 2011) and miR-210 is a critical component of the hypoxic response (Fasanaro et al., 2008). The sequestration of LARP4 into stress granules upon arsenite treatment (Yang et al., 2011) may prevent the stabilization of miR-210, thereby releasing the repression of PTBP1 and allowing for apoptosis in the presence of stress. We also observed BUD13 interaction with the miR-210 locus, which appears to suppress processing of the precursor to the mature based on the observation that BUD13 suppression increases precursor processing. Interestingly, a subset of our candidates cause both up- and downregulation of miR loci; for example, RBFOX2 stabilizes the miR-18a, miR-20a, miR-144, and miR-126 loci but destabilizes the miR-92a-1 locus. Similarly, BUD13 promotes mature miR-144 and represses mature miR-210, ILF3 stabilizes the miR-144 locus but destabilizes the miR-1304 locus, and LARP4 stabilizes the miR-210 locus but destabilizes the miR- 223 locus. This duality in function suggests that RBPs may have context or cofactor-dependent effects specific to a target.

In addition to identifying previously uncharacterized endogenous regulators of miR biogenesis, we have also demonstrated the power of a generalized approach for screening any eCLIP, or CLIP-like (e.g., PAR-CLIP, iCLIP, ChIP) dataset for binding to any set of transcriptomic intervals (e.g., annotations of alternatively spliced exons, stop codons, A-to-I editing sites, or post- transcriptional modifications such as m^1A, m^6A, or pseudouridine, or secondary structure). Not only is our approach scalable, but it can also be applied at virtually no cost when screening ENCODE or other published datasets against any standard or

custom annotation. Furthermore, the screening of eCLIP data has two critical advantages over other approaches. The first being that crosslinking allows for highly stringent washing to capture direct, endogenous RBP:RNA interactions while removing both indirect interactions as well as erroneous post-lysis interactions. The second is the IP step that results in significant enrichment of these interactions, which is critical if either the RBP or its target is lowly expressed. One disadvantage of our approach is that the scope is limited to the RBPs for which we had eCLIP data and is presently further limited to the cell types in which these assays were performed. As our data and previous data (Treiber et al., 2017) have shown, many RBP:miR interactions are cell type specific. Additionally, validation by RBP knockdown followed by sequencing and northern blot analysis is non-trivial at scale, particularly for cell types difficult to modify with RNAi. In summary, we are confident that our rapid approach that leverages rigorous, existing datasets allows for accurate identification of miR-interacting RBPs and has broad application to quickly identifying candidate interactors of myriad other RNA characteristics and modification.

Our study identified a surprising number of putative miR-inter-acting RBPs, laying the groundwork for both further validation of the 106 remaining candidates as well as the need to explore the potential for other RBPs to influence miR biogenesis and regulation. Our study complements an earlier screen that used synthetic pre-miR baits to pull down RBPs from lysates. This study allowed for detection of pre-miR-specific RBP interactors, as well as RBPs that bind pre-miRs through association of a protein complex. We conducted several comparative analyses (Figure S5) and found that few RBP:miR interactions were positively identified in both studies. This was in large part due to differences in scope of the two approaches but may also be due to technical differences. For example, crosslinking in eCLIP allows for more stringent washes to remove post-lysis and indirect interactions. Furthermore, while the approach in Treiber et al. has the specificity of RBP:pre-miR interactions, the bait strategy captured these interactions in a non-endogenous context. Interestingly, both approaches identified many cell-type-specific RBP:miR interactions, suggesting that further studies will need to consider cellular context when characterizing additional miR-interacting candidates.

From these observations, we conclude that an abundance of RBPs either directly or indirectly influence miR biogenesis, providing a valuable resource for further discovery of both miR-regulating RBPs as well as a computational platform for comprehensive analysis of RBP interactions with other regions or characteristics of interest. In order to fully explore the role of RBPs in miR biogenesis, additional studies must be extended to the full repertoire of both RBPs as well as relevant developmental and pathophysiological contexts. Furthermore, RNA binding domain mutation of these candidates followed by miR target analysis will confirm which candidates regulate miR locus processing through direct binding. Overall, our findings indicate that there are potentially dozens of RBPs with previously uncharacterized roles in miR biogenesis and regulation, and given the association of miRs with development and disease there is a great need for identifying these RBP:miR interactions.

STAR★METHODS

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

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-GAPDH (1:10,000 for WB)	Abcam	Cat# ab8245; RRID: AB_2107448
Mouse anti-α-TUB (1:5,000 for WB)	Abcam	Cat# ab7291; RRID: AB_2241126
Rabbit anti-ILF3 (1:10,000 for WB)	Bethyl	Cat# A303–651A; RRID: AB_11204576
Rabbit anti-LIN28B (1:2,000 for WB)	Bethyl	Cat# A303–588A; RRID: AB_11125329
Rabbit anti-PRPF8 (1:10,000 for WB)	Bethyl	Cat# A303–921A; RRID: AB_2620270
Rabbit anti-PTBP1 (1:5,000 for WB)	MBL International	Cat# RN011P; RRID: AB_1570645
Rabbit anti-SF3B4 (1:10,000 for WB)	Bethyl	Cat# A303–950A; RRID: AB_2620299
Rabbit anti-DDX3X (1:5,000 for WB)	Bethyl	Cat# A300–474A; RRID: AB_451009
Rabbit anti-SLTM (1:1,000 for WB)	Bethyl	Cat# A302–834A; RRID: AB_10632127
Rabbit anti-LARP7 (1:1,000 for WB)	Bethyl	Cat# A303–723A; RRID: AB_11205813
Rabbit anti-LARP4 (1:1,000 for WB)	Bethyl	Cat# A303–900A; RRID: AB_2620250
Rabbit anti-DKC1 (1:1,000 for WB)	Genetex	Cat# GTX109000; RRID: AB_11165396
Rabbit anti-RBFOX2 (1:1,000 for WB)	Bethyl	Cat# A300–864A; RRID: AB_609476
Rabbit anti-BUD13 (1:1,000 for WB)	Bethyl	Cat# A303–320A; RRID: AB_10952849
anti-FLAG beads for immunoprecipitation of recombinant protein	Clontech	635695
Chemicals, Peptides, and Recombinant Proteins		
polybrene	Sigma-Aldrich	H9268
TrypLE	Thermo Fisher Scientific	12604–013
Protease Inhibitor Cocktail III	Millipore Sigma	539134
TRIzol Reagent	Invitrogen	15596026
Expresshyb solution	clontech	636831
T4 PNK	NEB	M0201S
Antarctic phosphatase	NEB	M0289S
6% TBE retardation gel	Thermo Fisher Scientific	EC6365BOX
recombinant ILF3	Abcam	ab132543
Critical Commercial Assays		
Illumina TruSeq Small RNA library preparation kit	Illumina	RS-200–0012
miScript II RT Kit	QIAGEN	218160
miScript SYBR Green PCR Kit	QIAGEN	218073
Hs_RNU6–2_11 miScript Primer Assay	QIAGEN	MS00033740
Hs_mir-144_PR_1 miScript Precursor Assay	QIAGEN	MP00000924
Hs_mir-210_PR_1 miScript Precursor Assay	QIAGEN	MP00001505
T7 MegaShortScript kit	Thermo Fisher Scientific	AM1354
Lipofectamine3000	Thermo Fisher Scientific	L3000001

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ECL	Thermo Fisher Scientific	32106
ECL+	Thermo Fisher Scientific	32132
mirVana	Ambion	AM1560
Deposited Data		
Raw small RNA-Seq Data	This paper	GEO: GSE102497
Experimental Models: Cell Lines		
K562	ATCC	CCL-243
HepG2	ATCC	HB-8065
HEK293T	ATCC	CRL-1573
Oligonucleotides		
shRNAs for RBPs, see Table S1	N/A	N/A
Northern probe hsa-miR-20a-5p (5'-TACCTGCACTATAAGCACTTTA-3')	This paper	N/A
Northern probe has-miR-210–3p (5'-CAGCCGCTGTCACACGCACA-3')	This paper	N/A
Northern probe hsa-miR-144–5p (5'-CTTACAGTATATGATGATATCC-3')	This paper	N/A
205 bp of the miR-144 locus (5'-TAATACGACTCACTATAGGGAGAccct gacctgtcctgttctgcccccagcccctcacagtgcttttcaagccatgcttcctgtgcccccag tggggccctggctgggatatcatcatatactgtaagtttgcgatgagacactacagtatagatga tgtactagtccgggcacccccagctctggagcctgacaaggaggacaggagagatgctgca agcccaagaa-3')	This paper	N/A
171 bp of the miR-20a locus (5'-TAATACGACTCACTATAGGGAGAttat tgtgtcgatgtagaatctgcctggtctatctgatgtgacagcttctgtagcactaaagtgcttatag tgcaggtagtgtttagttatctactgcattatgagcacttaaagtactgctagctgtagaactcca gcttcggcctgtcgcccaatcaaactgtcctgtt-3')	This paper	N/A
g-block of [KpnI][BUD13][2xFLAG][STOPx2][BsiWI] (5'- cagcttggtaccatgg cggcagctccgccgctttccaaggccgagtatctgaagcgttacttgtccggggcagatgccgg cgtcgaccggggatctgagtccggtcgcaagcgtcgcaaaaagcggccgaagcctggcggg gccggcggcaagggaatgcggattgtggatgatgatgtgagctggacagctatctccacaacc aaactagaaaaggaggaagaggaagatgatggagatttgcctgtggtggcagagtttgtggatg agcggccagaagaggtaaagcagatggaggcctttcgttccagtgccaaatggaagcttctgg gaggccacaacgaagacctaccctcaaacagacattttcgtcacgataccccggattcatctcc taggagggtccgtcatGGCaccccagatccatctcctaggaaggaCCGAGTTGACATC CTGGCTCAATTTCCTAGGAAGGACCTTCATAGCATCCAGGGTCCttctcc cctcagaGGGGCACGACACGATAGCGACACTTCACCCCCAAGACGCAT CAGGCACGACTCATCAGATACATCACCACCAaggagggcccgtcatgattctc cagatccttctcccccaaggaggcctcagcataattcttcaggtGCATCCCCGCGAAGA GTACGGCATGACTCCCCAGATCCAAGCCCACCCCGACGAGCCCGCC ACGGCAGTTCAgatatctcttcccccagaagggtccataacaactcccctgacacatcta ggaggactcttggctcttcagacacacagcaactcagaagggcccgtcatgactcccctgatt tggctcctaatgtcacttattccctgcccagaaccaaaagtggtaaagccccagaaagagcct ctagcaagacttctccacattggaaggagtcaggagcctcccatttgtcattcccaaagaaca gcaaatatgagtatgaccctgacatctctcctcccacgaaaaaagcaagcaaaatcccatttgg agacaagaagcagcttgattccaaaggtgactgccagaaagcaactgattcagacctttcttct ccacggcataaacaaagtccagggcaccaggattctgattcagatctgtcacctccacggaa tagacctagacaccggagctctgattctgacctctctccaccaaggaggagacagaggacc	This paper	N/A

