Downloaded from rnajournal.cshlp.org on July 24, 2018 - Published by Cold Spring Harbor Laboratory Press

Assessing the functional association of intronic miRNAs with their host genes

AVITAL STEIMAN-SHIMONY, ORR SHTRIKMAN, and HANAH MARGALIT

Department of Microbiology and Molecular Genetics, IMRIC, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem 9112102, Israel

ABSTRACT

In human, nearly half of the known microRNAs (miRNAs) are encoded within the introns of protein-coding genes. The embedment of these miRNA genes within the sequences of protein-coding genes alludes to a possible functional relationship between intronic miRNAs and their hosting genes. Several studies, using predicted targets, suggested that intronic miRNAs influence their hosts' function either antagonistically or synergistically. New experimental data of miRNA expression patterns and targets enable exploring this putative association by relying on actual data rather than on predictions. Here, our analysis based on currently available experimental data implies that the potential functional association between intronic miRNAs and their hosting genes is limited. For host-miRNA examples where functional associations were detected, it was manifested by either autoregulation, common targets of the miRNA and hosting gene, or through the targeting of transcripts participating in pathways in which the host gene is involved. This low prevalence of functional association is consistent with our observation that many intronic miRNAs have independent transcription start sites and are not coexpressed with the hosting gene. Yet, the intronic miRNAs that do show functional association with their hosts were found to be more evolutionarily conserved compared to other intronic miRNAs. This might suggest a selective pressure to maintain this architecture when it has a functional consequence.

Keywords: miRNA; intron; post-transcriptional regulation

INTRODUCTION

MicroRNAs (miRNAs) are small RNA molecules, ~22 nucleotides (nt) in length, which play important roles in post-transcriptional regulation of gene expression. MiRNAs have been discovered in most metazoan organisms as well as in some viruses, and most of them were shown to function as negative regulators of gene expression. They carry out their regulatory function while assembled with Argonaute into miRNA induced silencing complexes (miRISCs), guiding the RISCs to target mRNAs with complementary sequences. Downregulation of gene expression by miRNAs is achieved by translational repression and mRNA degradation, initiated by deadenylation, followed by decapping and 5′–3′ exonucleolytic decay (Jonas and Izaurralde 2015). Multitudes of data exist detailing the interactions and effects of miRNAs in numerous cellular processes (Inui et al. 2010), including organism development (Asli et al. 2008) and the immune system (Lee et al. 2014). In addition, impairment of miRNA function was shown to be associated with various diseases (Li et al. 2014).

MiRNAs are mainly transcribed by RNA polymerase II as large RNA precursors called primary miRNAs (pri-miRNAs), which undergo several steps of processing in the nucleus and cytoplasm before they reach their mature functional form. The pri-miRNA, which consists of a stem of 33–35 base pairs (bp), a terminal loop, and single-stranded RNA segments at both the 5′ and 3′ sides, is cropped by Drosha, an RNase III enzyme leaving the smaller (~65 nt) hairpin-shaped precursor miRNA (pre-miRNA). The pre-miRNA is exported into the cytoplasm where it is cleaved by Dicer to its mature form, which is then loaded onto Argonaute to form the miRISC (Ha and Kim 2014). Roughly half of the miRNAs are encoded in intragenic regions, mainly in the introns of protein-coding and noncoding genes, as well as within exons of noncoding genes (Rodriguez et al. 2004; Hinske et al. 2010). Clusters of several miRNAs, encoded within a single genomic region, have also been identified in both intergenic and intronic regions (Altuvia et al. 2005).

MiRNAs residing in introns present an interesting genomic architecture not common in eukaryotes, where two or more different molecules are encoded within the same gene. This organization is reminiscent of bacterial operons, where, often, operon genes are involved in the same cellular process. It is furthermore reminiscent of a recently-discovered functional module in prokaryotes, where a protein-coding gene hosts a regulatory small RNA (sRNA) in its 3′UTR,

Corresponding author: hanahm@ekmd.huji.ac.il

Article is online at http://www.rnajournal.org/cgi/doi/10.1261/rna.064386.117. Freely available online through the *RNA* Open Access option.

© 2018 Steiman-Shimony et al. This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

Downloaded from rnajournal.cshlp.org on July 24, 2018 - Published by Cold Spring Harbor Laboratory Press

