Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs

Citation: Leung, Anthony K L et al. "Genome-wide Identification of Ago2 Binding Sites from Mouse Embryonic Stem Cells with and Without Mature microRNAs." Nature Structural & Molecular Biology 18.2 (2011): 237–244.

As Published: http://dx.doi.org/10.1038/nsmb.1991

Publisher: Nature Publishing Group

Persistent URL: http://hdl.handle.net/1721.1/74622

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Published in final edited form as:
Nat Struct Mol Biol. 2011 February ; 18(2): 237–244. doi:10.1038/nsmb.1991.

Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs

Anthony K L Leung[1,5], **Amanda G Young**[1,2,4,5], **Arjun Bhutkar**[1], **Grace X Zheng**[1,3,4], **Andrew D Bosson**[1,2], **Cydney B Nielsen**[2,4], and **Phillip A Sharp**[1,2]

[1]The Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[3]Computational and Systems Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

Abstract

MicroRNAs (miRNAs) are 19-22nt non-coding RNAs that post-transcriptionally regulate mRNA targets. To identify endogenous miRNA binding sites, we performed photo-crosslinking immunoprecipitation using Ago2 antibodies, followed by deep-sequencing of RNAs (CLIP-seq) in mouse embryonic stem cells (mESCs). We also performed CLIP-seq in *Dicer*$^{-/-}$ mESCs that lack mature miRNAs, allowing us to define whether the association of Ago2 with the identified sites was miRNA-dependent. A significantly enriched motif, GCACUU, was identified only in wild-type mESCs in 3′ untranslated and coding regions. This motif matches the seed of a miRNA family that constitutes ~68% of the mESC miRNA population. Unexpectedly, a G-rich motif was enriched in sequences crosslinked to Ago2 in the presence and absence of miRNAs. Expression analysis and reporter assays confirmed that the seed-related motif confers miRNA-directed regulation on host mRNAs and that the G-rich motif can modulate this regulation.

miRNAs are key regulators of gene expression in fundamental processes including cell proliferation, cell death, cell differentiation and cellular responses to the environment[1-3]. These short non-coding RNAs guide a ribonucleoprotein complex, containing a member of the conserved Argonaute (Ago) protein family, to sites predominantly in the 3′UTRs of their target mRNAs, resulting in the destabilization of the message and/or inhibition of translation[4,5]. Biochemical and computational studies have shown that base-pairing between the "seed" (2nd -7th nucleotide) at the 5′ end of the miRNA and mRNA target is

Correspondence should be addressed to P.A.S. sharppa@mit.edu.
[4]Present address: Salk Institute for Biological Studies, La Jolla, California, USA (A.G.Y.); Howard Hughes Medical Institute and Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California, USA (G.X.Z.); Michael Smith Genome Sciences Centre, Vancouver, British Columbia, Canada (C.B.N.).
[5]These authors contributed equally to this work.

Accession codes. Microarray and short RNA sequence files have been deposited at the Gene Expression Omnibus database repository under accession number GSE25310. BED files for clusters in all libraries are available for download from http://rowley.mit.edu/pubs/Ago2_CLIP/Ago2_CLIP.html

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

Author Contributions: A.K.L. and A.G.Y. designed and performed the experiments; A.K.L., A.G.Y. and P.A.S. wrote the paper; A.D.B. performed experiments; A.B., C.B.N, and G.X.Z. performed the bioinformatics analyses. All authors reviewed and approved the manuscript.

Competing Interest Statement: The authors declare no competing financial interests.

important for this regulation in animals[6–10]. Comparative genomic analysis for miRNA seed sites in 3′UTRs suggests that miRNAs regulate ~60% of all mammalian mRNAs[11]. Moreover, both comparative genomic analysis and emerging data from a handful of genes suggest that miRNAs also target coding sequences[8,12,13], but the prevalence of this interaction is unclear. Therefore, recent efforts[14–16], including the study presented here, have aimed at identifying *bona fide* miRNA binding sites on a genome-wide scale in samples from whole mouse brain and whole-animal nematodes. However, one challenge of these studies is to deconvolute the miRNA-target relationships in the mixed cell types from these samples[15,16].

Results

CLIP-seq identified miRNA-dependent and -independent sites crosslinked to Ago2

In this study, we dissect the miRNA-target relationship in a homogeneous cell population – mouse embryonic stem cells (mESCs) – with defined miRNA characteristics[17–19]. RNA tags photo-crosslinked to Ago2 in these cells were isolated by immunoprecipitation and subjected to deep-sequencing (CLIP-seq)[15,16,20,21]. Importantly, no RNA species were detectable by autoradiography in Ago2 immunoprecipitates without crosslinking, suggesting that cloned RNA tags require crosslinking and thus are in direct association with Ago2 (Fig. 1a **and** Supplementary Fig. 1a). In addition, we performed a parallel analysis in derivative mESCs that lacks *Dicer* and hence mature miRNAs[22]. Unexpectedly, we identified specific RNAs crosslinked to Ago2 in *Dicer*$^{-/-}$ cells (Fig. 1a), indicating that Ago2 can associate with RNAs in a miRNA-independent manner[23,24].

Approximately 24.5M sequenced RNA tags from 3 wild-type mESC libraries representing two biological replicates (WT1A, WT1B, WT2) and 10.6M tags from two *Dice*r$^{-/-}$ libraries (KO1, KO2) were processed and mapped to the mouse genome (Supplementary Methods and Supplementary Fig. 1b-d). Across all libraries, 79% reads uniquely matched to the genome, 21% mapped to non-unique locations and 0.05% could not be aligned.

miRNAs crosslinked to Ago2 in mESCs were identified by screening reads with unique and repeat matches to the genome against non-coding RNA databases (Supplementary Fig. 1e; Supplementary Methods). Mature miRNAs are significantly enriched, as expected[17,18], in Ago2-crosslinked samples from wild-type cells compared with *Dicer*$^{-/-}$ cells. The *miR-290~295* cluster, *miR-467* family, and *miR-302~367* cluster (most members share the AAGUGC seed) represents the largest fraction (~68%) of the Ago2-crosslinked mature miRNA population[17–19,25] (Fig. 1b and Supplementary Fig. 1e), and the Ago2-CLIP and whole cell miRNA populations were positively correlated (Fig. 1b). While WT2 library had more reads mapping to ncRNA and repetitive regions than WT1 libraries, the distribution of crosslinked miRNAs is similar between the libraries (Supplementary Fig. 1e). The specificity of CLIP-seq method is shown by the absence of Ago2 crosslinking to the highly abundant rRNAs (~0.2%) and tRNAs (~0.2%).

For each library, the remaining tags that mapped uniquely to 3′UTRs were subjected to a data processing pipeline that consists of four filtering steps (Fig. 1c and Supplementary Table 1): First, identical reads were collapsed as a single read to eliminate potential PCR bias, and overlapping reads were then clustered (*Clustering filter*, Fig. 1c). 25nt flanking regions were added to the clusters in case an RNase cleavage separated where the miRNA bound and Ago2 crosslinked. Clusters were further considered only if they were significantly enriched over background levels (*Normalization filter*, Fig. 1c; Supplementary Methods). Third, to select for a reproducible signal, only the clusters that overlapped with at least one other cluster from a *Normalized* biological replicate library were considered (*Multi-Library filter*, Fig. 1c). Fourth, the remaining WT clusters that had overlaps with clusters from either *Normalized Dicer*$^{-/-}$ library (Knockout/KO) were removed (*Knockout filter*, Fig. 1c). Finally, after removing duplicates from technical replicates WT1A and 1B, 430 clusters in the combined WT libraries (244 in WT1[A+B] and 186 in WT2), of average length 81nt, passed all four filters. Various sets of clusters from different filtering steps were then subjected to motif enrichment analysis using two independent approaches.