REAGENT or RESOURCE	SOURCE	IDENTIFIER
aaatcttctgattctgacctgtccccgcctcgaaggagtcagcctcctggaaagaaggctgca		
cacatgtattctggggctaaaactgggttggtgttaactgacatacagcgagaacagcagga		
gctcaaggaacaggatcaagaaaccatggcatttgaagctgaatttcaatatgctgaaaccgt		
atttcgagataagtctggtcgtaagaggaatttgaaactcgaacgtttagagcaaaggaggaa		
agcagaaaaggactcagagagagatgagctgtatgcccagtggggaaaagggcttgccca		
gagccggcaacagcaacaaaatgtggaggatgcaatgaaagagatgcaaaagcctctggc		
ccgctatattgatgacgaagatctggataggatgctaagagaacaggaaagagagggggac		
cctatggccaacttcatcaagaagaataaggccaaggagaacaagaataaaaaagtgaga		
cctcgctacagtggtccagcacctcctcccaacagatttaatatctggcctggatatcgctggg		
acggagtggacagatccaatggatttgaacagaagcgctttgccaggcttgccagcaagaag		
gcagtggaggaacttgcctacaaatggagtgttgaggatatgGATTATAAGGATGATG		
ATGATAAAGATTATAAGGACGACGATGACAAGTAATAAcgtacgcg acgg-3')		
EMSA miR-144 locus: 5'-GCCCCCAGCCCCTCACAGTGCTTTTCAAGCCATGCTTCCTGTGCCCCCAGTGGGGCCCTG-3'	This paper	N/A
EMSA miR-20a locus: 5'TAAAGTGCTTATAGTGCAGGTAGTGT-3'	This paper	N/A
Recombinant DNA		
pEF5-FRT-V5-DEST	Thermo Fisher Scientific	V602020
Software and Algorithms		
Cutadapt version 1.8.1	Martin, 2011	https://cutadapt.readthedocs.io/en/stable/; RRID:SCR_011841
Bowtie version 1.1.1 (mapping) version 1.0.0 (removal of repetitive elements)	Langmead et al., 2009	http://bowtie-bio.sourceforge.net/;RRID:SCR_005476
Bedtools version v2.22.0–27-g6ae9016	Quinlan and Hall, 2010	http://bedtools.readthedocs.io/en/latest/;RRID:SCR_006646
DESeq2 version 1.10.1	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Other		
eCLIP peak bed files	ENCODE	https://www.encodeproject.org
RNA-Seq data	ENCODE	https://www.encodeproject.org
microRNA annotations (miRBase v20, GRCh37.p5 genome-build, NCBI_Assembly: GCA_000001405.6)	miRBase	miRBase v20; GRCh37.p5 genome-build;NCBI_Assembly:GCA_000001405.6
Mirtron annotations	Ladewig et al., 2012	http://genome.cshlp.org/cgi/doi/10.1101/gr.133553.111
Hg19 annotations	GENCODE	http://www.gencodegenes.org/releases/19.html
RepBase version 18.05	Jurka et al., 2005	http://www.girinst.org/repbase/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Gene W. Yeo, Ph.D., M.B.A. (geneyeo@ucsd.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture—Human chronic myelogenous leukemia cells (K562; female) were cultured in RPMI1640 supplemented with 10% FBS. Human hepatocellular carcinoma cells (HepG2; male) and human embryonic kidney cells (HEK293T) were cultured in DMEM supplemented with 10% FBS. All cell lines were maintained at 37°C in 5% CO_2. K562 and HepG2 cells were purchased from ATCC and all lines were routinely tested for mycoplasma using MycoAlert PLUS (Lonza).

METHOD DETAILS

eCLIP data, RNA-Seq data, and miRBase annotations—All eCLIP data and RNA-Seq was obtained from https://www.encodeproject.org. The $\log_2$(fold enrichment) and multiple-hypothesis adjusted p values for significant CLIPper defined peaks were calculated as described previously (Van Nostrand et al., 2016). Pre-miR and mature miRNA annotations were downloaded from miRBase.org (miRBase v20, GRCh37.p5 genome-build, NCBI_Assembly: GCA_000001405.6).

Generating a metric for candidate selection—smInput-normalized BED files for eCLIP biological replicates were combined into a single bedtool of shared peaks using bedtools intersect, where a shared peak was defined as at least one intersecting nucleotide. The miRBase annotation for miRNAs was filtered for pre-miRNAs and converted into a bedtool. eCLIP peaks at miRNA loci were then identified by using bedtools intersect to determine eCLIP peaks where either 100% of the peak overlapped with an annotated pre-miR (filled circles in scatterplots), or at least 1bp up to (peak length - 1bp) overlapped with an annotated pre-miR (empty circles in scatterplots). This procedure distinguished clusters that are more likely to be pre-miR versus pri-miR binders, respectively. Volcano plots of these pre-miR-intersecting eCLIP peaks were then generated using the eCLIP $\log_2$(fold enrichment) and padj (Van Nostrand et al., 2016) values. We selected $\log_2$(fold enrichment) and $-\log_{10}$(padj) cutoffs such that 25% of RBPs had eCLIP peaks within 10 or more miR loci ($\log_2$(fold enrichment) of 2 and a $-\log_{10}$(padj) of 3). For the volcano plots, if an RBP had multiple eCLIP clusters within a miR locus, we plotted the best cluster as defined first by the lowest padj, then the highest fold enrichment values, with a preference for clusters with 100% overlap with the pre-miR within that miRNA locus. miRNA locus regions of origin were determined by the genic location of the annotated pre-miR using gencode hg19 annotations and published mirtron annotations (Ladewig et al., 2012). These were assigned in the following order: snoRNA, lncRNA, mirtron, distal intron (500bp), proximal intron (500bp), coding exons, and uncategorized.

Lentiviral preparation—Viral packaging of the selected shRNAs was performed using the pLKO.1 lentiviral vector containing a shRNA against the RBP of interest or a control againstEGFP. shRNA plasmids were co-transfected intoHEK293T cells with viral packaging plasmids MDL, RevRSV, and VSVG using polyethyleneimine (PEI). The media was changed after 5 hours and virus-containing media was harvested at 48 h post-transfection, filtered through 0.22 μm filter and either used fresh or stored at −80°C. TRC IDs for shRNA hairpins are: BUD13: TRCN0000074896, DDX3X: TRCN0000000003, DKC1: TRCN0000039738, ILF3: TRCN0000329787, LARP4: TRCN0000161048, LARP7: TRCN0000122544, LIN28B: TRCN0000144508, PRPF8: TRCN0000075112, PTBP1: TRCN0000001062, RBFOX2: TRCN0000074544, SF3B4: TRCN0000000039, SLTM: TRCN0000135106.

RBP knockdown—HepG2 cells were seeded into 6-well plates at a density of 5×10^5 cells/well and expanded overnight to 50%−60% confluency. K562 cells were seeded into 6-well plates at a density of 1.4×10^6 cells/well. 200 μl/well (HepG2) or 400 μl/well (K562) of filtered shRBP or shControl lentiviral supernatant was added to each well with 8 μg/ml

polybrene (Catalog Number H9268, Sigma-Aldrich). Media was changed at 24, 72, and 95 hours post infection and supplemented with 3 μg/ml puromycin. Six days post infection the cells were harvested. HepG2 cells were rinsed with PBS and detached using TrypLE (12604–013). Each well of the 6-well plate was split into two aliquots. K562 cells were divided into four aliquots/well. Cell aliquots were then pelleted and either snap frozen and stored at −80°C or immediately processed for total RNA, small RNA, or western blot.