and both may function in the same cellular process (Chao and Vogel 2016; Melamed et al. 2016). Such organization defines a multilayered functional module encoded in a single gene, where the encoded protein can be involved in transcriptional regulation or protein–protein interaction while the hosted sRNA functions as a post-transcriptional regulator. If the production of the two molecules derived from a single gene is coupled, this architecture guarantees tight coordination between the two regulation layers. However, in case they are transcribed independently, such coordination is not guaranteed. Indeed, sRNAs produced from 3′ UTRs of protein-coding genes were classified into two groups: Type I, which are transcribed independently of their host mRNA and seem to hold an independent function; and type II, which are processed out of their host transcript and whose function seems to support the function of their host mRNA (Miyakoshi et al. 2015; Melamed et al. 2016). In eukaryotes, small nucleolar RNAs (snoRNAs) provide an example of type II organization, where many of them are encoded in the introns of genes encoding ribosomal proteins, and their transcription is dependent on the transcription of the host gene (Boivin et al. 2017). As snoRNAs are involved in modifications of ribosomal RNAs, their embedment in ribosomal genes suggests a genomic organization underlying modules that coordinate the expression of protein components of the ribosome and the expression of noncoding RNAs regulating the biogenesis of RNA components of the ribosome. Since at least half of the miRNAs are encoded within protein-coding transcripts, it is intriguing to investigate whether their embedment within protein-coding genes might have functional consequences, and to explore the extent and types of possible functional associations in host-miRNA modules. Potentially, miRNAs can affect their hosts' function at multiple junctures. Most directly, an intronic miRNA can post-transcriptionally autoregulate its host by targeting the host gene's transcript. Another form of functional association between the host and the intronic miRNA can be through the regulation of common targets, where the regulatory effects of the miRNA and host gene can be either synergistic or antagonistic. Finally, miRNAs may be associated with their host function by targeting genes involved in pathways in which the host gene is involved.

The role of intronic miRNAs in regulating their host mRNA, as well as their involvement in the same pathways as their hosts, has been previously explored to some extent using predicted targets of miRNAs (Hinske et al. 2010; Lutter et al. 2010; Gao et al. 2012). Recent high-throughput advances in miRNA target identification, along with experimental data of transcription start sites (TSS) and of regulatory interactions the hosting genes are involved in, open the way to study the functional associations between protein-coding mRNAs and their hosted miRNAs using experimental data. Transcriptome-wide experimentally validated targets of miRNAs were first reported by using the high-throughput CLASH method (Helwak et al. 2013), and subsequently by similar methods, iPAR-CLIP (Grosswendt et al. 2014) and CLEAR-CLIP (Moore et al. 2015). These methods involve crosslinking of Argonaute and bound RNAs, ligation of the RNAs, sequencing, and identification of chimeric fragments representing miRNA-target pairs. The iPAR-CLIP study also looked for RNA chimeras in other data sets, which might be generated by an endogenous ligase, and support miRNA-target pairs (Grosswendt et al. 2014). The data generated by CLASH, as well as other similar methods, provide purely experimentally determined networks of direct interactions between miRNAs and their targets. In contrast, other high-throughput methods either identify Arognaute-bound mRNA transcripts and miRNAs, but require computational prediction to identify the miRNA-target pairs (Chi et al. 2009; Hafner et al. 2012), or find targets of a specific miRNA (Tan et al. 2014; Pickl et al. 2016). In addition, a recent annotation of cell-specific TSSs of miRNAs based on H3K4me3 and DNase I hypersensitivity sites enables the assessment of the cotranscription of the hosting gene and its intronic miRNA based on experimental data (Hua et al. 2016). In the current study, we set out to uncover the extent of possible involvement of intronic miRNAs in their hosts' function, using experimentally validated targets (Helwak et al. 2013; Grosswendt et al. 2014; Moore et al. 2015), experimental expression data of miRNA and hosting genes (Landgraf et al. 2007; Uhlén et al. 2015; Panwar et al. 2017) and TSSs based on experimental data (Hua et al. 2016). We find that many of the intronic miRNAs are transcribed independently of their hosting genes and that functional coordination between protein-coding genes and their hosted miRNAs seems to be less widespread than anticipated.

RESULTS

Characteristics of intronic miRNAs

We organized a data set of intronic miRNAs and characterized their sequence and expression properties. We determined intronic miRNAs using the human hairpin pre-miRNA coordinates from miRBase V21 (Kozomara and Griffiths-Jones 2014), and the coordinates of the introns of protein-coding genes with the same strand orientation as the miRNA according to the RefSeq human genome annotation (build hg38; Pruitt et al. 2005). A total of 866 pre-miRNAs were found to be encoded within the introns of 786 protein-coding genes. Among these, 158 pre-miRNAs were found to be clustered in 67 introns.

Evolutionary conservation

An evolutionary conservation score was computed for each intronic miRNA by recording the positional phastCons100ways conservation score (Siepel et al. 2005) and averaging it over all positions of the hairpin pre-miRNA sequence (Fig. 1). The distributions of conservation

Downloaded from rnajournal.cshlp.org on July 24, 2018 - Published by Cold Spring Harbor Laboratory Press