First, significantly enriched motifs were identified in 3′UTR-mapped clusters from WT and KO sets independently (Fig. 1d, Supplementary Table 2 for all motif analyses). The motif discovery tool MEME26 was used to search for significantly enriched motifs of variable lengths over background (Supplementary Table 2a; Supplementary Methods). We found significant enrichment for G-rich motifs in clusters from both WT and KO libraries, suggesting that Ago2 may be associated with G-rich sequences independently of miRNAs. The G-rich motifs identified independently in WT and KO libraries have an average Pearson correlation of ~0.80, suggesting a high degree of similarity between the motifs (Supplementary Table 3a). Therefore, we defined a consensus G-rich motif by performing MEME analysis on Ago2-crosslinked clusters that overlapped between WT and KO libraries. This motif was highly statistically enriched (E-value=2.9×10^{-386}) and present in 87% of the common clusters between *Normalized* WT and KO libraries. Examination of individual libraries showed that this consensus G-rich motif was present at approximately equal frequency in sequences crosslinked to Ago2 from wild-type and *Dicer*$^{-/-}$ mESCs, again suggesting that its association with Ago2 is miRNA independent (Supplementary Table 3b).

One of the two significantly enriched miRNA-dependent motifs identified in 3′UTRs was GCACU[UG] (79 instances from 430 WT clusters). GCACUU (48/79 GCACU[UG] motifs) is complementary to the seed AAGUGC of several highly expressed miRNA families in mESCs. The only other statistically enriched miRNA-dependent motif in the selected clusters was CCAGCC (51 instances). However, unlike GCACUU, this motif is not complementary to any miRNA sequence with appreciable expression in mESCs.

To independently investigate enrichment of motifs from clusters within each individual CLIP library (Fig. 1e, Supplementary Fig. 2a, Supplementary Table 2b for top 20 motifs), we used an enumerative approach that guarantees global optimality by statistical overrepresentation and avoids the problem of being trapped at local optima inherent in most general motif-finding algorithms27. Briefly, we measured the statistical significance of the occurrence of all possible *n*-mer sequences within each library compared to their occurrence in sequences drawn randomly given a background distribution. This independent analysis confirmed the significant enrichment of G-rich hexamers (≥3Gs, red dots, Fig. 1e, Supplementary Fig. 2a) out of all possible hexamers in WT and KO libraries, as demonstrated by their high p-values and z-scores at a false discovery rate (FDR) < 0.5% (Supplementary Methods for derivations). The three hexamers encompassed in the consensus G-rich motif are amongst the top 7 significantly enriched hexamers in WT and KO libraries (black circle, Fig. 1e). Enrichment was observed exclusively in WT libraries for several non-G rich hexamers (blue dots, Fig. 1e), including GCACUU (black dot) and CCAGCC (light-blue dot). 29 non-G rich hexamers matched to miRNA seeds, but these miRNAs are associated with Ago2 at a median frequency of 0.003% ($p < 0.05$; Supplementary Fig. 1f, Supplementary Table 2b, c). Several other miRNA seed-matching hexamers occurred with high frequency, but were not observed significantly more than expected by chance and thus were not further considered (Supplementary Table 2c). The miRNA-dependent motif GCACUU is one of the top significantly enriched non-G rich hexamers in all WT libraries, including WT2 that had a lower proportion of 3′UTR-mapping clusters. The enrichment of GCACUU is particularly apparent after applying the *Knockout filter*, where common clusters between WT and KO libraries, many of which contain G-rich hexamers (red dots), are removed from the WT set. In effect, the *Knockout filter* reveals GCACUU as the most significantly enriched non-G rich hexamer in mESCs expressing miRNAs (black dot, left-most panel vs. right-most panel Fig. 1e). We also observed enrichment of 7mers and 8mers containing GCACUU that match the extended seed region[6,8–10] of the AAGUGC-seed family (Supplementary Table 2d).

Sequences mapping to coding sequences (CDS) were also subjected to the same data processing pipeline, resulting in a set of 197 clusters (106 in WT1[A+B], 91 in WT2). As in the case of 3′UTR clusters, G-rich motifs were highly significantly enriched by MEME analysis in CDS clusters from both WT and KO libraries (constituting ~25% and ~30% of clusters, respectively; data not shown). Moreover, GCACUU hexamer was observed in the CDS clusters from wild-type libraries (22 instances in 197 clusters; Supplementary Table 2a), but not KO libraries. Similarly, in the enumerative analysis of individual libraries, G-rich hexamers were highly enriched in both WT and KO libraries and GCACUU was enriched only in WT libraries (Supplementary Fig. 2b, Supplementary Table 2e, 3c). Unlike 3′UTR-mapped clusters, both MEME and enumerative analyses indicated no enrichment for CCAGCC in the CDS-mapped clusters.

Ago2-CLIP genes exhibit a miRNA-dependent gene expression signature

mRNAs targeted by miRNAs are often destabilized, resulting in a lower abundance of targeted transcripts in wild-type cells as compared to *Dicer*$^{-/-}$ cells[4,28,29]. We used mRNA expression of two sets of Ago2-CLIP 3′UTR GCACUU transcripts in wild-type and *Dicer*$^{-/-}$ mESCs to determine if their stability is miRNA-regulated. These two sets included the high-confidence “Overlap” set, comprised of 43 genes that passed the *Normalization* and *Multi-Library filters*, and a more inclusive “All” set, comprised of 201 genes that passed the *Normalization filter* for any WT library. The log$_2$ fold expression change (LFC) between wild-type and *Dicer*$^{-/-}$ mESCs was compared to the LFC of a control set of genes that lacked the GCACUU-motif. The Ago2-CLIP 3′UTR GCACUU-motif genes from both “Overlap” and “All” sets showed significantly more downregulation in wild-type mESCs relative to *Dicer*$^{-/-}$ mESCs, as compared to the control gene set (Fig. 2a-d; Supplementary Table 4 for statistics and Supplementary Table 5 for gene lists). These results independently support that these mRNAs physically bound to Ago2 are *in vivo* miRNA targets in mESCs.

Given that miRNA-dependent changes in mRNA expression have previously been shown for high-confidence predicted targets based on computational analysis of conservation and context around the seed site (TargetScan 5.1[8,11,30]), the properties of these predicted targets of the AAGUGC seed-related family were compared with the mRNAs identified by Ago2-CLIP. Comparison of expression levels in wild-type mESCs of Ago2-CLIP genes and predicted targets showed that the Ago2-CLIP 3′UTR GCACUU-motif genes tend to be more highly expressed (Supplementary Fig. 3). This is not surprising, as biochemical enrichment protocols tend to more effectively sample highly expressed genes.

To further compare properties of the predicted targets and CLIP-identified mRNAs other than expression level, two expression-matched and 3′UTR length-matched sets of predicted targets for the AAGUGC-seed family were generated. The first set, “All predicted targets”, contains 799 TargetScan GCACUU-containing predicted targets. Compared with this predicted set, both Ago2-CLIP “Overlap” and “All” gene sets have significantly greater miRNA-dependent changes in expression (Fig. 2a and 2c), suggesting that the CLIP-identified mRNAs possess features in addition to the miRNA seed match requirement.

To assess the importance of conservation and context around the seed site, we created two gene sets (“Conserved predicted targets”) containing the highest-confidence bioinformatically predicted targets, which are first ranked by branch length (i.e.

conservation), then by context score (scored combinatorially by its site-type, 3′pairing, local AU content, and position within the 3′UTR[8,11]) (Fig. 2b, d). These two sets were comparable to the corresponding Ago2-CLIP "Overlap" and "All" sets in terms of gene number, expression level, and 3′UTR length. No statistically significant difference in miRNA-dependent gene expression change was observed between the Ago2-CLIP "Overlap" gene set and the "All" gene set and their corresponding "Conserved predicted targets" sets (Fig. 2a, c). Yet, the CLIP-identified GCACUU sites from the "Overlap" and "All" sets are generally less conserved and surrounded by a relatively less favorable sequence context than the "Conserved predicted targets" (Fig. 2b, d; Supplementary Table 4d-f). Taken together, our results suggest that the "All" set and the smaller "Overlap" gene sets represent high confidence sets of miRNA-regulated mRNAs and that there are factors, besides conservation and context around the GCACUU seed motif, that govern which sites miRNAs target and/or are bound by Ago2 in mESCs.