Western blotting—Either fresh cell pellets or frozen cell pellets thawed on ice were resuspended in lysis buffer (50 mMTrisHCl pH 7.4,100 mM NaCl, 1% NP-40,0.1% SDS, 0.5% sodium deoxycholate, 1:200 Protease Inhibitor Cocktail III (EMD Millipore)).The suspensions were then sonicated for five cycles of 30 s intervals at 4°C. Protein was quantified using the BCA assay (Pierce). 30 μg of protein was heat denatured and run on 4% −12% NuPAGE Novex Bis-Tris protein gels (ThermoFisher) and transferred to PVDF membranes. The membranes were then blocked in 5% milk in TBS buffer for 1h at 25°C. Membranes were incubated in primary antibody diluted in 5% milk TBS-tween overnight at 4°C. Primary antibodies used were: GAPDH (Abcam ab8245; 1:10,000), α-TUB (Abcam ab7291, 1:5000), ILF3 (Bethyl A303–651A; 1:10,000), LIN28B (Bethyl A303–588A; 1:2,000), PRPF8 (Bethyl A303–921A; 1:10,000), PTBP1 (MBL RN011P; 1:5,000), SF3B4 (Bethyl A303–950A; 1:10,000), DDX3X (Bethyl A300–474A; 1:1000), SLTM (Bethyl A302–834A; 1:1000), LARP7 (Bethyl A303–723A; 1:1000), LARP4 (Bethyl A303–900A; 1:1000), DKC1 (Genetex GTX109000; 1:1000), RBFOX2 (Bethyl A300–864A; 1:1000), BUD13 (Bethyl A303–320A; 1:1000). Membranes were washed and probed with a 1:10,000 dilution of secondary antibody prepared in 5% milk TBS-tween for 1 hour at 25°C, washed, and developed with ECL or ECL+ (Pierce).

Total RNA isolation—Cell pellets were resuspended in 1ml TRIzol Reagent (Invitrogen) then either stored at −80°C or immediately processed for total RNA according to the manufacturer instructions.

Small RNA isolation—Either fresh cell pellets or frozen cell pellets thawed on ice were enriched for small RNAs using the mirVana (Ambion) kit, following the manufacturer protocol for enrichment of small RNAs.

Small RNA-seq—Either 1 μg of total RNA or 500–1000 ng of small RNAs were used as input for the Illumina TruSeq Small RNA library preparation kit. Libraries were validated by Tapestaion (Agilent), quantified by Qubit (Thermo Fisher) and sequenced on the Illumina HiSeq 4000 platform for 50 cycles. Small RNA-Seq reads were trimmed of adapters using Cutadapt version 1.8.1 (Martin, 2011) with parameters -f fastq-match-read-wildcards-times 2 -e 0.0 -O 5-quality-cutoff 6 -m 18 -b GUUCAGAGUUCUACAGUCCGA CGAUC -b GCCTTGGCACCCGAGAATTCCA - b TGGAATTCTCGGGTGCCAAGG -b GAAUUCCACCACGUUCCCGUGG -b AAT GATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA. Reads were then mapped against a database of repetitive elements derived from RepBase (version 18.05, Jurka et al., 2005) using Bowtie (Langmead et al., 2009) (version 1.0.0) with parameters -S -q -p 16 -e 100 -l 20. Reads that did not map to Repbase sequences were aligned to the hg19

human genome (UCSC assembly) using Bowtie (Langmead et al., 2009) (version 1.1.1) with parameters -p 8 -k 1 -m 10 -l 25-best-chunkmbs 128-sam. SAM files were then converted to BAM files, sorted and indexed using samtools (Li et al., 2009). Counts were calculated with feature Counts (Liaoet al., 2014). For visualization on the UCSC Genome Browser, BAM files were RPM (reads per million) normalized against the total number of usable reads in that dataset and converted to bedgraph format with genomeCoverageBed from Bedtools v2.22.0–27-g6ae9016 (Quinlan and Hall, 2010) then to bigwig for generation of trackhubs. Experiments using shRBPs were performed in biological duplicate, and shControl in biological quadruplicate for each cell line.

Differential expression—Differential expression of small RNA-Seq data was calculated using DESeq2 version 1.10.1 (Love et al., 2014); duplicate shRBP knockdown samples were paired with 4 replicates of shControl treated cells for each cell line. miRNAs were considered significantly changed between shControl and shRBP if the Benjamini-Hochberg adjusted $p < 0.05$. Differential expression of encode RNA-Seq data was calculated using DESeq2 version 1.10.1 (Love et al., 2014); shControl and shRBP were both performed in duplicate.

Northern blot analysis—2–5 μg of total RNA or small RNAs were diluted in formamide buffer, heat denatured at 95°C for 5 minutes, and immediately run on 15% NuPAGE NovexTBE-Urea gels (ThermoFisher) in 0.5XTBE. Size markers were ssRNAlow range ladder (NEB N0364S) and mi-croRNA marker (NEB N2102S). The gel was stained with SybrGold (ThermoFisher S11494) for 5 minutes and imaged to confirm RNA integrity. RNA was then electro-transferred to a hybond plus membrane XL (GE Lifesciences) in 0.5X TBE. The membrane was cross-linked at 120 mJ/cm^2 then blocked in pre-warmed Expresshyb solution (Clontech) for 1h at 42°C. During blocking, the probes were labeled as follows: 3 μL of 6 μM DNA oligonucleotide probe, 4 μL of 5X T4 PNK buffer, 5 μL of ^{32}P-γ-ATP, 1 μL of T4 PNK (NEB M0201S) and 7 mL of molecular biology grade water were mixed and incubated at 37°C for 1 hour with rotation. The reaction was purified using a nucleotide removal kit (QIAGEN), denatured at 95°C for 5 minutes then placed on ice for 2 min. The blocking solution was removed and replaced with fresh, pre-warmed Expresshyb solution containing the radiolabeled/denatured probe. The membrane was incubated with probe overnight at 37°C with rotation. The membrane was washed twice quickly with wash I (2X SSC, 0.05% SDS), then two times for 30 minutes in wash I. The membrane was then washed twice for 15 minutes in wash II (0.1X SSC, 0.1% SDS). Excess liquid was removed from the membrane, which was then placed in a sheet protector and exposed to maximum sensitivity autoradiography film (Sigma Z363022) in the presence of an intensifying screen at −80°C. Target probe sequences: hsa- miR-20a-5p (5’-TACCTGCACTATAAGCACTTTA-3’), has-miR-210–3p (5’-CAGCCGCTGTCACACGCACA-3’) and hsa-miR-144–5p (5’-CTTACAGTATATGATGATATCC-3’).

Pre-miR qPCR—Total RNA was isolated using TRIzol reagent (Invitrogen) according to manufacturer instructions. RNA was then reverse transcribed using themiScript II RT Kit (QIAGEN) according to manufacturer instructions. The qPCR was then performed using the

miScriptSYBR Green PCR Kit, Hs_RNU6–2_11 miScript Primer Assay (control), Hs_mir-144_PR_1 miScript Precursor Assay, and Hs_mir-210_PR_1 miScript Precursor Assay (QIAGEN) according to manufacturer instructions.

Electrophoretic mobility shift assay: ILF3—RNA was *in vitro* transcribed from a gBlock using the T7 MegaShortScript kit (Thermo AM1354). The gBlock template contained the T7 promoter sequence as well as 205 bp of the miR-144 locus (5'-TAATACGACTCACTATAGGGAGAccctgacctgtcctgttctgcccc cagcccctcacagtgcttttcaagccatgcttcctgtgcccccagtggggccctggctgggatatcatcatatactgtaagtttgcgatg agacactacagtatagatgatgtactag tccgggcacccccagctctggagcctgacaaggaggacaggagagatgctgcaagcccaagaa-3') or 171 bp of the miR-20a locus (5'-TAATACGACTCA CTATAGGGAGAttattgtgtcgatgtagaatctgcctggtctatctgatgtgacagcttctgtagcactaaagtgcttatagtgca ggtagtgtttagttatctactgcattatgagca cttaaagtactgctagctgtagaactccagcttcggcctgtcgcccaatcaaactgtcctgtt-3'). The 5' ends of the IVT product were de-phosphorylated using Antarctic phosphatase (NEB M0289S) with 2 pmol 5' ends in a 20 μL reaction. The reaction was incubated at 37°C for 60 minutes then the enzyme was deactivated at 80°C for 2 minutes. The RNA was then radioactively labeled as follows: 20 μL de-phosphorylated RNA, 5 μL 10X T4 PNK reaction buffer, 3 μL ^{32}P-γ-ATP, 2 μL T4 PNK (NEB M0201S) in a 50 μL reaction were mixed and incubated at 37°C for 30 minutes. During incubation, the 6% TBE retardation gel (ThermoFisher EC6365BOX) was pre-run at 100V in 0.5X TBE buffer. Labeled RNA oligos were then cleaned by spin column (GE Illustra microspin G-25) and EDTA was added to 0.1mM. RNA was then scintillation counted, and 30,000 CPM/binding reaction was heated to 95°C for 1 minute and immediately chilled on ice. These pre-heated, radiolabeled oligos were then incubated with decreasing amounts (0.1, 0.04, 0.02, or 0 ug) of recombinant ILF3 (Abcam ab132543), 4 μL 5X EMSA binding buffer (5X: 100mM HEPES pH 7.9, 375 mM KCl, 2.5 mM DTT, 0.05% Tween 20, 50% glycerol), and water up to 20 μl. The binding reaction was incubated on ice for 25 minutes, followed by the addition of loading buffer (final concentration: 0.03 mg/ml heparin, 0.017% bromophenol blue, 0.017% xylene cyanol, 0.5 mM DTT), and an additional 5-minute incubation on ice. Samples were then resolved on the pre-run retardation gel at 100V. The gel was then dried onto filter paper under vacuum for 1 h at 80°C, exposed to high-sensitivity film in the presence of an intensifier screen overnight at −80°C, and imaged.