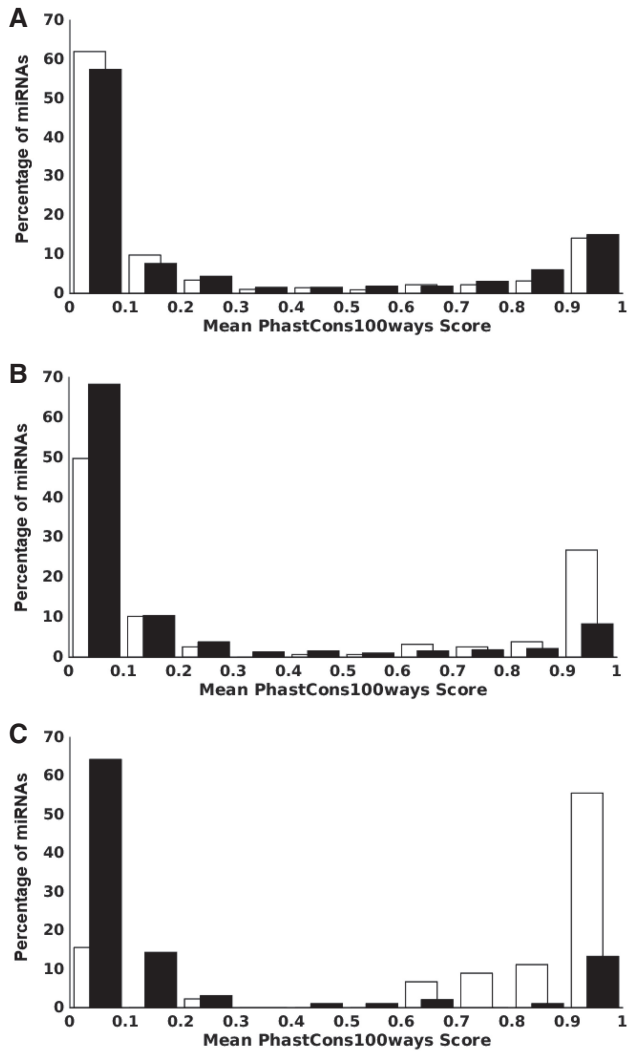

FIGURE 1. Evolutionary conservation of hairpin precursor miRNAs. The conservation score of each hairpin pre-miRNA is represented by the mean of phastCons100ways scores over its nucleotide sequence. (*A*) The distribution of the conservation scores of the intronic pre-miRNAs (white bars) is similar to the distribution of the nonintronic human pre-miRNAs (black bars). (*B*) Clustered intronic miRNAs (white bars) are statistically significantly more conserved than nonclustered intronic miRNAs (black bars; Wilcoxon rank sum test, $P \leq 3.2 \times 10^{-3}$). (*C*) Intronic miRNA that were found to be clustered in the same homologous transcripts in both mouse and human (white bars) are statistically significantly more conserved than other clustered intronic miRNAs (black bars; Wilcoxon rank sum test, $P \leq 1.8 \times 10^{-8}$).

scores of intronic and nonintronic hairpin pre-miRNAs were similar, both had a mean conservation score near 0, and ~20% were shown to have a score close to 1 (Fig. 1A). Among the intronic miRNAs, those encoded in clusters were more conserved than those that were not in clusters (Wilcoxon rank sum test, $P \leq 3.2 \times 10^{-3}$; Fig. 1B). Intronic miRNAs were also searched for in the mouse transcriptome. Five-hundred seventy-nine intronic pre-miRNAs were found in 500 protein-coding genes. One-hundred seventeen mouse host genes were also host genes in human, 91 of which have the same intronic miRNAs in human and in mouse. Of the 158 clustered intronic pre-miRNAs in human, 45 were also found to be clustered in the homologous gene in mouse, and were shown to be more highly conserved than the other clustered intronic miRNAs (Wilcoxon rank sum test $P \leq 1.8 \times 10^{-8}$; Fig. 1C). In *Caenorhabditis elegans*, 58 intronic pre-miRNAs were found in 55 protein-coding genes, with only one host-intronic miRNA homolog between human and *C. elegans*.

Transcription of intronic miRNAs and their host genes

Previous studies reported conflicting data as to the correspondence between the expression level of the hosting gene and the hosted miRNA. While some studies, using small data sets, found correlation between the expression patterns of the intronic miRNAs and their hosting genes (Baskerville and Bartel 2005; Hinske et al. 2010; França et al. 2016), other studies using more extensive data have demonstrated a low correlation between the two (Monteys et al. 2010; Marsico et al. 2013; Budach et al. 2016). To examine the association between the transcription of the hosting gene and hosted miRNA, we first surveyed TSS data of intronic miRNAs from a recent study by Hua et al. (2016). The study identified cell-specific TSSs for over 80% of miRNAs recorded in miRbase (Kozomara and Griffiths-Jones 2014) in 54 cell lines, based on H3K4me3 and DNase I hypersensitivity sites in addition to conservation and other sequence features. Furthermore, for intronic miRNAs, Hua and coworkers defined the TSSs as host dependent or independent relying on the distance of the intronic miRNA's candidate TSS from its host TSS. These data have enabled us to examine the span of cell lines in which the hosted miRNAs shared their TSSs with the hosting genes, and from this to infer the scale at which their expression is coordinated. To this end we computed the Jaccard index, which provides a measure ranging from 0 to 1 for the extent of shared TSSs of a host-miRNA pair over all cell lines where data were available for the respective pair (0: no shared TSS in any cell line; 1: shared TSS in all cell lines, Fig. 2A). Of the 638 intronic miRNAs in that data set, 332 intronic miRNAs were found to be transcribed solely by independent promoters in all cell lines checked, 242 were transcribed independently in some cell lines but also with the host transcript in other cell lines, and 64 intronic miRNAs were transcribed solely by the host genes' promoters in the given data. This distribution was not dependent on the number of cell lines for which data were available (Fig. 2B). In total, out of 17888 combinations of intronic miRNA and cell line for which data were available, in 3311 combinations the intronic miRNAs were found to share TSSs with their hosts. Our results suggest that in the majority of cases the intronic miRNAs are transcribed independently of their hosts.