We also sought to determine whether the GCACUU-motifs identified in CDS were associated with a miRNA-dependent gene expression signature. To this end, expression of *Normalization filtered* Ago2-CLIP CDS GCACUU-motif genes from all WT libraries (excluding those with GCACUU in the 3′UTR) was compared to a set of controls that lacks GCACUU in the CDS. The 80 Ago2-CLIP CDS GCACUU-motif genes showed miRNA-dependent downregulation in mRNA expression compared with the control set (Fig. 2e). Interestingly, other expression-matched CDS GCACUU-motif genes ("Predicted set", Fig. 2e) showed a similar profile as the Ago2-CLIP identified set and a significant downregulation compared with the control. This indicates that the presence of the GCACUU motif in CDS, as in the case of 3′UTR, is associated with a miRNA-dependent gene expression signature[8,12].

The expression profile difference between wild-type and *Dicer*$^{-/-}$ mESCs was further examined for genes with the G-rich motif, whose association with Ago2 appears to be miRNA-independent, and with the CCAGCC motif, whose association with Ago2 might be miRNA-dependent. Neither the G-rich motif nor the CCAGCC motif is complementary to any miRNA sequence with appreciable expression in mESCs. We compared those CCAGCC-containing genes that passed the *Normalization filter* (excluding those containing GCACUU) with expression-matched sets of all mouse genes that do not contain CCAGCC (control) in the 3′UTR (Fig. 2f). Surprisingly, we observed a significant downregulation of gene expression for the CCAGCC-containing genes in wild-type mESCs relative to *Dicer*$^{-/-}$ mESCs. This expression difference appears to be specific to those Ago2-CLIP CCAGCC-containing mRNAs as other mRNAs containing CCAGCC did not have a similar change ("Predicted set" in Supplementary Fig. 4). For the G-rich motif from Fig. 1d, we compared the expression change for Ago2-CLIP genes that contain matches to this motif, but lack GCACUU in their 3′UTRs, and passed the *Normalization* and *Multi-library filters* (Fig. 2f) with a set of 3′UTRs randomly chosen from the mouse genome that was matched for expression level, dinucleotide CG composition, and 3′UTR length. As is the case for GCACUU- and CCAGCC-containing genes, a significant increase in gene expression was observed upon deletion of *Dicer* for these G-motif containing genes identified by Ago2-CLIP. Such observed change could be due to the presence of other miRNA seed matches in the 3′UTRs of Ago2-CLIP G-rich motif genes. However, excluding those G-rich motif genes harboring seed matches to abundant mESC miRNAs did not affect the aggregate gene expression change of the G-rich motif gene set (Supplementary Fig. 4).

Interestingly, the degrees of change in mRNA expression observed for G-motif or CCAGCC containing genes were not significantly different from those observed for the Ago2-CLIP GCACUU-motif genes (Fig. 2f) and their expression-matched predicted GCACUU set (cf. Fig. 2a). Correlated with this, previous data suggests that the effect of miRNAs can be mimicked by miRNA-independent tethering of Argonaute proteins to reporter mRNAs[31,32]. Thus, the observed miRNA-dependent expression changes for Ago2-CLIP genes could be due to the close proximity between Ago2 and the crosslinked mRNA targets.

GCACUU-containing clusters are sufficient to confer miRNA-mediated repression

Next, we sought to determine whether the GCACUU-containing regions that crosslinked to Ago2 are sufficient to confer miRNA-dependent repression on luciferase reporter transgenes in the presence of endogenous levels of the corresponding miRNAs. Since only four genes[33-36] have been validated as GCACUU seed match targets in mESCs, it was difficult to evaluate our dataset with the existing literature. Instead, the ~80nt Ago2-CLIP cluster sequence was inserted into the 3′UTR of luciferase and the expression of this construct was compared to an equivalent construct with the GCACUU motif mutated to CCUCAU. The ratio of wild-type to mutant construct expression was evaluated in 3 cellular states: (1) wild-type (endogenous miRNA levels), (2) *Dicer*$^{-/-}$ mESCs (no mature miRNAs) and (3) *Dicer*$^{-/-}$ mESCs transfected with a miR-295 mimic, as illustrated in Fig. 3a. In each cellular state, the relative repression was calculated by normalizing to the ratio in *Dicer*$^{-/-}$ cells. We found that 8 out of 8 Ago2-CLIP 3′UTR GCACUU-motifs showed significant miRNA-dependent repression in wild-type cells but not in *Dicer*$^{-/-}$ cells (Fig. 3b). However, the repression in *Dicer*$^{-/-}$ cells was restored by addition of a miR-295 mimic, suggesting that a member from this mESC-specific miRNA cluster (with AAGUGC seed) is sufficient to provide the specificity for such regulation. Additionally, Ago2-CLIP-identified binding sites were present in three genes that have previously been shown to be regulated by the AAGUGC-related miRNA family (*E2f1*[37], *Pten*[38], *Cdkn1a*[33]). These data show that the Ago2-CLIP 3′UTR-bearing GCACUU-motif sites are indeed endogenous targets for direct regulation by miRNAs in mESCs and the short fragment of ~80nt containing such sites is sufficient to confer mESC-specific miRNA-mediated repression through *miR-290~295*.

To determine whether the Ago2-CLIP CDS GCACUU-motif sites are sufficient for miRNA regulation, we inserted the CDS cluster sequence (~80nt), or a seed mutant equivalent, in the 3′UTR of luciferase. 7 out of 8 clusters containing CDS GCACUU-motifs conferred downregulation on the luciferase reporter (Fig. 3c), suggesting that these sequences are recognized by the endogenous miRNA machinery even in the heterologous context of the 3′UTR.

miRNA regulation can be modulated by G-rich motifs associated with Ago2

The enrichment of the G-rich motif in Ago2-CLIP sequences from both wild-type and *Dicer*$^{-/-}$ mESCs (Fig. 1d) suggests that it is likely a miRNA-independent binding site for Ago2. This binding preference has not previously been described for Ago2, so we used an independent method to confirm the miRNA-independent association of Ago2 with the set of the G-rich motif containing mRNAs. We transfected *Dicer*$^{-/-}$ mESCs with an HA-tagged Ago2 construct, immunoprecipitated Ago2 by anti-HA antibodies, isolated the bound mRNA, and hybridized it to Affymetrix microarrays. We also performed microarray analysis on total RNA from *Dicer*$^{-/-}$ mESCs. The enrichment of mRNAs in the Ago2-IP was determined by comparing expression values between Ago2-IP and total RNA. We then determined whether the set of genes enriched in the Ago2-IP from *Dicer*$^{-/-}$ mESCs significantly overlapped with the sets of Ago2-CLIP G-motif containing genes. We found that 1.6-2.1 fold more genes overlapped between the Ago2-IP set and the Ago2-CLIP G-motif set than expected by chance (Supplementary Table 6a). These data support the observation that the G-motif containing genes identified by Ago2-CLIP are likely bound to Ago2 or its associated protein complex in a miRNA-independent manner in mESCs.

We previously determined that the CLIP-identified 3′UTR GCACUU mRNAs have a miRNA-dependent expression change comparable to the high-confidence predicted targets, despite being less conserved and in a less favorable sequence context (Fig. 2a-d). To explore whether the G-motif is a feature in these sequences that can contribute to miRNA-dependent regulation, we focused on a new miRNA target, *Txnip*, identified in this study. We validated this target by demonstrating both its endogenous mRNA and protein levels are regulated by *Dicer* in mESCs (Fig. 4a), similarly to previously validated *miR-290~295* target, *Cdkn1a*33. One of the Ago2-CLIP clusters identified in *Txnip* was amongst the most repressed in our 3′UTR luciferase assay (Fig. 3b). This provides a good range of sensitivity to test whether neighboring G-rich motifs affect the miRNA-dependent activity of GCACUU seed sites (Fig. 4b).