Tagging and purification of BUD13–2xFLAG—The pEF5-FRT-V5-DEST gateway vector (ThermoFisher V602020) was first digested with KpnI (2 μl KpnI-HF, 5 μl 10X cutsmart buffer, 5 μg plasmid, 13 μl water) for 2 h at 37°C. NaCl was added to 100 mM, and 2 μl BsiWI was also added followed by a 2 h incubation at 55°C. the digested backbone was then gel purified using the QIAGEN gel extraction kit according to manufacturer instructions. A g-block of [KpnI][BUD13][2xFLAG][STOPx2][BsiWI] was purchased from IDT (5'- cagcttggtaccatggcggcagctccgccgcttt ccaaggccgagtatctgaagcgttacttgtccggggcagatgccggcgtcgaccggggatctgagtccggtcgcaagcgtcgcaa aaagcggccgaagcctggcggggccg gcggcaagggaatgcggattgtggatgatgatgtgagctggacagctatctccacaaccaaactagaaaaggaggaagaggaaga tgatggagatttgcctgtggtggcagag

tttgtggatgagcggccagaagaggtaaagcagatggaggcctttcgttccagtgccaaatggaagcttctgggaggccacaacga
agacctaccctcaaacagacattttcgtc
acgataccccggattcatctcctaggagggtccgtcatGGCaccccagatccatctcctaggaaggaCCGAGTTGACAT
CCTGGCTCAATTTCCTAGGAAG
GACCTTCATAGCATCCAGGGTCCttctcccctcagaGGGGCACGACACGATAGCGACACT
TCACCCCCAAGACGCATCAGGCACGAC
TCATCAGATACATCACCACCAaggagggcccgtcatgattctccagatccttctcccccaaggaggcctcagcata
attcttcaggtGCATCCCCGCGAAGA
GTACGGCATGACTCCCCAGATCCAAGCCCACCCCGACGAGCCCGCCACGGCAGTT
CAgatatctcttcccccagaagggtccataacaactcc
cctgacacatctaggaggactcttggctcttcagacacacagcaactcagaagggcccgtcatgactcccctgatttggctcctaatgt
cacttattccctgcccagaaccaaaag
tggtaaagccccagaaagagcctctagcaagacttctccacattggaaggagtcaggagcctcccatttgtcattcccaaagaacag
caaatatgagtatgaccctgacatctctc
ctccacgaaaaaagcaagcaaaatcccattttggagacaagaagcagcttgattccaaaggtgactgccagaaagcaactgattcag
acctttcttctccacggcataaacaaa
gtccagggcaccaggattctgattcagatctgtcacctccacggaatagacctagacaccggagctctgattctgacctctctccacca
aggaggagacagaggaccaaatcttc
tgattctgacctgtccccgcctcgaaggagtcagcctcctggaaagaaggctgcacacatgtattctggggctaaaactgggttggtg
ttaactgacatacagcgagaacagcag
gagctcaaggaacaggatcaagaaaccatggcatttgaagctgaatttcaatatgctgaaaccgtatttcgagataagtctggtcgtaa
gaggaatttgaaactcgaacgtttagag
caaaggaggaaagcagaaaaggactcagagagagatgagctgtatgcccagtggggaaaagggcttgcccagagccggcaaca
gcaacaaaatgtggaggatgcaatga
aagagatgcaaaagcctctggcccgctatattgatgacgaagatctggataggatgctaagagaacaggaaagagagggggaccc
tatggccaacttcatcaagaagaataag
gccaaggagaacaagaataaaaaagtgagacctcgctacagtggtccagcacctcctcccaacagatttaatatctggcctggatat
cgctgggacggagtggacagatccaa
tggatttgaacagaagcgctttgccaggcttgccagcaagaaggcagtggaggaacttgcctacaaatggagtgttgaggatatgG
ATTATAAGGATGATGATGATA
AAGATTATAAGGACGACGATGACAAGTAATAAcgtacgcgacgg-3′) and digested as above. The digested g-Block and backbone were then ligated as follows: 2 μL 10XT4 DNA ligase buffer, 50 ng insert, 37.5 ng backbone, 1 μLT4 DNA ligase in a 20 μL reaction which was incubated for 1 h at room temperature then at 65°C for 10 minutes and transformed.

1×10^7 HEK293T cells were plated on a PDL-coated 15 cm dish and cultured with DMEM + 10% FBS overnight. For transfection, the Lipofectamine3000 kit (Invitrogen) was used according to the manufacturer's instructions. 24 h post-transfection the media was replaced, and 48 h post-transfection the cells were counted and harvested. 2×10^7 cells were snap frozen per tube. 2×10^7 cells were lysed in 1200 μL of cold low-detergent lysis buffer (50mM Tris-HCl, 150mM NaCl, 0.5% NP-40, and 10% glycerol) and incubated for 15 minutes on ice. The cells were then sonicated for 5 minutes (30 s on/off). Afterward the cells were centrifuged at 20,000 × g for 20 minutes at 4°C and the supernatant was taken for further processing. 25 μl was saved as input and the remaining lysate was incubated with 100 μl of anti-FLAG beads (Clontech 635695) that had been washed twice with low-

deterged lysis buffer. The lysate was incubated with the beads for 2 h at 4°C and then separated with a magnetic rack. 25 μl was taken as IP-supernatant and the remaining was discarded. The beads were washed three times with high salt buffer (1 M NaCl, 50 mM Tris-HCl, 1 mM EDTA, 1% NP-40 and 10% glycerol) and then washed three times with TBS-G (TBS with 10% glycerol). In the first elution step, the beads were incubated in 60 μl EMSA binding buffer with 100 μg/ml 3x FLAG-peptidefor 4 minutes at 4°C. The supernatant was then taken as Elution 1. To get Elution 2, the beads were then incubated again with 60 μl EMSA binding buffer with 250 μg/ml 3x FLAG-peptide for 30 minutes at 4°C. The remaining beads were re-suspended in 60 μl TBS-G and incubated at 70°C for 10 minutes to elute the remaining BUD13–2xFLAG that is precipitated on the beads to control for the elution efficiency. To quantify the relative eluted amount western blot was performed using 5% (3 μl) of elution 1 and elution 2, and 80% (20 μl) of the remaining fractions.

Electrophoretic mobility shift assay: BUD13–2xFLAG—RNA oligonucleotides for BUD13 target (miR-144 locus: 5'-GCCCCCAGCCCCTCACAGTGCTTTTCAAGCCATGCTTCCTGTGCC CCCAGTGGGGCCCTG-3') and non-target (miR-20a locus: 5'TAAAGTGCTTATAGTGCAGGTAGTGT-3') were purchased from IDT. Oligonucleotides were radioactively labeled as follows: 50 pmol RNA oligonucleotide, 5 μL of 10X T4 PNK reaction buffer, 3 mL of ^{32}P-γ-ATP, 2 μL of T4 PNK (NEB M0201S) in a 50 μL reaction were mixed and incubated at 37°C for 30 minutes. During incubation, the 6% retardation gel (ThermoFisher EC6365BOX) was pre-run at 100V in 0.5XTBE buffer. Labeled RNAoligos were then cleaned by spin column (GE Illustra microspin G-25) and EDTA was added to0.1mM. RNA was then scintillation counted, and 10,000 CPM/binding reaction was heated to 95°C for 5 minutes and immediately chilled on ice. These pre-heated, radiolabeled oligos were then incubated with increasing amounts (0, 1, 2, 4, 10 μl) of purified BUD13–2XFLAG, 4 μL 5X EMSA binding buffer (5X: 100mM HEPES pH 7.9, 375 mM KCl, 2.5 mM DTT, 0.05% Tween 20, 50% glycerol, 500 ng/μl tRNA), and water up to 20 μl. The binding reaction was incubated on ice for 25 minutes, followed by the addition of loading buffer (final concentration: 0.03 mg/ml heparin, 0.017% bromophenol blue, 0.017% xylene cyanol, 0.5 mM DTT), and an additional 5-minute incubation on ice. Samples were then resolved on the pre-run retardation gel at 100V. The gel was then dried onto filter paper under vacuum for 1 h at 80°C, exposed to high-sensitivity film in the presence of an intensifier screen overnight at −80°C, and imaged.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses, values for sample size, and software used are detailed in the Method Details sections above, associated with each experiment, as well as in the figure legends. To summarize, eCLIP fold change and p-adjusted values were calculated as previously described (Van Nostrand et al., 2016). Fold change and p-adjusted values for small RNA-Seq differential expression were calculated with DESeq2 version 1.10.1 (Love et al., 2014) using a Benjamini-Hochberg correction. Unless otherwise indicated in the Method Details, figure panel, or figure legend, statistical significance was defined as a p-adjusted < 0.05. For bar plots, error bars represent 1 standard deviation, and unless otherwise indicated n = 2 biological replicates (Figure 5D n = 4 biological replicates); asterisks (*) indicating p values

are indicated in the figure panel. For small RNA-sequencing and RNA-Seq analysis, count data was normalized to read depth (reads per million, RPM). For qPCR analysis, fold change in expression was calculated as $2^{-\Delta\Delta CT}$. Northern and western blots were quantified using ImageJ software, and fold change expression was calculated versus loading control then versus shControl. All statistical calculations were conducted using R (version 3.2.2) and python (version 2.7.12).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors would like to thank members of the Yeo lab, especially Tomas Bos and Fernando Martinez for critical reading of the manuscript, and Gabriel Pratt for critical feedback of the eCLIP analysis. This work was supported by grants from the NIH (HG004659, HG007005). G.W.Y. is an Alfred P. Sloan Research Fellow. J.K.N. is an Achievement Rewards for College Scientists Foundation Fellow and was partially supported by the Genentech Foundation fellowship and the University of California, San Diego, Cancer Training Program through an institutional training grant from the National Institute of General Medical Sciences (T32CA067754).