To further explore the association between the transcription patterns of the hosting genes and hosted miRNAs we turned to analyze expression data. Initially, we used miRNA expression data from Landgraf et al. (2007) and host gene

Downloaded from rnajournal.cshlp.org on July 24, 2018 - Published by Cold Spring Harbor Laboratory Press

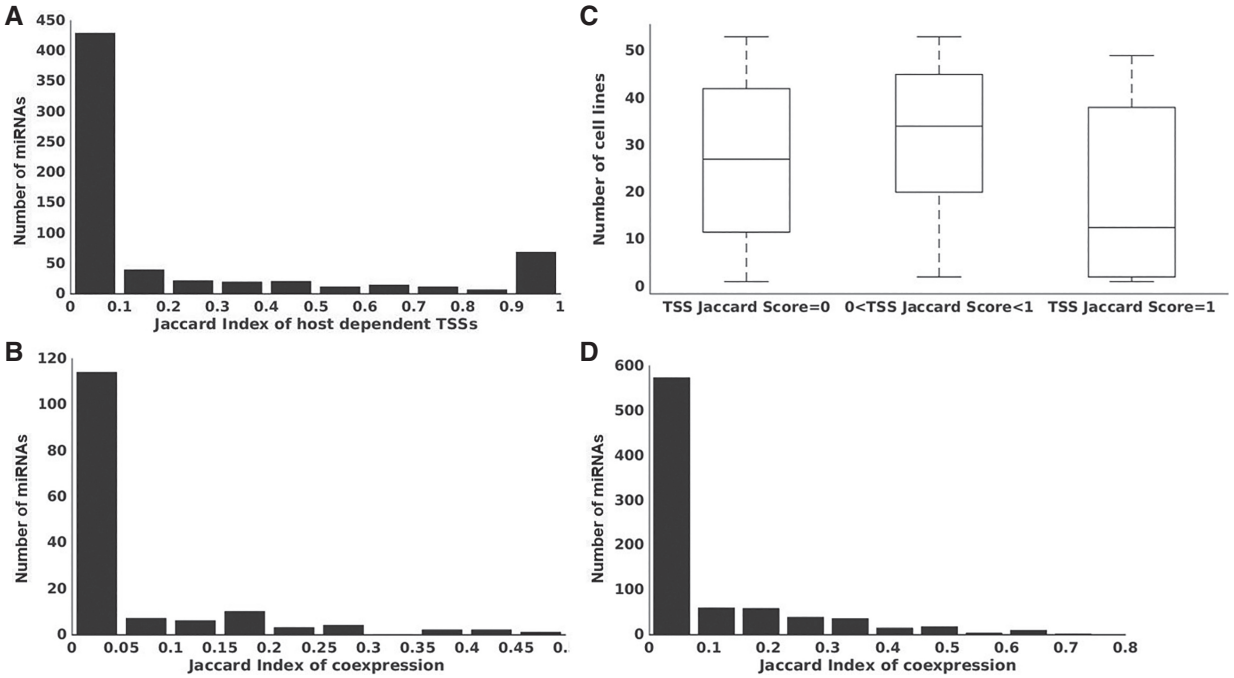

FIGURE 2. Coexpression and cotranscription of intronic miRNAs with their host genes. (*A*) Distribution of Jaccard index of host dependent TSSs for 638 intronic miRNAs over 54 cell lines, indicating little cotranscription of miRNA and host. (*B*) Distribution of the Jaccard index in *A* is independent of the number of cell lines in which data were available. Intronic miRNAs with TSS data were divided into three groups according to their Jaccard index for TSSs as follows: Jaccard index of 0 (*left*); Jaccard index above 0 and below 1 (*center*); Jaccard index of 1 (*right*). Box plots were made for each group showing the distribution of the number of cell lines in which each intronic miRNA had TSS data. It is clear that the low/high Jaccard indices are not due to small/large number of cell lines for which data were available. (*C*) Distribution of Jaccard index of coexpression of 149 intronic pre-miRNAs with their hosts over 28 cell lines based on pre-miRNA expression data from Landgraf et al. (2007), indicating little coexpression of miRNA and host. (*D*) Similarly, distribution of Jaccard index of coexpression of 829 intronic pre-miRNAs with their hosts over 14 cell lines based on miRNA expression data from miRmine (Panwar et al. 2017), also indicating little coexpression.