The relationship between this G-rich motif and seed-motif was investigated as in Fig. 3 using the following luciferase constructs: (1) the WT cluster, (2) with GCACUU seed motif mutated to CCUCAU, (3) a mutant G-rich motif where all Gs are mutated to Cs (not to alter AU content8,10), (4) or both motifs being mutated. For Txnip “A” cluster, repression was the strongest with wild-type GCACUU seed motif and G-motif (Fig. 4c). Interestingly, the repression was relieved by 50% when the G-motif was mutated (Fig. 4c). However, in the absence of GCACUU, the presence of G motif alone did not confer repression (Fig. 4c). We extended this analysis by investigating the contributions of G-rich motifs to miRNA-dependent repression in another cluster from *Txnip*, Txnip “B”, (Fig. 4d) and a cluster from *Ei24* (Fig. 4e). These clusters each have multiple G-rich motifs; mutation of each G-rich motif individually has varying impact on repression by the GCACUU seed site, with deletion of all G-rich motifs having the greatest effect on repression (Fig. 4f). Similar loss of miRNA-dependent repression was also observed when the G-rich motif of Txnip “A” cluster and Ei24 cluster is deleted (Supplementary Fig. 5a). Taken together, these data suggest that the G-rich motif is important for the full activity of the miRNA seed site, but does not contribute activity in the absence of the miRNA seed site.

Given that the CLIP-identified G-rich motif modulated the miRNA-mediated repression in the three clusters examined, we further investigated the general features of this motif, including its composition, conservation, and location within the Ago2-associated sequence. We searched for enrichment of shorter motifs in the 3′UTR clusters and found that the original 8mer G motif is comprised of enriched G-rich 4mers and 5mers (Supplementary Fig. 5b). Next, we analyzed the conservation of the 8mer G-motif. We determined an average conservation score for all G motifs based on the phastCons conservation39 of each nucleotide within the motifs and compared with a background set of 8mers (Supplementary Methods). We found that G-motifs are generally more conserved than random sequences ($p<1E-06$) (Fig. 4g). We also analyzed nucleotide positional conservation of an alignment of G-motifs with 10 nt flanks at either end. The level of conservation decreased immediately after the 3′end of the motif whereas the higher level of conservation persists in the 10 nt 5′ of the motif (Supplementary Fig. 5c). Interestingly, further MEME analyses suggest that the 8mer G-motif is likely embedded in an extended G-motif (Supplementary Fig. 5b). The excess conservation observed for G-motifs was true for all 3′UTR clusters, including those lacking the GCACUU motif (Supplementary Fig. 5d). Thus, the excess conservation of the G-motif is not a bystander effect from being near this particular miRNA seed match, but rather the G-motif has attributes of a functional regulatory element40.

Another common feature of this G-motif is that it tends to be present in the 5′ half of the sequence that is crosslinked to Ago2 (Supplementary Fig. 5e-f). In contrast, there is no positional bias for the GCACUU motif (Supplementary Fig. 5f). In cases where both motifs are present in the Ago2-CLIP sequences, there are no biases as to whether the G-motif is 5′ or 3′ of the GCACUU motif (data not shown). The activity of the examined G-motifs is independent of its location relative to GCACUU (c.f. Fig. 4b, d-e), suggesting that the vicinity, rather than the directionality, is important for modulating miRNA repression.

Discussion

Photo-crosslinking followed by Ago2 immunoprecipitation, Ago2-CLIP, was used to identify miRNA binding sites in mESCs. We found significantly enriched motifs in 3′UTRs and CDS that correspond to miRNA seed matches, representing 201 and 103 potential mESC miRNA targets in 3′UTRs and CDS, respectively. In regards to the latter point, this study is in agreement with other studies that the presence of miRNA binding sites in CDS is more widespread than has been previously considered and nearly as prevalent as in 3′UTRs[14–16]. Here we provided gene expression data suggesting that these CDS sites regulate mRNA stability much like 3′UTR sites. Moreover, these sites can be recognized by miRNAs at endogenous expression levels and confer repression in a heterologous 3′UTR[8,13,41,42].

Two other Ago-CLIP studies have identified potential miRNA targets in mammalian cells and tissue. Our study differs from those by analyzing mRNAs associated with endogenous Ago2 in a mostly homogenous cell population of mESCs, whereas Chi *et al.*[15] performed CLIP on brain extracts using endogenous Ago antibodies and Hafner *et al.*[14] performed CLIP in 293 cells using HA-tagged Ago1-4 and crosslinking by a photoactivatable nucleotide. Independently of the variations in the CLIP technique and data analysis, these studies, as well as our studies, identified similar numbers of targets for each miRNA seed family (several hundred), which is comparable to the number of moderately conserved targets predicted for each miRNA seed family by TargetScan[11] (Supplementary Notes for cross-comparison with other CLIP datasets).

There are several previously published reports of miRNA-regulated mRNAs in mESCs that we could compare to the Ago2-CLIP 3′UTR GCACUU-containing genes[35, 43]. Only miR-294 (member of AAGUGC seed family) regulated mRNAs described by Melton *et al.*[43] showed significant overlap with the Ago2-CLIP 3′UTR GCACUU-containing mRNAs (Supplementary Table 6b for all comparisons).

Unlike other cell types, including those used in other Ago2-CLIP studies[14,44], mESCs appear to be dominated by a single miRNA seed family that is probably responsible for most of the miRNA regulation in this cell type. Essentially all of the GCACUU-motif containing CLIP 3′UTR clusters conferred miRNA-dependent regulation when tested in luciferase reporter assays in the presence of endogenous levels of miRNAs. This suggests that the stringency of our filtering criteria resulted in selection of a high confidence set of GCACUU-containing mRNAs that most likely are *bona fide* miRNA targets in mESCs. Previous studies have already shown that this miRNA family plays important roles in mESCs, including maintaining pluripotency, self-renewal and cell cycle control[33–35,43,45]. But, few targets have been identified and validated. This study identifying a few hundred new miRNA targets by Ago2-CLIP is a significant step in the exploration of this biology.

To understand the extent of miRNA-regulated pathways represented by the Ago2-CLIP 3′UTR GCACUU-motif genes ("All" set, 201 genes), we performed pathway enrichment analysis (Supplementary Methods) and compared this set with the top 201 "Conserved predicted targets" and all mRNAs expressed in mESCs that contain GCACUU hexamer in the 3′UTR ("All predicted targets", 2969 genes). 37 and 11 pathways were significantly enriched in the CLIP and "Conserved predicted targets" sets, respectively (Supplementary Figure 6a and Supplementary Table 7). The pathways significantly enriched in CLIP included "Early S-phase" (4 genes), a pathway in which *miR-290~295* has been previously implicated33, and "TGF-beta receptor signaling" (5 genes), a pathway where *miR-290~295* has not been implicated.

The genes identified by Ago2-CLIP in "TGF-beta receptor signaling" pathway (p-value 0.013) include two intracellular pathway inhibitors, the cytoplasm-localized *Smad7* and the nucleus-localized *Skil*, and an extracellular inhibitor, *Lefty1*46. Our reporter assay confirmed that these 3 genes are indeed targeted by miR-295 (Supplementary Fig. 6b). We extended this analysis to *Lefty2*, a gene that was not identified in the CLIP results, but is homologous to *Lefty1* and contains the GCACUU hexamer, and showed that it is also targeted by miR-295 (Supplementary Fig. 6b). Correlated with this, *miR-302* and *miR-430*, which are related in miRNA seed to *miR-290~295*, have been shown, respectively, in human ESCs47 and zebrafish embryos48 to regulate differentiation through targeting *Lefty* homologs. Here using a genome-wide approach, we found that the *miR-290~295* regulates not only the extracellular *Lefty* homologs, but also additional inhibitory nodes of the TGF-beta pathway localized in different cellular compartments (Supplementary Fig. 6c). This coordinate inhibition, as observed for other miRNAs49‑51, might confer robustness in this signaling network.

We unexpectedly identified a G-rich motif in most of the sequences associated with Ago2 regardless of the miRNA status in the cell. We believe this is a true biological association, rather than a technical artifact, based on the following observations. First, this motif is conserved above the general 3′UTR background even when matched for sequence content. Second, the genes containing this G-rich motif have significant overlap with the set of genes enriched in HA-tagged Ago2 immunoprecipitates from *Dicer*$^{-/-}$ mESCs. Third, we only observe G bias in genic sequences, and not miRNA or intergenic sequences crosslinked to Ago2. Lastly, the enrichment of G residues is not likely due to CLIP itself as there are no described G biases in the literature for any of the steps involved (Supplementary Notes for further discussion).