REFERENCES

Alawi F, and Lin P (2010). Loss of dyskerin reduces the accumulation of a subset of H/ACA snoRNA-derived miRNA. Cell Cycle 9, 2467–2469. [PubMed: 20519940]

Bartel DP (2009). MicroRNAs: Target recognition and regulatory functions. Cell 136,215–233. [PubMed: 19167326]

Chen Y, Zubovic L, Yang F, Godin K, Pavelitz T, Castellanos J, Macchi P, and Varani G (2016). Rbfox proteins regulate microRNA biogenesis by sequence-specific binding to their precursors and target downstream Dicer. Nucleic Acids Res. 44, 4381–4395. [PubMed: 27001519]

Denli AM, Tops BB, Plasterk RH, Ketting RF, and Hannon GJ (2004). Processing of primary microRNAs by the Microprocessor complex. Nature 432, 231–235. [PubMed: 15531879]

Dziembowski A, Ventura AP, Rutz B, Caspary F, Faux C, Halgand F, Laprévote O, and Séraphin B (2004). Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. EMBO J. 23, 4847–4856. [PubMed: 15565172]

Fasanaro P, D'Alessandra Y, Di Stefano V, Melchionna R, Romani S, Pompilio G, Capogrossi MC, and Martelli F (2008). MicroRNA-210 modulates endothelial cell response to hypoxia and inhibits the receptor tyrosine kinase ligand Ephrin-A3. J. Biol. Chem. 283, 15878–15883. [PubMed: 18417479]

Ge J, Crosby SD, Heinz ME, Bessler M, and Mason PJ (2010). SnoRNA microarray analysis reveals changes in H/ACA and C/D RNA levels caused by dyskerin ablation in mouse liver. Biochem. J. 429, 33–41. [PubMed: 20423331]

Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, and Shiekhattar R (2004). The Microprocessorcomplexmediatesthegen- esis of microRNAs. Nature 432, 235–240. [PubMed: 15531877]

Han J, Lee Y, Yeom KH, Kim YK, Jin H, and Kim VN (2004). The Drosha-DGCR8 complex in primary microRNA processing. Genes Dev. 18, 3016–3027. [PubMed: 15574589]

Hao J, Duan F-F, andWang Y (2017). MicroRNAs and RNA binding protein regulators of microRNAs in the control of pluripotency and reprogramming. Curr. Opin. Genet. Dev. 46, 95–103. [PubMed: 28753462]

Helwak A, Kudla G, Dudnakova T, and Tollervey D (2013). Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell 153, 654–665. [PubMed: 23622248]

Heo I, Joo C, Cho J, Ha M, Han J, and Kim VN (2008). Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. Mol. Cell 32, 276–284. [PubMed: 18951094]

Heo I, Joo C, Kim Y-K, Ha M, Yoon M-J, Cho J, Yeom K-H, Han J, and Kim VN (2009). TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. Cell 138, 696–708. [PubMed: 19703396]

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, and Walichiewicz J (2005). Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. 110, 462–467. [PubMed: 16093699]

Kim B, Jeong K, and Kim VN (2017). Genome-wide mapping of DROSHA cleavage sites on primary microRNAs and noncanonical substrates. Mol. Cell 66, 258–269. [PubMed: 28431232]

Kozomara A, and Griffiths-Jones S (2011). miRBase: Integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 39, D152–D157. [PubMed: 21037258]

Ladewig E, Okamura K, Flynt AS, Westholm JO, and Lai EC (2012). Discovery of hundreds of mirtrons in mouse and human small RNA data. Genome Res. 22, 1634–1645. [PubMed: 22955976]

Langmead B, Trapnell C, Pop M, and Salzberg SL (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25. [PubMed: 19261174]

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, and Durbin R; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. [PubMed: 19505943]

Liao Y, Smyth GK, and Shi W (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930. [PubMed: 24227677]

Lin E, Kuo P-H, Liu Y-L, Yang AC, Kao C-F, and Tsai S-J (2016). Association and interaction ofAPOA5, BUD13, CETP, LIPA and health-related behavior with metabolic syndrome in a Taiwanese population. Sci. Rep. 6, 36830. [PubMed: 27827461]

Loffreda A, Rigamonti A, Barabino SM, and Lenzken SC (2015). RNA-binding proteins in the regulation of miRNAActivity:Afocus on neuronal functions. Biomolecules 5, 2363–2387. [PubMed: 26437437]

Lovci MT, Ghanem D, Marr H, Arnold J, Gee S, Parra M, Liang TY, Stark TJ, Gehman LT, Hoon S, et al. (2013). Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. Nat. Struct. Mol. Biol. 20, 1434–1442. [PubMed: 24213538]

Love MI, Huber W, and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. [PubMed: 25516281]

Macias S, Plass M, Stajuda A, Michlewski G, Eyras E, and Caceres JF (2012). DGCR8 HITS-CLIP reveals novel functionsforthe microprocessor. Nat. Struct. Mol. Biol. 19, 760–766. [PubMed: 22796965]

Martin M (2011). Cutadapt removes adapter sequences from high- throughput sequencing reads. EMBnet J. 17, 10–12.

Michlewski G, and Caceres JF (2010). Antagonistic role of hnRNP A1 and KSRP in the regulation of let-7a biogenesis. Nat. Struct. Mol. Biol. 17, 1011–1018. [PubMed: 20639884]

Nam Y, Chen C, Gregory RI, Chou JJ, and Sliz P (2011). Molecular basis for interaction of let-7 microRNAs with Lin28. Cell 147, 1080–1091. [PubMed: 22078496]

Persengiev S, Kondova I, Otting N, Koeppen AH, and Bontrop RE (2011). Genome-wide analysis of miRNA expression reveals a potential role for miR-144 in brain aging and spinocerebellar ataxia pathogenesis. Neurobiol. Aging 32, e17–e27.

Quinlan AR, and Hall IM (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. [PubMed: 20110278]

Ren K, Liu QQ, An ZF, Zhang DP, and Chen XH (2017). MiR-144 functions as tumor suppressor by targeting PIM1 in gastric cancer. Eur. Rev. Med. Pharmacol. Sci. 21, 3028–3037. [PubMed: 28742204]

Scott MS, Avolio F, Ono M, Lamond AI, and Barton GJ (2009). Human miRNA precursors with box H/ACA snoRNA features. PLoS Comput. Biol. 5, e1000507.

Shiohama A, Sasaki T, Noda S, Minoshima S, and Shimizu N (2007). Nucleolar localization of DGCR8 and identification of eleven DGCR8-associ-ated proteins. Exp. Cell Res. 313, 4196–4207. [PubMed: 17765891]

Sundararaman B, Zhan L, Blue SM, Stanton R, Elkins K, Olson S, Wei X,Van Nostrand EL, Pratt GA, Huelga SC, et al. (2016). Resourcesforthe comprehensive discovery offunctional RNA elements. Mol. Cell 61, 903–913. [PubMed: 26990993]

Treiber T, Treiber N, Plessmann U, Harlander S, Daiss JL, Eichner N, Lehmann G, Schall K, Urlaub H, and Meister G (2017). A compendium of RNA-binding proteins that regulate microRNA biogenesis. Mol. Cell 66, 270–284. [PubMed: 28431233]

Triboulet R, Pirouz M, and Gregory RI (2015).Asingle Let-7 microRNA by-passes LIN28-mediated repression. Cell Rep. 13, 260–266. [PubMed: 26440890]

Valentin-Vega YA, Wang YD, Parker M, Patmore DM, Kanagaraj A, Moore J, Rusch M, Finkelstein D, Ellison DW, Gilbertson RJ, et al. (2016). Cancer-associated DDX3X mutations drive stress granule assembly and impair global translation. Sci. Rep. 6, 25996. [PubMed: 27180681]

Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundararaman B, Blue SM, Nguyen TB, Surka C, Elkins K, et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat. Methods 13, 508–514. [PubMed: 27018577]

Yang R, Gaidamakov SA, Xie J, Lee J, Martino L, Kozlov G, Crawford AK, Russo AN, Conte MR, Gehring K, and Maraia RJ (2011). La-related protein 4 binds poly(A), interactswith the poly(A)-binding protein MLLEdomain viaavariant PAM2w motif, and can promote mRNAstability. Mol. Cell. Biol. 31, 542–556. [PubMed: 21098120]

Yoshida T, Kato K, Oguri M, Horibe H, Kawamiya T, Yokoi K, Fujimaki T, Watanabe S, Satoh K, Aoyagi Y, et al. (2011a). Association of polymor- phismsofBTN2A1 and ILF3with myocardial infarction in Japanese individuals with or without hypertension, diabetes mellitus or chronic kidney disease. Int. J. Mol. Med. 27, 745–752. [PubMed: 21347509]

Yoshida T, Kato K, Oguri M, Horibe H, Kawamiya T, Yokoi K, Fujimaki T, Watanabe S, Satoh K, Aoyagi Y, et al. (2011b). Association of polymor-phismsofBTN2A1 and ILF3with myocardial infarction in Japanese individuals with different lipid profiles. Mol. Med. Rep. 4, 511–518. [PubMed: 21468600]