expression data based on RNA-seq data in the Human Protein Atlas (Uhlén et al. 2015). We had expression data for 149 intronic pre-miRNAs and their hosts in 28 different cell lines, which we could use in our analysis (Materials and Methods). A Jaccard index of coexpression was calculated for each host-miRNA pair (without taking into account the level of expression). Our analysis found that most intronic miRNAs for which expression data were available were not coexpressed with their hosts. Only 38 of the intronic miRNAs with expression data were found to be coexpressed with their host gene in at least one cell line, and even within those 38 host-miRNA pairs the average value of the Jaccard Index was low (0.190, Fig. 2C). We repeated this analysis with more recent miRNA expression data from the database miRmine, which contains data curated from publicly available human miRNA-seq studies (Panwar et al. 2017). The data include samples from 16 different human tissues, and 24 human cell lines, of which six human tissues and eight cell lines coincide with tissues and cell lines in the RNA-seq data of the Human Protein Atlas (Uhlén et al. 2015). Hence, the coexpression of 829 intronic miRNA-host pairs were compared in those 14 tissues and cell lines (Fig. 2D). Similar to the coexpression analysis based on Landgraf et al. (2007) data (Fig. 2C), here too the majority of intronic miRNAs (533/829) were not coexpressed with their hosts, with only 26 having a Jaccard coexpression score $\geq$0.5.

Intronic miRNAs known as mirtrons have been shown to bypass processing by Drosha through the utilization of the splicing machinery (Ruby et al. 2007; Westholm and Lai 2011). Using computational analysis of small RNA expression data, Wen et al. (2015) listed 478 loci in human that have the potential to generate splicing-derived miRNAs, 235 of which were included in our data set of 866 intronic pre-miRNAs. Based on their biogenesis, it is conceivable that the transcription and expression of these splicing-derived intronic miRNAs will be more coupled with their host transcripts compared to canonically derived intronic miRNAs. Hence, we repeated the analysis of all the studied characteristics separately for the mirtrons and compared them to other intronic miRNAs (Supplemental Fig. S1). While mirtrons were found to be less evolutionarily conserved than other intronic miRNAs, they did not differ from the other intronic miRNAs in the distribution of the Jaccard index for shared TSS and coexpression (Supplemental Fig. S1). These results are consistent with results of previous analyses showing that the expression of quite a few mirtrons is uncoupled with that of the hosting genes (Wen et al. 2015) and that more than half of the mirtrons for which the TSS was

Downloaded from rnajournal.cshlp.org on July 24, 2018 - Published by Cold Spring Harbor Laboratory Press

determined had independent TSSs (Marsico et al. 2013), suggesting that the processing of these miRNAs might involve multiple alternative mechanisms (Marsico et al. 2013). In the subsequent analyses all intronic miRNAs for which relevant data were available were included.

While our analyses suggest that in general intronic miRNAs are most often transcribed independently and are not coexpressed with their hosts, there are instances of shared TSSs and of coexpression in specific tissues, which might allude to functional associations between the hosting genes and the embedded miRNAs. To look into such possible associations we investigated the targets of the intronic miRNAs, to evaluate their involvement in specific regulatory circuits and in cellular pathways. Below we describe possible types of functional associations between hosting genes and hosted miRNAs, assess their extent, and provide actual examples of such scenarios.

Extent and examples of potential functional associations between intronic miRNAs and their hosts

Intronic miRNAs can potentially influence their host genes' function in multiple ways, including autoregulation of the host gene transcript, regulation of the host gene targets or interactors, and regulation of components of pathways involving the host gene. Previous studies based on predicted targets of miRNAs investigated the functional role of intronic miRNAs on their host (Hinske et al. 2010; Lutter et al. 2010; Gao et al. 2012); however, the predictions are not cell specific such that the extent of the interactions in the frame of a cell is unclear. Hence, we turned to analyze the functional role of intronic miRNAs in a specific cell line using experimentally validated targets obtained by CLASH (involving Argonaute-1 purification), applied to HEK293 cells, CLEAR-CLIP (involving Argonaute-2 immunoprecipitation), applied to Human HUH7.5 cell line, and the iPAR-CLIP analysis applied to several data sets (based on either Argonaute-1 or Argonaute-2 immunoprecipitation) over a number of human cell lines (Helwak et al. 2013; Grosswendt et al. 2014; Moore et al. 2015). The CLASH interactome in HEK293 cells includes 361 mature miRNAs, out of which 193 are intronic miRNAs encoded in 129 protein-coding genes, providing a model system that nicely represents the fraction of intronic miRNAs in the total miRNA repertoire. These 193 intronic miRNAs interact with 3929 targets in the CLASH network. The scale of the CLEAR-CLIP interactome is similar to that of the CLASH and includes 436 mature miRNAs, out of which 188 are intronic miRNAs that interact with 3663 mRNA transcripts. Grosswendt et al. (2014) identified by iPAR-CLIP analysis a network of 273 pre-miRNAs out of which 121 are intronic and interact with 2336 mRNAs. In the latter study, the data came from a number of data sets which used different cell lines, including HEK293 (Kishore et al. 2011; Memczak et al. 2013), human embryonic stem cells (Lipchina et al. 2011), EBV-infected lymphoblastoid cell lines (Skalsky et al. 2012), and primary effusion lymphoma cell lines (Gottwein et al. 2011), and is also at the same scale as the CLASH network. We thus carry out our main analysis using the CLASH data and refer to the two other studies for comparison.