Yet it remains unclear whether crosslinking to this G-rich sequence is due to Ago2 itself or a binding partner of Ago2. Given that UV-crosslinking forms covalent bonds between protein and RNA that are within angstroms, a potential binding partner would have to be in close proximity to Ago2 and the mRNA target. Indeed, several proteins that co-immunoprecipitate with Ago2 have binding preference for G-rich sequences, including HNRNP-H and FMRP52‑54, although we only observed one Ago2-dependent RNAprotein complex close to the molecular weight of native Ago2 in the CLIP procedure. Alternatively, Ago2 itself could have a previously unidentified preference for binding G-rich sequences. In either case, when a G-rich sequence occurs near a miRNA binding site, it could give the Ago2/miRNA complexes a higher affinity for this region and thus lead to increased probability that the mRNA is targeted for degradation and/or inhibition of translation. In three cases examined, this G-rich motif modulates the level of miRNA-dependent regulation by the *miR-290~295*-related seed motif, but imparts no regulation by itself. Therefore, identification of this association indicates the value of Ago2-CLIP data from *Dicer*$^{-/-}$ mESCs as an invaluable background in delineating *bona fide* microRNA targets.

Methods

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank C. Burge, J. Wilusz, A. Ravi and members of the Sharp laboratory for critical comments, M. Lindstrom for illustration, T. Cybulski for technical help, A. Leshinsky and R. Cook for running the Solexa sequencing samples in the KI Biopolymer & Proteomics Core Facility, M. Luo and L. Smeester for microarray technical assistance in the MIT Department of Biology Biomicrocenter, and C. Whittaker for bioinformatics support in the Bioinformatics & Computing Core Facility at the Koch Institute. A.K.L was supported by a special fellowship from the Leukemia and Lymphoma Society. A.G.Y was partially supported by a David H. Koch graduate fellowship. This work was supported by United States Public Health Service grants R01-GM34277 and R01-CA133404 from the National Institutes of Health, P01-CA42063 from the National Cancer Institute to P.A.S and partially by Cancer Center Support (core) grant P30-CA14051 from the National Cancer Institute.

References

1. Ambros V. The functions of animal microRNAs. Nature. 2004; 431:350–5. [PubMed: 15372042]
2. Leung AK, Sharp PA. MicroRNA functions in stress responses. Mol Cell. 2010; 40:205–15. [PubMed: 20965416]
3. Stefani G, Slack FJ. Small non-coding RNAs in animal development. Nat Rev Mol Cell Biol. 2008; 9:219–30. [PubMed: 18270516]
4. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009; 136:215–33. [PubMed: 19167326]
5. Carthew RW, Sontheimer EJ. Origins and Mechanisms of miRNAs and siRNAs. Cell. 2009; 136:642–55. [PubMed: 19239886]
6. Brennecke J, Stark A, Russell RB, Cohen SM. Principles of microRNA-target recognition. PLoS Biol. 2005; 3:e85. [PubMed: 15723116]
7. Doench JG, Sharp PA. Specificity of microRNA target selection in translational repression. Genes Dev. 2004; 18:504–11. [PubMed: 15014042]
8. Grimson A, et al. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell. 2007; 27:91–105. [PubMed: 17612493]
9. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005; 120:15–20. [PubMed: 15652477]
10. Nielsen CB, et al. Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. Rna. 2007; 13:1894–910. [PubMed: 17872505]
11. Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 2009; 19:92–105. [PubMed: 18955434]
12. Baek D, et al. The impact of microRNAs on protein output. Nature. 2008; 455:64–71. [PubMed: 18668037]
13. Tay Y, Zhang J, Thomson AM, Lim B, Rigoutsos I. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. Nature. 2008; 455:1124–8. [PubMed: 18806776]
14. Hafner M, et al. Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. Cell. 2010; 141:129–141. [PubMed: 20371350]
15. Chi SW, Zang JB, Mele A, Darnell RB. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature. 2009; 460:479–86. [PubMed: 19536157]
16. Zisoulis DG, et al. Comprehensive discovery of endogenous Argonaute binding sites in Caenorhabditis elegans. Nat Struct Mol Biol. 2010; 17:173–9. [PubMed: 20062054]
17. Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes Dev. 2008; 22:2773–85. [PubMed: 18923076]
18. Calabrese JM, Seila AC, Yeo GW, Sharp PA. RNA sequence analysis defines Dicer's role in mouse embryonic stem cells. Proc Natl Acad Sci U S A. 2007; 104:18097–102. [PubMed: 17989215]
19. Houbaviy HB, Murray MF, Sharp PA. Embryonic stem cell-specific MicroRNAs. Dev Cell. 2003; 5:351–8. [PubMed: 12919684]

20. Licatalosi DD, et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature. 2008; 456:464–9. [PubMed: 18978773]

21. Yeo GW, et al. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nat Struct Mol Biol. 2009; 16:130–7. [PubMed: 19136955]

22. Leung AK, Calabrese JM, Sharp PA. Quantitative analysis of Argonaute protein reveals microRNA-dependent localization to stress granules. Proc Natl Acad Sci U S A. 2006; 103:18125–30. [PubMed: 17116888]

23. Tan GS, et al. Expanded RNA-binding activities of mammalian Argonaute 2. Nucleic Acids Res. 2009; 37:7533–45. [PubMed: 19808937]

24. Yoda M, et al. ATP-dependent human RISC assembly pathways. Nat Struct Mol Biol. 2010; 17:17–23. [PubMed: 19966796]

25. Ciaudo C, et al. Highly dynamic and sex-specific expression of microRNAs during early ES cell differentiation. PLoS Genet. 2009; 5:e1000620. [PubMed: 19714213]

26. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol. 1994; 2:28–36. [PubMed: 7584402]

27. Sinha S, Tompa M. Discovery of novel transcription factor binding sites by statistical overrepresentation. Nucleic Acids Res. 2002; 30:5549–60. [PubMed: 12490723]

28. Behm-Ansmant I, Rehwinkel J, Izaurralde E. MicroRNAs silence gene expression by repressing protein expression and/or by promoting mRNA decay. Cold Spring Harb Symp Quant Biol. 2006; 71:523–30. [PubMed: 17381335]

29. Farh KK, et al. The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. Science. 2005; 310:1817–21. [PubMed: 16308420]

30. Lim LP, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature. 2005; 433:769–73. [PubMed: 15685193]

31. Djuranovic S, et al. Allosteric regulation of Argonaute proteins by miRNAs. Nat Struct Mol Biol. 17:144–50. [PubMed: 20062058]

32. Pillai RS, Artus CG, Filipowicz W. Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. Rna. 2004; 10:1518–25. [PubMed: 15337849]

33. Wang Y, et al. Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. Nat Genet. 2008; 40:1478–83. [PubMed: 18978791]

34. Benetti R, et al. A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases. Nat Struct Mol Biol. 2008; 15:268–79. [PubMed: 18311151]

35. Sinkkonen L, et al. MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. Nat Struct Mol Biol. 2008; 15:259–67. [PubMed: 18311153]

36. Foshay KM, Gallicano GI. miR-17 family miRNAs are expressed during early mammalian development and regulate stem cell differentiation. Dev Biol. 2009; 326:431–43. [PubMed: 19073166]

37. O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT. c-Myc-regulated microRNAs modulate E2F1 expression. Nature. 2005; 435:839–43. [PubMed: 15944709]

38. Xiao C, et al. Lymphoproliferative disease and autoimmunity in mice with increased miR-17-92 expression in lymphocytes. Nat Immunol. 2008; 9:405–14. [PubMed: 18327259]

39. Siepel A, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 2005; 15:1034–50. [PubMed: 16024819]

40. Xie X, et al. Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. Nature. 2005; 434:338–45. [PubMed: 15735639]

41. Kloosterman WP, Wienholds E, Ketting RF, Plasterk RH. Substrate requirements for let-7 function in the developing zebrafish embryo. Nucleic Acids Res. 2004; 32:6284–91. [PubMed: 15585662]