Zhao L, Mao Y, Zhao Y, and He Y (2016). DDX3X promotes the biogenesis of a subset of miRNAs and the potential roles they played in cancer development. Sci. Rep. 6, 32739. [PubMed: 27586307]

Highlights

- Analysis of eCLIP datasets identifies 116 candidate RBPs that bind miRNA loci
- ~68% of identified direct RBP:miR locus interactions are cell line specific
- Knockdown of candidate RBPs affects microRNA levels
- Biochemical assays reveal steps in microRNA processing that are affected

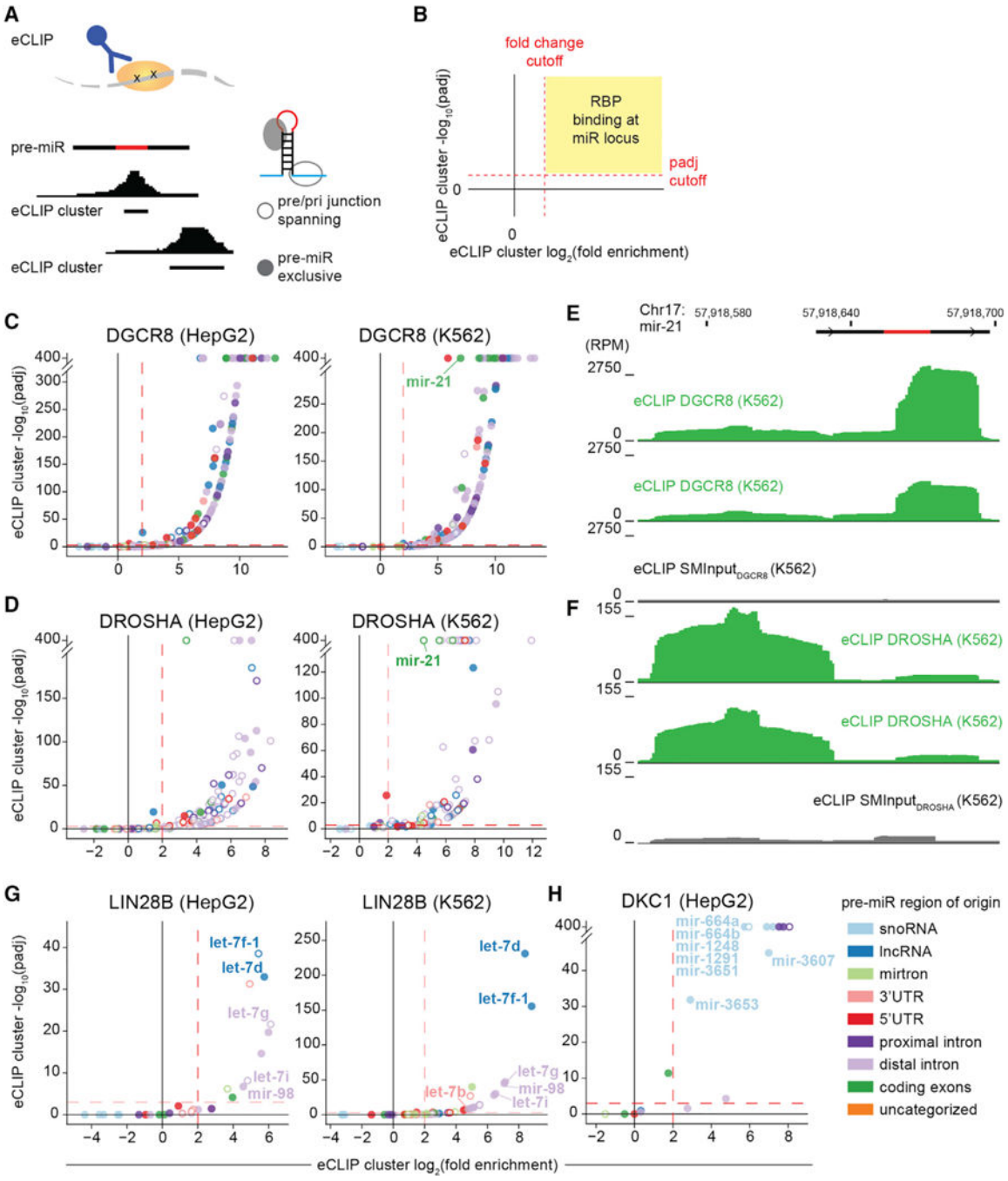


Figure 1. Identification of RBPs with Clusters at Pre-miR Loci

(A and B) Experimental approach for the identification and validation of direct RBP regulation of miR processing. For each of 181 eCLIP experiments characterizing RNA binding of 126 RBPs in HepG2 and K562 cell lines, eCLIP clusters obtained from encodeproject.org were filtered for those intersecting miRBase annotated pre-miRs(A). Thedata werethen visualized as volcano plots ofeCLIPclusterlP/SMInput $\log_2$(fold enrichment) versus-$\log_{10}$(padj) (B). Clustersthat are exclusiveto pre-miRs are represented by filled circles (pre-miR exclusiveclusters), and clustersthat intersect pre-miRs by at least 1 nt

but are not exclusivetothe pre-miR are represented by open circles (pre-/pri-miR junction spanning clusters).

(C and D) Volcano plots of eCLIP clusters in HepG2 and K562 cell lines for known regulators of miR processing: (C) DGCR8; (D) DROSHA. Cutoffs of eCLIP IP/SMInput $\log_2$(fold enrichment) = 2 and $-\log_{10}$(padj) = 3 are denoted by a dashed red line.

(E and F) UCSC genome browsertracks of eCLIP data for (E) DGCR8 in K562 and (F) DROSHA in K562 forthe mir-21 locus demonstrating a pre-miR exclusive cluster (E) and a pre/pri-miR junction spanning cluster (F). The mir-21 locus is highlighted in the K562 volcano plot in (C) and (D).

(G and H) Volcano plots of eCLIP clusters in HepG2 and K562 cell lines for known regulators of a specific subset of miRs: (G) LIN28B (H) DKC1. Cutoffs of eCLIP IP/SMInput $\log_2$(fold enrichment) = 2 and $-\log_{10}$(padj) = 3 are denoted by a dashed red line. Known binding targets of LIN28B (D) and DKC1 (E) are highlighted. eCLIP data in K562 were not available for DKC1.

eCLIP IP/SMInput fold enrichment and padj were calculated as in Van Nostrand et al. (2016). padj = adjusted p value. Volcano plot clusters are colored based on the (non)-genic origin of the intersecting pre-miR based on ENSEMBL annotations and mirtron annotations from Ladewig et al. (2012). Volcano plot filled circles represent pre-miR exclusive clusters while open circles represent pre/pri-miR junction spanning clusters.