Autoregulation: intronic miRNAs may target their hosts' transcripts

An intronic miRNA can regulate its host directly by targeting the host gene transcript, and thereby allow for the adjustment and maintenance of host transcript levels (Fig. 3). Examples of this form of negative feedback have been observed in a few small-scale studies (Dill et al. 2012; Han et al. 2012; Kos et al. 2012; Liu et al. 2013; Chamorro-Jorganes et al. 2014; Hinske et al. 2015; Qian et al. 2016). A previous study by Hinske et al. (2015), using predicted targets, found that ~20% of intronic miRNAs target their host. Predicted targets, however, do not take into account coexpression of the miRNAs with their hosts in specific cell lines, and therefore it would be informative to study this question using the experimentally determined targets. Of the 193 mature miRNAs included in the CLASH interactome, only two were found to regulate their host genes' transcripts directly: hsa-miR-615-3p was found to target *HOXC4*, and hsa-miR-93-3p was found to target *MCM7*. Chuang et al. (2012) found the expression of hsa-miR-93-3p and *MCM7* to be inversely related in leiomyomas. *MCM7* encodes the MCM7 protein, which is part of the MCM complex, the putative replicative helicase essential for cell cycle replication initiation and elongation (Rebhan et al. 1998). Hence, miR-93-3p may play an essential role in regulating replication and cell cycle progression. No evidence was found for miR-615-3p regulation of *HOXC4* in any publication to date. Based on the coexpression of miR-615-3p and *HOXC5* (a gene that shares a 5′ intron with *HOXC4*), Quah and Holland (2015) speculated that miR-615-3p and HOXC5 have complementary functions, but the possibility of autoregulation has not been discussed.

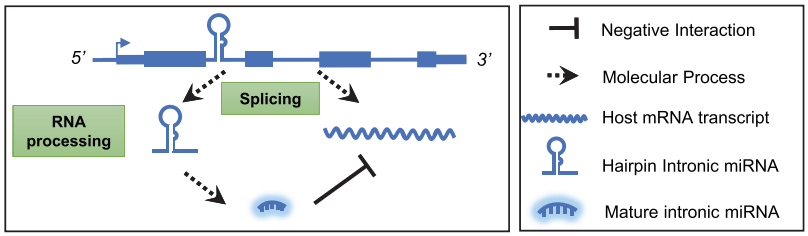

FIGURE 3. Autoregulation of a host gene by an intronic miRNA (schematic representation). An intronic miRNA is transcribed and processed into a mature miRNA that may interact with the transcript of its host gene.

Downloaded from rnajournal.cshlp.org on July 24, 2018 - Published by Cold Spring Harbor Laboratory Press

Considering our previous findings that intronic miRNAs are generally not transcribed with their hosts, it is not surprising that we have few examples of intronic miRNAs targeting their hosts. In accord with that we also find a relatively small number of autoregulatory intronic miRNAs in the other two experimental human data sets, two out of 121 in Grosswendt et al. (2014) data and 17 out of 188 in the CLEAR-CLIP data (Moore et al. 2015). While based on the available data, autoregulation of the hosts by intronic miRNAs seems to be rare, it is possible that in other cell systems miRNAs may be more extensively cotranscribed with their hosts, and in those systems there may be more cases of autoregulation.

Shared targets: intronic miRNAs may target transcripts whose genes/proteins are targeted by the host

Another form of functional association between the host and the intronic miRNA can be through the regulation of common targets (Fig. 4). Host genes can encode proteins that function as transcription factors (TFs) or that interact with other proteins either transiently or stably. In the case of host genes that encode TFs, their intronic miRNAs can influence their function by base-pairing with the transcripts of the target genes (Fig. 4A). If the TF positively regulates its target, then the miRNA antagonizes its host's function, defining a circuit similar in functionality to a mixed incoherent feed forward loop, which might be suitable when short pulses of expression are needed (Alon 2007). On the other hand, if the host TF negatively regulates transcription of its target, then the intronic miRNA acts in collaboration with the host gene, accelerating down-regulation by providing an additional, post-transcriptional, layer of regulation, similar to the dynamics achieved by mixed coherent feed forward loops (Nitzan et al. 2017). In case of host genes that encode proteins that interact with other proteins, their intronic miRNAs can affect their function by targeting the