42. Gu S, Jin L, Zhang F, Sarnow P, Kay MA. Biological basis for restriction of microRNA targets to the 3′ untranslated region in mammalian mRNAs. Nat Struct Mol Biol. 2009; 16:144–50. [PubMed: 19182800]

43. Melton C, Judson RL, Blelloch R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. Nature. 2010; 463:621–6. [PubMed: 20054295]

44. Landgraf P, et al. A mammalian microRNA expression atlas based on small RNA library sequencing. Cell. 2007; 129:1401–14. [PubMed: 17604727]

45. Judson RL, Babiarz JE, Venere M, Blelloch R. Embryonic stem cell-specific microRNAs promote induced pluripotency. Nat Biotechnol. 2009; 27:459–61. [PubMed: 19363475]

46. Zovoilis A, Smorag L, Pantazi A, Engel W. Members of the miR-290 cluster modulate in vitro differentiation of mouse embryonic stem cells. Differentiation. 2009; 78:69–78. [PubMed: 19628328]

47. Rosa A, Spagnoli FM, Brivanlou AH. The miR-430/427/302 family controls mesendodermal fate specification via species-specific target selection. Dev Cell. 2009; 16:517–27. [PubMed: 19386261]

48. Choi WY, Giraldez AJ, Schier AF. Target protectors reveal dampening and balancing of Nodal agonist and antagonist by miR-430. Science. 2007; 318:271–4. [PubMed: 17761850]

49. Li X, Cassidy JJ, Reinke CA, Fischboeck S, Carthew RW. A microRNA imparts robustness against environmental fluctuation during development. Cell. 2009; 137:273–82. [PubMed: 19379693]

50. Marson A, et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell. 2008; 134:521–33. [PubMed: 18692474]

51. Tsang J, Zhu J, van Oudenaarden A. MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. Mol Cell. 2007; 26:753–67. [PubMed: 17560377]

52. Caudy AA, Myers M, Hannon GJ, Hammond SM. Fragile X-related protein and VIG associate with the RNA interference machinery. Genes Dev. 2002; 16:2491–6. [PubMed: 12368260]

53. Edbauer D, et al. Regulation of Synaptic Structure and Function by FMRP-Associated MicroRNAs miR-125b and miR-132. Neuron. 2010; 65:373–384. [PubMed: 20159450]

54. Hock J, et al. Proteomic and functional analysis of Argonaute-containing mRNA-protein complexes in human cells. EMBO Rep. 2007; 8:1052–60. [PubMed: 17932509]

55. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. Nat Genet. 2007; 39:1278–84. [PubMed: 17893677]

56. Blanchette M, et al. Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res. 2004; 14:708–15. [PubMed: 15060014]

57. Ule J, Jensen K, Mele A, Darnell RB. CLIP: a method for identifying protein-RNA interaction sites in living cells. Methods. 2005; 37:376–86. [PubMed: 16314267]

58. Hafner M, et al. Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. Methods. 2008; 44:3–12. [PubMed: 18158127]

59. Calabrese JM, Sharp PA. Characterization of the short RNAs bound by the P19 suppressor of RNA silencing in mouse embryonic stem cells. RNA. 2006; 12:2092–102. [PubMed: 17135486]

60. Bailey TL, Gribskov M. Combining evidence using p-values: application to sequence homology searches. Bioinformatics. 1998; 14:48–54. [PubMed: 9520501]

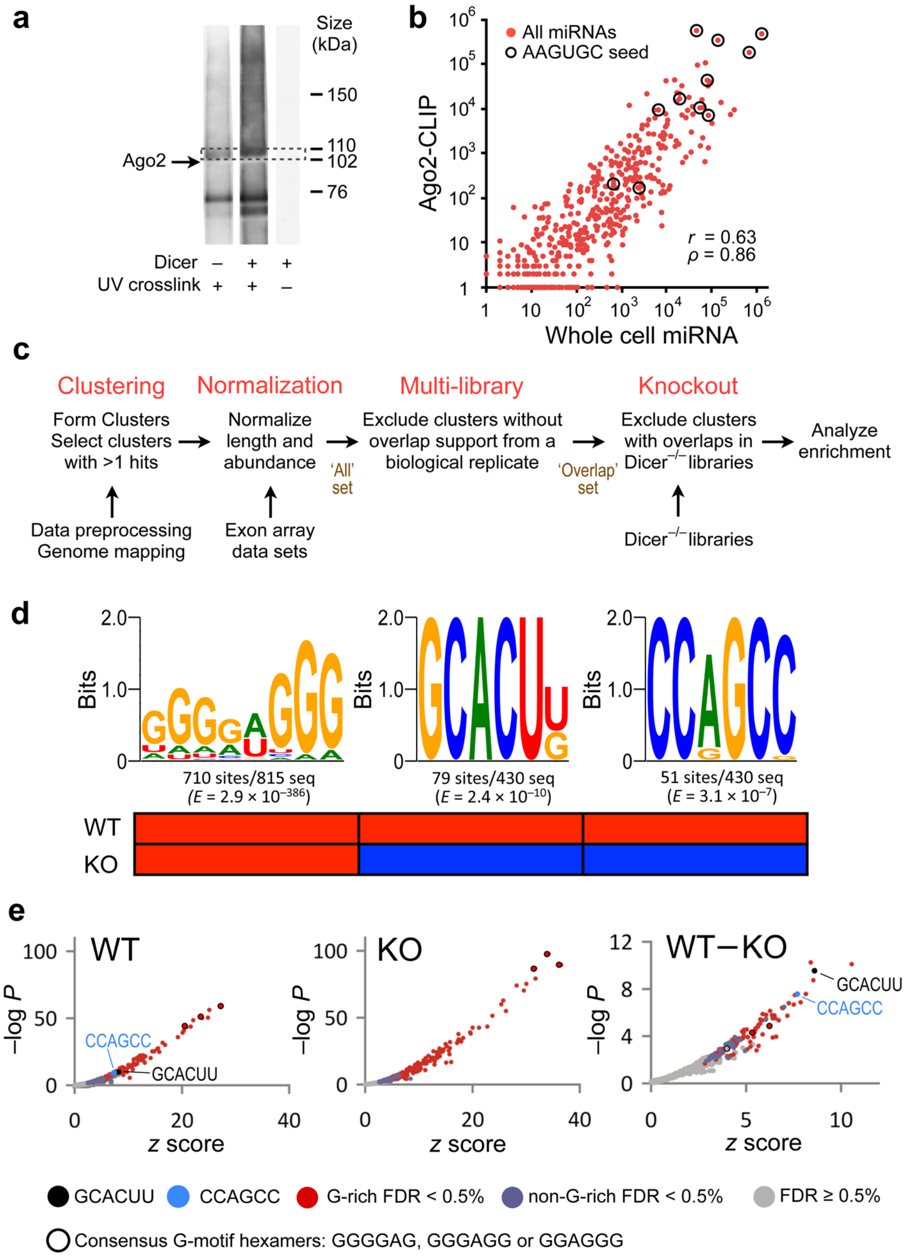

Figure 1. Identification of miRNA-dependent and –independent motifs associated with Ago2