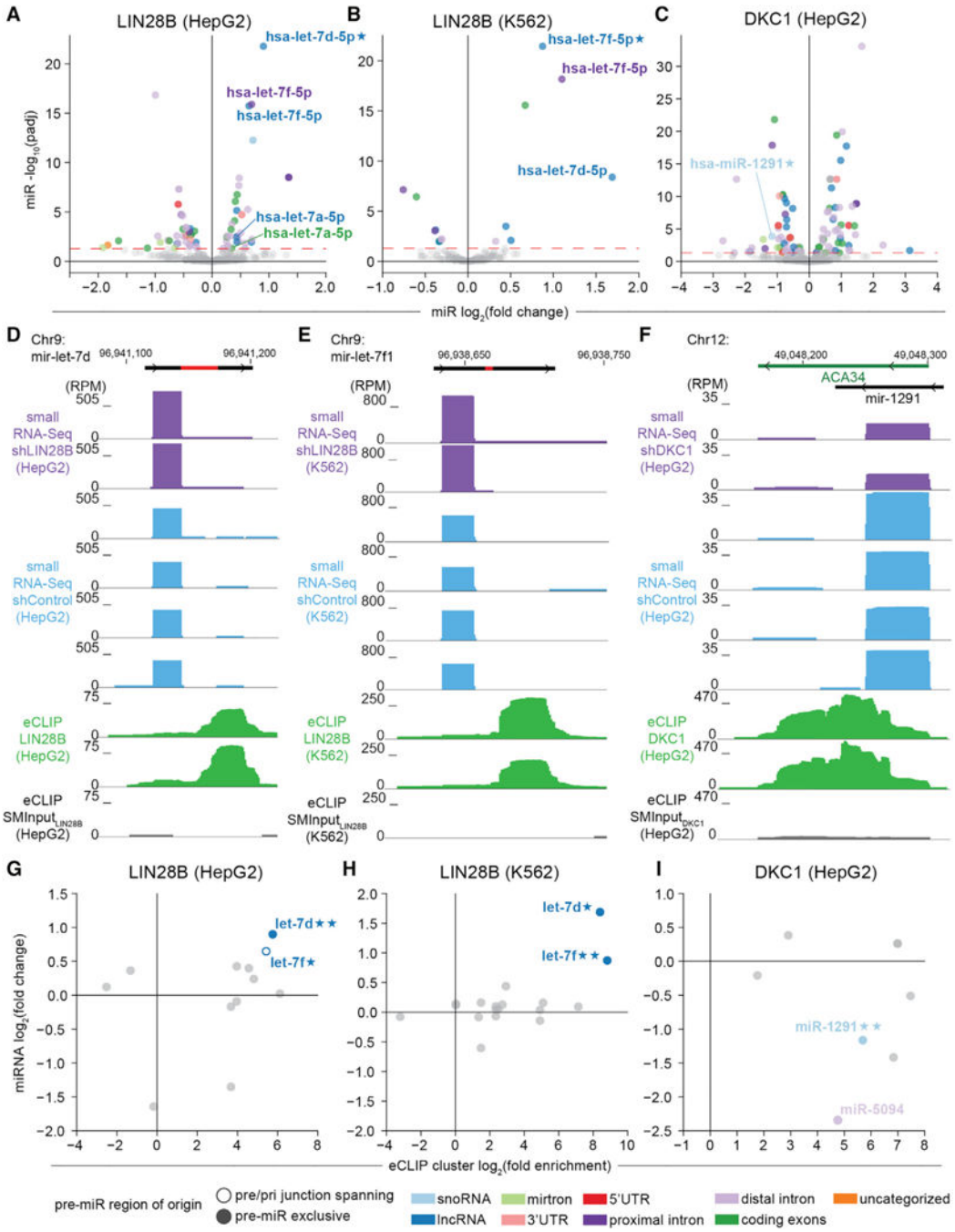


Figure 2. Suppression of LIN28B and DKC1 Causes Misregulation of Target miRs

(A-C) Volcano plots of small RNA-seq-identified mature miRs in control versus RBP knockdown cells. Small RNA-seq data were generated in nontargeting control (quadruplicate) and lentiviral shRNA knockdown of RBPs (duplicate) for LIN28B (A and B) and DKC1 (C). The miR knockdown/control $\log_2$(fold change) and padj were calculated based on DESEQ2 analysiswith Benjamini-Hochberg correction. Dotted red line represents padj = 0.05 cutoff. Labeled miRsare known targets of the RBP, and those viewed as browser

tracks in (D)-(F) are further denoted by a star ★). miRs are colored as in Figure 1 volcano plots.

(D-F) Representative UCSC genome browser tracks of normalized read density (RPM) of eCLIP data and small RNA-seq data for LIN28B in HepG2 cells (D), LIN28B in K562 cells (E), and DKC1 in HepG2 cells (F). eCLIP smInput (gray) and eCLIP IP (green) tracks show RBP binding at the miR locus for let-7d (D), let-7f1 (E), and mir-1291 (F). Small RNA-seq non-targeting controls (blue) and RBP knockdown tracks (purple) show upregulation of LIN28B-interacting miRs upon RBP knockdown (D and E) and downregulation of a DKC1-interacting sno-miR upon RBP knockdown (F). Controls were performed in quadruplicate and shRNA knockdowns in duplicate. Pre-miR hairpins are annotated as in Figure 1 with stems in black and loops in red.

(G-I) Integration of eCLIP and small RNA-seq data for LIN28B in HepG2 (G), LIN28B in K562 (H), and DKC1 in HepG2 (I) cells. eCLIP clusters at miR loci were further filtered forthose intersecting miRs detected by small RNA-seq. Forthese clusters, eCLIPcluster IP/smInput $\log_2$(fold enrichment) was plotted versus small RNA-seq miR $\log_2$(fold change). miRswith significant eCLIP clusters (padj < 0.05) that also changed significantly upon RBP knockdown (padj < 0.05; Benjamini-Hochberg correction) are highlighted and colored as in Figure 1. Known targets of the specified RBP as identified in Figure 1 are annotated with a star (★), and those plotted in (A)-(C) are annotated with a second star(★★). padj =adjusted pvalue. eCLIP cluster: IP/SMInput, SMInput, size-matched input; RPM, reads per million mapped reads. eCLIPexperiment= 1 RBP in 1 cell line. eCLIP/small RNA-seq integrated plot filled circles represent pre-miRexclusiveclusters, while open circles represent pre/pri-miR junction spanning clusters.

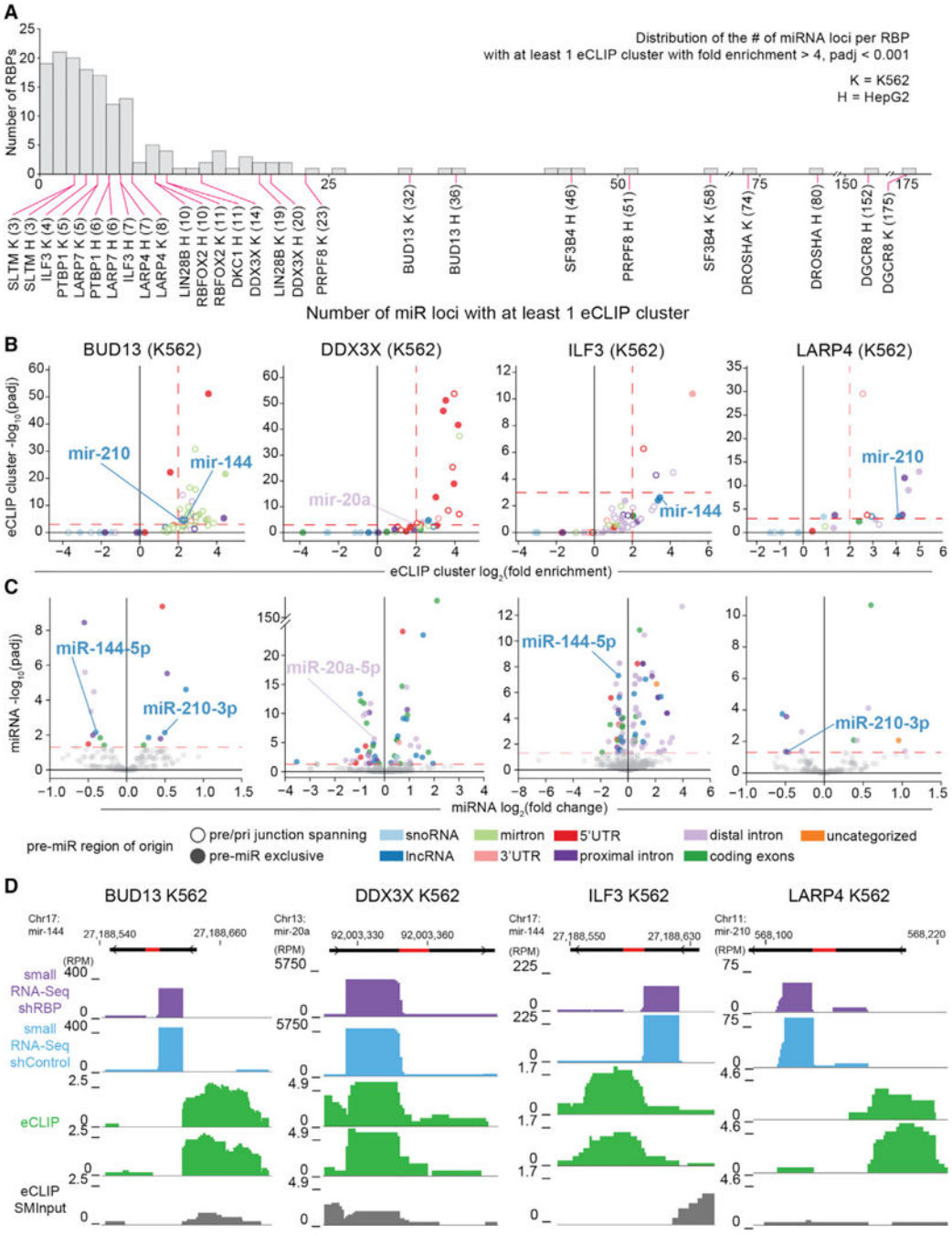


Figure 3. Identification of Putative miR-Binding RBPs by Knockdown and Small RNA-Seq

(A) Distribution ofthe numberof miR loci per RBPwith at least one miR locus eCLIP cluster passing the fold enrichment and p-adjusted cutoffs. The histogram is annotated forour positive controls (DGCR8, DKC1, LIN28B, and DROSHA) aswell asthe 10selected candidates (BUD13, DDX3X, ILF3, LARP4, LARP7, PRPF8, PTBP1, RBFOX2, SF3B4, and SLTM). The value in parentheses after each RBP is the number of eCLIP clusters intersecting miR loci.

(B) Volcano plots for eCLIP clusters at miR loci in K562 for BUD13, DDX3X, ILF3, and LARP4 (volcano plots for HepG2 and 6 additional candidates in Figure S2). Plotted eCLIP IP/SMInput $\log_2$(fold enrichment) versus $-\log-_{10}$(padj). Cutoffs of eCLIP IP/SMInput $\log_2$(fold enrichment) = 2 and $-\log-_{10}$(padj) = 3 denoted by dashed red line. Clusters colored as in Figure 1. miRs validated by northern blot in Figure 5 are highlighted.
(C) Volcano plots of mature miR $\log_2$ (fold change) versus $-\log_{10}$(padj) as determined by small RNA-seq in knockdown versus control cells. Fold change is knockdown/control. Analysis performed using DESeq2 with padj calculated using the Benjamini-Hochberg correction. Clusters colored as in Figure 1. padj = 0.05 denoted by dashed red line. miRs validated by northern blot in Figure 5 are highlighted and volcano plots for HepG2 as well as other candidates are plotted in Figure S2.
(D) Representative UCSC genome browser tracks of normalized read density (RPM) of eCLIP data and small RNA-seq data for BUD13, DDX3X, ILF3, and LARP4 in K562 cells. Tracks are colored as in Figure 2 and show the RBP-bound loci validated by northern blot in Figure 5. Representative small RNA-seq tracks selected from 2 knockdown and 4 control replicates.