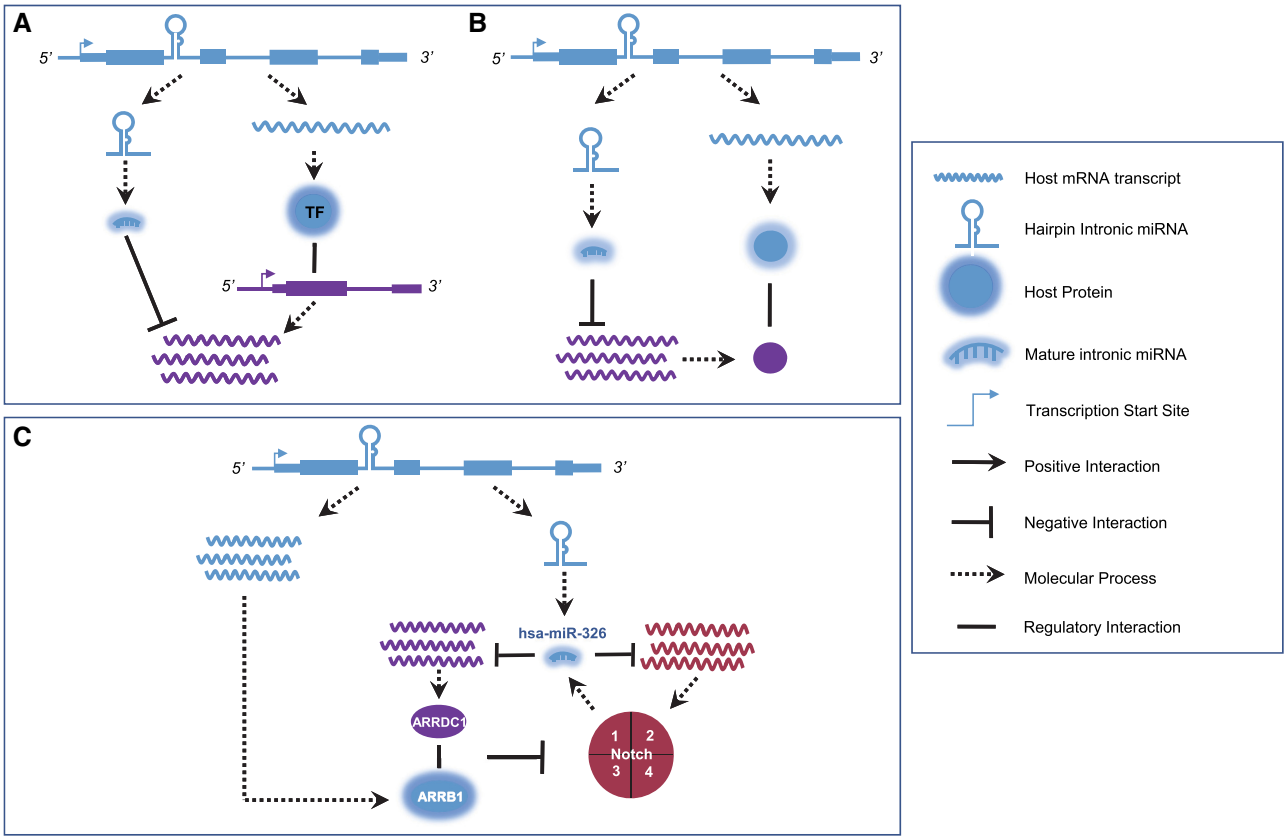

FIGURE 4. Intronic miRNAs may target transcripts whose genes/proteins are targeted by the host protein. (*A*) Schematic representation of a host protein that functions as a TF. The edge from the TF represents either positive or negative regulation. A host TF can positively/negatively regulate the transcription of a gene, while its intronic miRNA represses the transcript of the same gene, thus functioning antagonistically/synergistically to its host's function. (*B*) Schematic representation of a host protein involved in protein–protein interaction. The edge from the protein represents either positive or negative interaction. An intronic miRNA may affect its host antagonistically/synergistically by repressing the transcript of a gene whose protein product is positively/negatively regulated by interaction with the host protein. (*C*) Hsa-miR-326 and *ARRB1*, both in interaction with *ARRDC1* provide an example of a host protein interaction that is affected by its intronic miRNA antagonistically: ARRB1 protein is responsible for the regulation of agonist mediated G protein coupled receptors (GPCR) by mediating receptor sensitization and desensitization. ARRB1 forms a dimer with ARRDC1 to internalize and ubiquitylate the NOTCH receptor. hsa-miR-326 targets *ARRDC1* transcripts. A study by Kefas et al. (2009) found that hsa-miR-326 targets and inhibits both *NOTCH1* and *NOTCH2*, and that knockdown of *NOTCH* leads to increased expression of hsa-miR-326.