(**a**) Autoradiograph of ^{32}P-labeled RNA from Ago2 complexes immunoprecipitated from *Dicer*$^{-/-}$ mESCs, WT mESCs and WT mESCs without UV crosslinking. Bracketed area indicates the region of the blot that was excised and analyzed; arrow indicates where endogenous Ago2 migrates. (**b**) Log plot of mature miRNA cloning frequency in the WT1A Ago2-CLIP library vs. whole cell miRNA from wild-type mESCs (Pearson coefficient r = 0.63; Spearman coefficient ρ = 0.86). (**c**) Data processing pipeline. After preprocessing for linker matching and stripping, and subsequently mapping to the mouse genome, Ago2-CLIP sequence reads were subjected to the filtering steps as described in the text (Supplementary Methods and Supplementary Table 1 for more details). WT reads were subjected to all 4 filters, KO reads were subjected to first 3 filters, *Clustering, Normalization, and Multi-Library (KO1,KO2 overlaps)*. (**d**) Sequence logos and statistics of the top three significantly enriched motifs derived from motif analysis of Ago2-CLIP 3′UTR-mapping clusters using the motif tool MEME (Supplementary Methods for details). The number of sites containing each motif out of the number of clusters examined is as indicated. In the heatmap, red indicates significant enrichment ($E < 1 \times 10^{-5}$ cutoff) and blue indicates not significantly enriched in either the WT or KO libraries. The G-rich motif (far left) was highly enriched in all libraries analyzed, and the representative consensus sequence was taken from MEME analysis (width 4-8) on *Normalized* KO1 clusters that had overlaps with clusters from any *Normalized* WT library. The GCACU[UG] motif (middle) and CC[AG]GCC (far right) were significantly enriched by MEME analysis (width 6) in only the 430 clusters that passed all 4 filters in the combined wild-type libraries and not in the Ago2-CLIP 3′UTR clusters from any *Normalized* KO library. There were 48 instances of GCACUU and 31 instances of GCACUG represented in the GCACU[UG] motif. (**e**) Hexamer enrichment analysis by statistical overrepresentation within individual libraries, using an enumerative approach (Supplementary Methods). From left to right: *Normalized* and *Multi-library* filtered Ago2-CLIP WT1A (WT), *Normalized* and *Multi-library* filtered Ago2-CLIP KO2 (KO), and *Normalized*, *Multi-library* and *Knockout* filtered Ago2-CLIP WT1A (WT-KO). Refer to Supplementary Fig. 2 for hexamer enrichment analyses in other libraries. Two measures of significance are plotted: The X axis shows the z-score, which is a measure of the number of standard deviations the observed frequency of a hexamer exceeds its expectation. The Y axis is the negative $\log_{10}$ p-value (two-sided Fisher's exact test) of the hexamer enrichment above background. All hexamers with a false discovery rate (FDR) ≥ 0.5% are in grey and not considered as significantly enriched. Significantly enriched hexamers (FDR < 0.5%) are classified into two types: G-rich (≥3Gs, red) and non-G-rich (blue). Those hexamers that match the consensus G-rich motif derived from MEME analysis are circled. Hexamers only significantly enriched in WT are GCACUU (black dots) and CCAGCC (light blue dot). Note the change in scale for the far right plot.

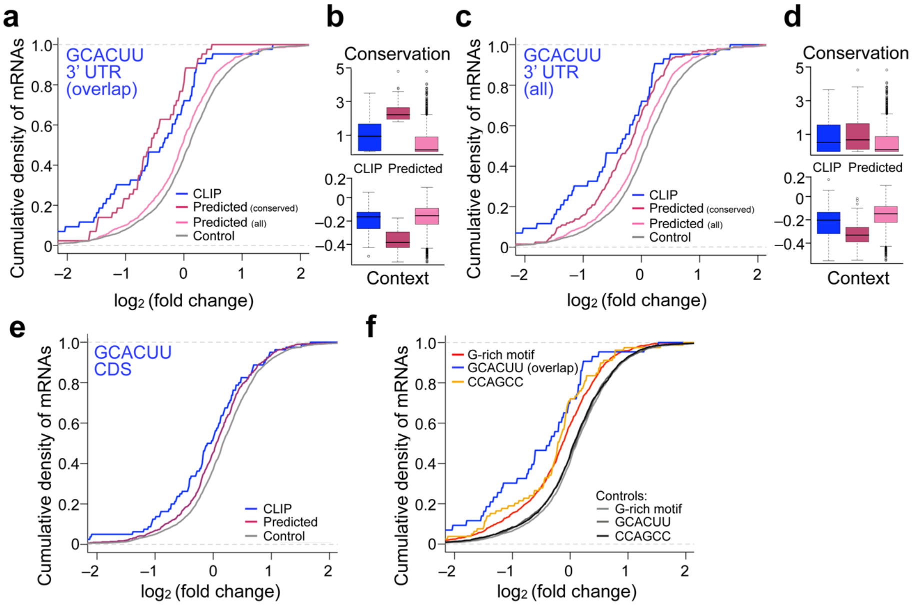

Figure 2. Ago2-CLIP genes exhibit a miRNA-dependent gene expression signature

(**a, c, e, f**) CDFs (cumulative density functions) of the $\log_2$ fold change (LFC) in mRNA expression between the wild-type (WT) and *Dicer*$^{-/-}$ (KO) mESCs. Fold change is ratio of expression level in WT cells to expression level in KO cells, such that a CDF with a more negative LFC compared to another CDF indicates that the distribution has lower mRNA expression in wild-type mESCs relative to *Dicer*$^{-/-}$ mESCs (i.e. population includes mRNAs that are more downregulated in presence of miRNAs). Refer to Supplementary Table 4 for detailed descriptions on statistics and Supplementary Table 5 for gene lists. (**b, d**) Box plots of branch lengths (top), i.e. "Conservation", and "Context" scores (bottom) computed by TargetScan 5.1. (**a-d**) Gene sets include: Ago2-CLIP 3′ UTR GCACUU-motif mRNAs (blue) that passed the *Normalization* and *Multi-Library filters* (**a-b,** "Overlap") or just the *Normalization filter* (**c-d,** "All") in any WT library; "Conserved predicted" expression-matched TargetScan GCACUU targets (dark pink), ranked by branch length (i.e. conservation), then by context score (**a**, Top-ranked 43 genes; **c**, Top-ranked 201 genes); "All predicted" expression-matched TargetScan GCACUU targets (light pink, 1469 genes); and "Control" mRNAs (grey). (**e**) Gene sets include: Ago2-CLIP CDS GCACUU-containing mRNAs that passed the *Normalization filter* in any WT library ("CLIP", blue), "Predicted" GCACUU-containing CDS mRNAs (pink) that were expression-matched to the "CLIP" set but lacked GCACUU in their 3′UTRs, and matched "control" mRNAs (grey). (**f**) Gene sets include: Ago2-CLIP 3′ UTR GCACUU-containing (blue) and G-motif-containing mRNAs (red) that passed the *Normalization* and *Multi-Library filters* in WT libraries, Ago2-CLIP 3′UTR CCAGCC-containing mRNAs (orange) that passed the *Normalization filter* in at least one WT library, and corresponding matched controls (grey). Any mRNAs with GCACUU in the 3′UTR or CDS were removed from the "G-motif" and "CCAGCC" sets. Error bar is SD.

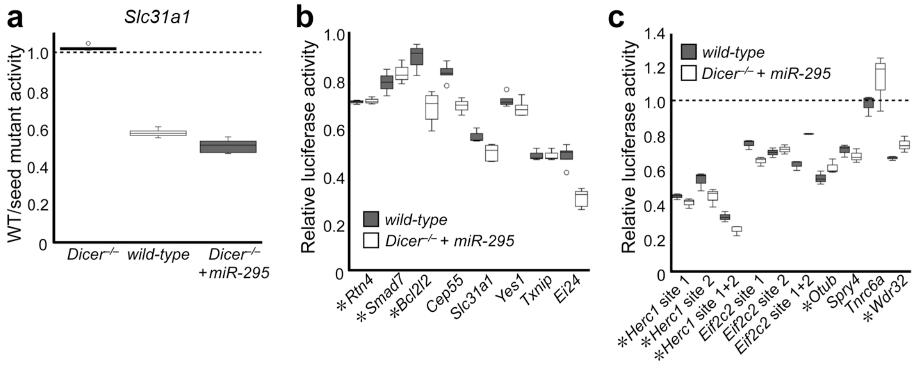

Figure 3. Ago2-CLIP identified GCACUU-motif containing cluster is sufficient to confer miR-295-mediated repression