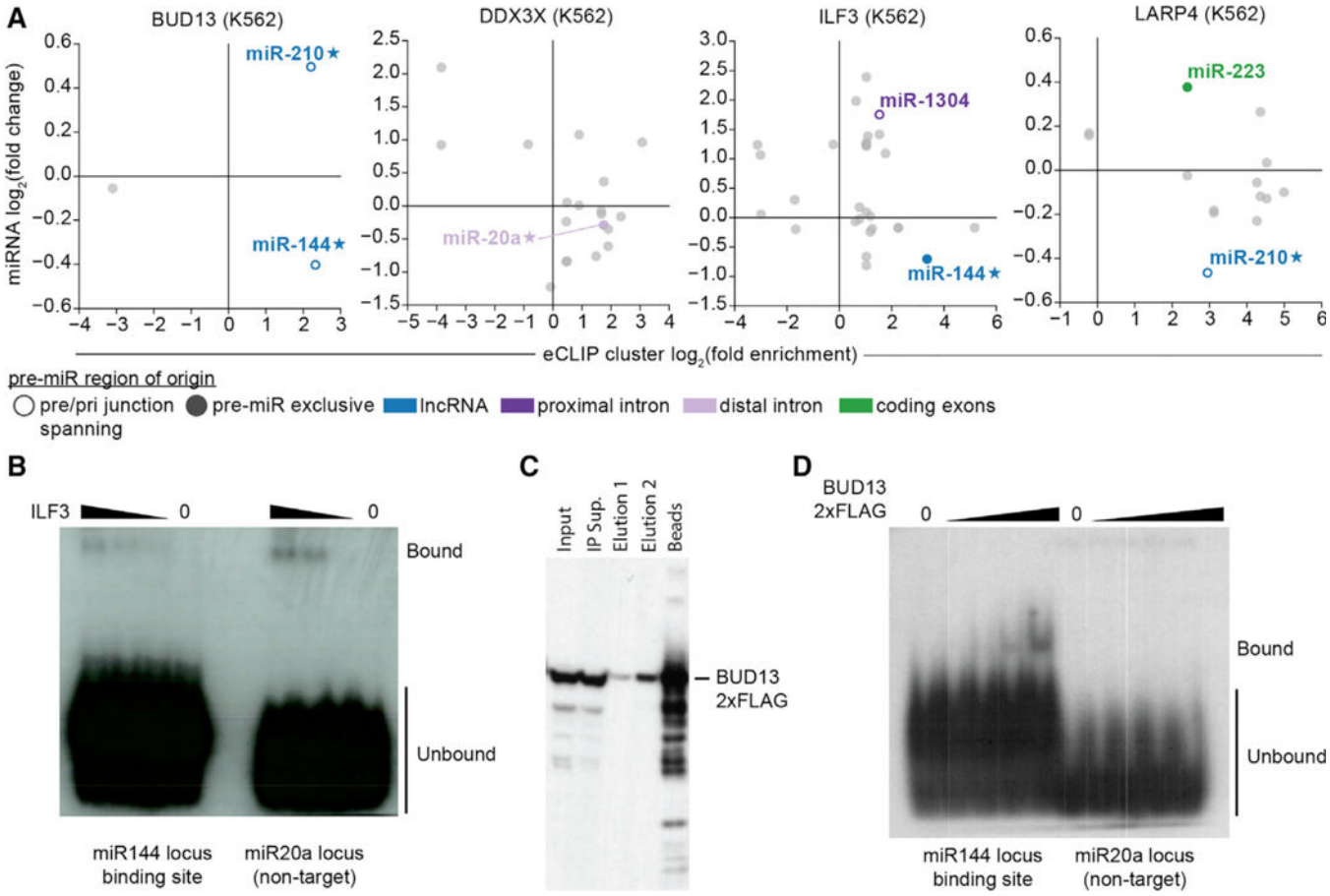


Figure 4. ILF3 and BUD13 Directly Interact with miR Loci Targets

(A) Integration of eCLIP and small RNA-seq data. eCLIP clusters at miR loci were further filtered for those intersecting miRs detected by small RNA-seq. For these clusters, eCLIP cluster IP/smInput $\log_2$(fold enrichment) was plotted versus small RNA-seq miR $\log_2$(fold change). miRs with significant eCLIP clusters (padj < 0.05) that also changed significantly upon RBP knockdown (padj < 0.05; Benjamini-Hochberg correction) are highlighted and colored as in Figure 1. miR loci validated by northern blot are further denoted by a star (★).

(B) Gel shift of recombinant ILF3 and ^{32}P-γ-ATP labeled miR-144 locus target and miR20a non-target.

(C) Purification of 2xFLAG-tagged BUD13. Elution 2 was used for the gel shift in (D).

(D) Gel shift of purified 2xFLAG-tagged BUD13 and ^{32}P-γ-ATP labeled miR-144 locus target and miR-20a locus non-target.

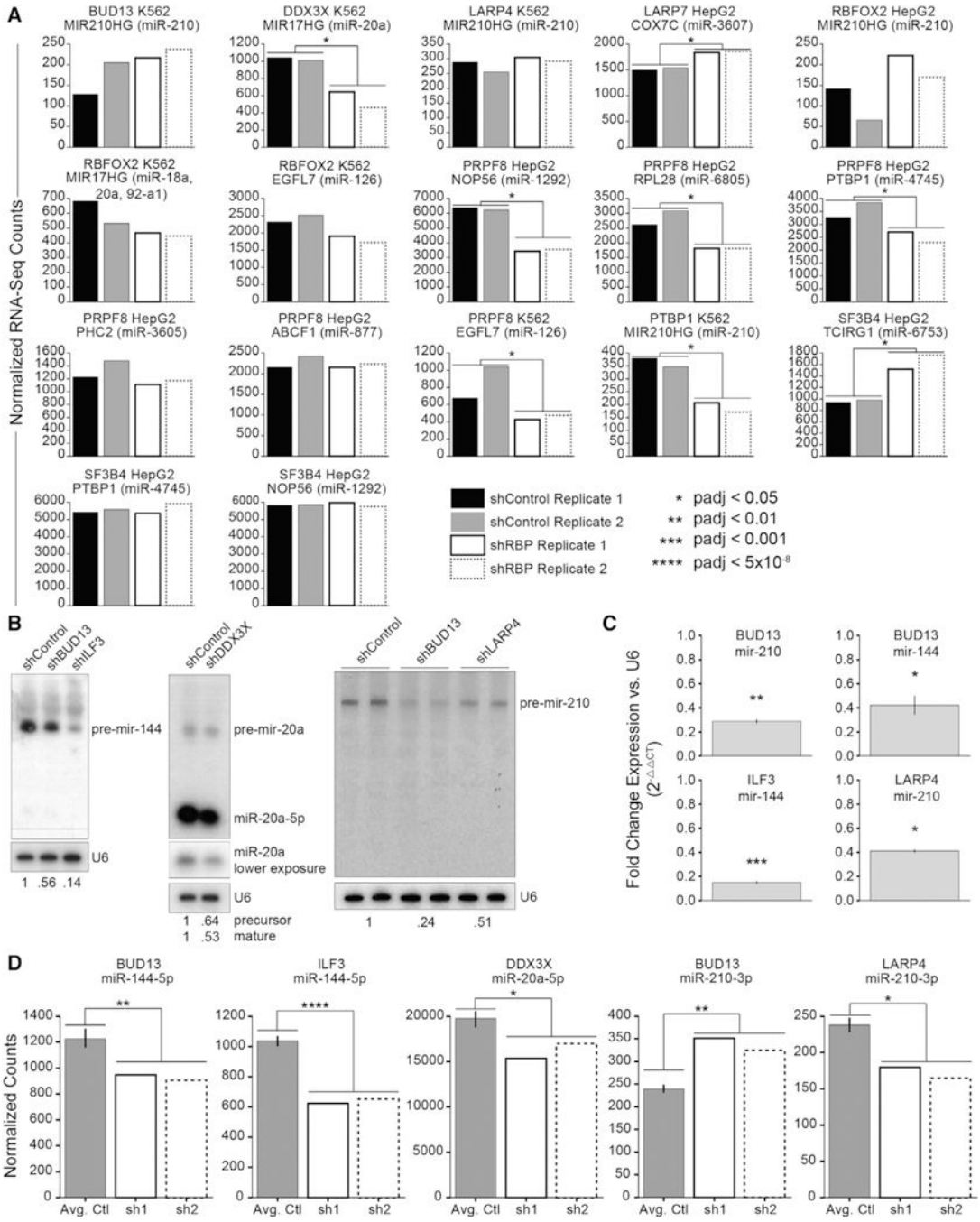


Figure 5. Identification of the Step of miR Biogenesis Regulation

(A)Bar plots ofdepth-normalized read counts as determined by RNA-seq in knockdown versus control cells. Highlighted arethe miR host genes (ifannotated) of the bound and regulated miR loci annotated in Figures 5A and S5.

(B) Northern blot analysis of bound and regulated miR loci upon RBP knockdown. Small RNAs (<200 bp) were isolated from control (shControl) and lentiviral shRNARBP knockdown K562 cellsand probed for pre- and mature miRs using 32-P γ-ATP-labeled

probes againstthe mature miRsequence. Probes against U6 were used as a loading control, and changes versus U6 and normalized to shControl are indicated below.
(C) Fold change in precursor expression determined by qPCR versus U6 control. Error bars represent 1 SD; p values are calculated by t test of biological duplicates.
(D) Bar plots of depth-normalized read counts as determined by small RNA-seq of shRBP/shControl mature guide miR of targets analyzed in (B). Error bars represent 1 SD of 4 biological replicates for shControl. Benjamini-Hochberg adjusted p values were calculated with DESeq2.