Box plots of relative activity of *Renilla* luciferase transgenes bearing CLIP clusters plus 25nt flanking regions in the 3′UTR. All raw *Renilla* luciferase values are normalized first to values from Firefly luciferase transfection control. Error bars are S.D. (**a**) Exemplary plot of a luciferase reporter assay testing an Ago2-CLIP 3′UTR cluster, *Slc31a1* plus 25nt flanking regions at both ends. Shown are box plots for the ratio of the expression level of a luciferase reporter containing a wild-type *Slc31a1* CLIP cluster in its 3′UTR to an identical luciferase reporter where the GCACUU seed was mutated to CCUCAU. This ratio is defined as "WT/Seed Mutant Activity" and calculated in three different cellular states: *Dicer*$^{-/-}$ mESCs (black), wild-type mESCs (white), and *Dicer*$^{-/-}$ mESCs transfected with miR-295 mimic (grey). For panels **b-c**, "Relative Luciferase Activity" is defined as "WT/Seed mutant activity" in specific cellular state normalized to the corresponding value in *Dicer*$^{-/-}$ mESCs. Genes that passed both the *Normalization* and *Multi-library filters* (i.e. "Overlap" set) are marked with an asterisk. All remaining genes passed the *Normalization filter* in one WT library. All box plots are from at least 3 independent experiments. (**b**) Relative luciferase activity of CLIP-identified 3′UTR clusters plus 25nt flanking regions in wild-type mESCs (grey) or *Dicer*$^{-/-}$ mESCs transfected with miR-295 mimic (white). Relative luciferase activity is significantly less than one in all genes examined (p<0.05, two-tailed paired t-test). (**c**) Relative luciferase activity of CLIP-identified CDS clusters plus 25nt flanking regions in wild-type mESCs (grey) or *Dicer*$^{-/-}$ mESCs transfected with miR-295 mimic (white). *Herc1* and *Eif2c2* each have two GCACUU motifs. Relative luciferase activity is significantly less than one for all genes examined except *Tnrc6a* (p<0.05, two-tailed paired t-test).

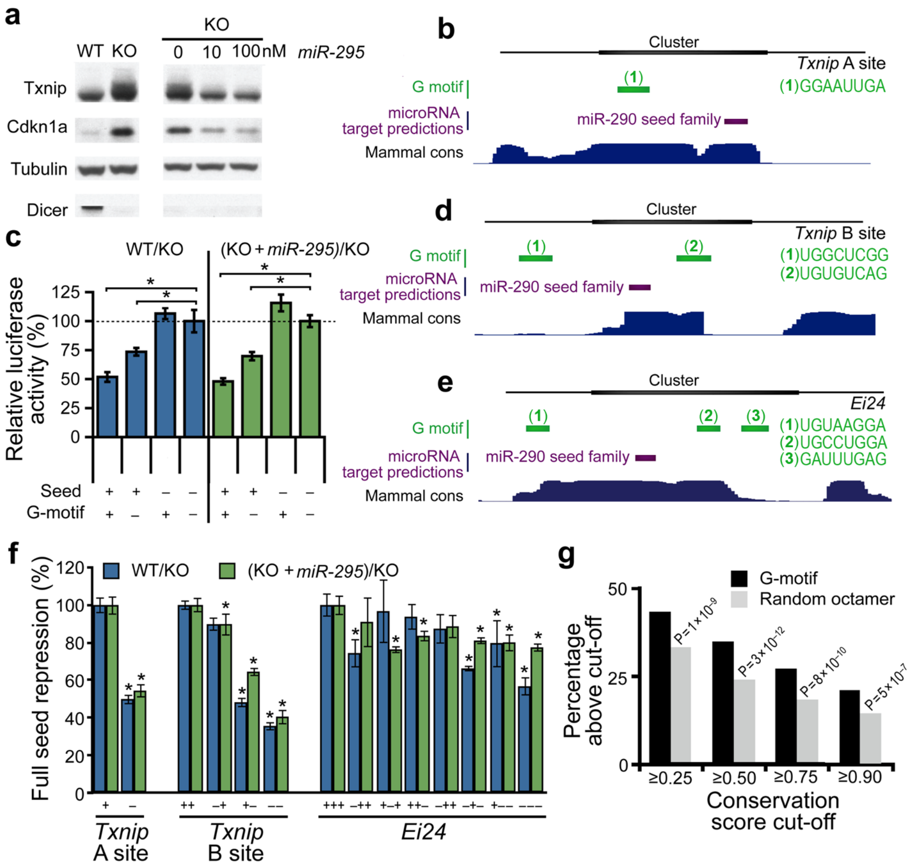

Figure 4. G-rich motif modulates miRNA-mediated repression

(**a**) Western blot of GCACUU-containing targets, Txnip and Cdkn1a, in wild-type mESCs (WT), *Dicer*$^{-/-}$ mESCs (KO) (left panel), and *Dicer*$^{-/-}$ mESCs transfected with miR-295 mimic at different concentrations (nM), 0 (control), 10, 100 (right panel). Tubulin loading control and *Dicer* genotype control are also shown. (**b, d, e**) Genomic schematic of Ago2-CLIP 3′UTR mapping clusters. The following features are depicted: the Ago2-CLIP cluster sequence (thick black bar) with flanking regions (thin black bar), 8mer G-rich motif(s) in the cluster (green bar and sequence in green to the right; Supplementary Methods for details), predicted *miR-290~295* cluster GCACUU binding site (purple) using multiple algorithms including TargetScan 5.111 and PITA55, UCSC genome browser 30-Way Mammalian Multiz alignment and conservation39,56 (blue, heights indicate degree of conservation at aligned position). Clusters depicted are *Txnip* "A" (**b**) and "B" (**d**) clusters from *Txnip* 3′UTR, and *Ei24* cluster (**e**) from *Ei24* 3′UTR. All G-rich motifs shown match a consensus G-rich motif p<0.05, Supplementary Methods for details. (**c**) Mutation of the Txnip "A" G-rich motif reduces GCACUU seed-mediated repression. Luciferase reporter constructs were created as in Fig. 3, using an Ago2-CLIP cluster, plus 25nt flanking regions, from the *Txnip* 3′UTR. Along with the GCACUU seed mutant construct, a G-motif mutant construct with all the Gs in the G-motif changed to Cs as well as a construct with both the seed and G motif mutated were created. The presence or absence of an intact GCACUU seed match or G-motif in the construct is indicated by + (WT) or - (mutated). Relative luciferase activity was calculated similarly as in Fig. 3 as the expression level of each indicated construct relative to the seed(-) G-motif(-) double mutant construct in either wild-type mESCs (WT) (left, blue) or *Dicer*$^{-/-}$ mESCs transfected with miR-295 mimic (right, green), normalized to the same ratio in *Dicer*$^{-/-}$ mESCs (KO). Values shown are averages of three independent experiments and p-values < 0.05 were deemed significant and marked as asterisk. (**f**) Mutation of G-rich motifs in Ago2-CLIP clusters reduces repression conferred by three independent miR-290 family seed sites. Similar data as in **c** but plotting the amount of seed-mediated repression on the luciferase reporter in the presence (+) or absence (-) of the G-motif in either wild-type mESCs (WT) (blue) or *Dicer*$^{-/-}$ mESCs transfected with miR-295 mimic (green), normalized to the same ratio in *Dicer*$^{-/-}$ mESCs (KO). 100% Full seed repression was set as one minus the ratio of the expression levels of the WT construct to the seed mutant construct (G motif +). The % full seed repression in the absence of the G-motif was calculated as one minus the ratio of the expression levels of the G-motif mutant construct to seed mutant, G-motif mutant construct (G-motif -) as a percent of 100% full seed repression (as described above). For *Txnip* “B” and *Ei24* clusters, the G-motifs “+” or “-” correspond to same motifs shown left to right in the schematics (**d**) and (**e**). Values shown are averages of three independent experiments and p-values < 0.05 were deemed significant and marked as asterisk. (**g**) Average phastCons score was determined for each G-motif (match score > 0, data for *Normalized* Ago2-WT1A shown in black) along with a constrained background set of random 8mers drawn from annotated mouse 3′UTRs (grey) (Supplementary Methods for explanation of background derivation). On the X axis are bins for phastCons score cutoffs, ≥0.25, ≥0.50, ≥0.75 and ≥0.90. On the Y axis is the percentage of total G-motifs or random 8mers that are in each phastCons score cutoff bin. Ago2-CLIP G-motifs are on average more conserved than random 8mers (p-values as indicated, Fisher's exact test). Error bar is SD.