Downloaded from rnajournal.cshlp.org on July 24, 2018 - Published by Cold Spring Harbor Laboratory Press

transcripts of the proteins that they interact with (Fig. 4B). Here too, the effect of the intronic miRNA on its host can be either synergistic or antagonistic, depending on the consequence of the interaction of the host protein with its associated protein.

We analyzed these theoretical functional circuits by compiling target data of host genes encoding TFs (based on the TRRUST reference database of human transcriptional regulatory interactions [Han et al. 2015]), and examining them in view of the relevant intronic miRNAs' targets (from the CLASH network [Helwak et al. 2013]). Out of 786 miRNA hosting genes, 45 genes were annotated as encoding TFs included in the TRRUST database and of those 45, seven were included in the CLASH network. Protein–protein interactions for the analysis were taken from the Biogrid Database (Chatr-Aryamontri et al. 2017), and examined in view of the relevant miRNA targets from the CLASH interactome (Helwak et al. 2013). Of the 17,997 proteins included in the Biogrid Database for human interactions, 682 were encoded by genes hosting miRNAs, and of those, 328 were included in the CLASH interactome, and were involved in 12,049 interactions in the Biogrid database. For both TFs and protein–protein interactions, the number of targets or interactors of the host that were found to be shared with intronic miRNA targets was assessed for statistical significance by the hypergeometric test (Supplemental Table S1). The tests took into account the total number of targets or interactors of the host protein expressed in HEK293 cells according to the RNA-seq data of the Human Protein Atlas (Uhlén et al. 2015), as well as the total number of targets of the intronic miRNA. The analysis of host TFs and their intronic miRNAs yielded no statistically significant results. In addition, there were no cases found where an intronic miRNA regulated a TF that regulated its host gene. There were three host-miRNA pairs with protein interaction data for the host, where the overlap between the host interactors and the targets of the intronic miRNA was statistically significant (hypergeometric test, $P \leq 0.05$ after multiple hypotheses correction; Table 1). Accordingly, in the two other human data sets, four (Grosswendt et al. 2014) and two (Moore et al. 2015) intronic miRNAs were found to statistically significantly share target/interactors with their hosts. One example of an antagonistic relationship found in our analysis was between *ARRB1* and its intronic miRNA hsa-miR-326 in the *NOTCH* pathway (Fig. 4C). The *NOTCH* pathway plays a key role in nervous system development and in brain tumors and thus requires tight regulation. Hsa-miR-326 can regulate *NOTCH* on two planes; by base-pairing with *NOTCH* mRNA it maintains the NOTCH levels, and by targeting *ARRDC1* it regulates the level of *NOTCH*'s signaling pathway. Kefas et al. (2009) found that transfection of hsa-miR-326 in glioma cell lines led to the reduction of tumorigenicity of the cells, thus strengthening the potential role that hsa-miR-326 plays in the pathway.

Intronic miRNA targets may participate in the same pathways as the hosts

Intronic miRNAs can target transcripts of genes that encode proteins involved in the same pathway as the host protein. Thus, they can affect the host function synergistically by targeting the transcripts of proteins antagonistic to the hosts' function (Fig. 5A; Barik 2008; Lai et al. 2016), or antagonistically, by targeting transcripts of proteins that cooperate with the host's function (Fig. 5B). We aimed to understand the capacity of these interactions by determining in which pathways this interchange occurred and whether the effects were synergistic or antagonistic to the host. Data of the proteins involved in each pathway were taken from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al. 2016). Each pathway that included a host protein was checked for proteins whose transcripts were targets of the relevant intronic miRNA based on the CLASH network (Helwak et al. 2013). Statistical significance was computed for each pathway by cumulative hypergeometric distribution that took into account the number of pathways that the host participates in, as well as the number of pathways the intronic miRNA's targets participate in, in reference to the total number of pathways found in the KEGG database for human. Out of the 293 human pathways found in KEGG, there were 23 pathways in which nine host-intronic miRNA pairs were found to participate together; however, none were statistically significant (Supplemental Table S2).

TABLE 1. Intronic miRNAs and host genes with shared targets/interactors

miRNA	Host gene	Common interactors/targets	Host interactors/targets	miRNA targets	Cumulative hypergeometric	PhastCons100ways mean conservation score
hsa-miR-186-5p	*ZRANB2*	*ARRB1, HDGF, SRPK1*	18	235	1.26×10^{-2}	0.98[a]/1.0[b]
hsa-miR-326	*ARRB1*	*ARRDC1, FASN, HSPA1B, PKM2, THRAP3*	243	41	3.44×10^{-3}	0.97[a]/1.0[b]
hsa-miR-339-5p	*C7ORF50*	*PAXIP1, RPA1*	18	76	1.42×10^{-2}	0.72[a]/1.0[b]

[a]Score calculated for pre-miRNA.
[b]Score calculated for mature miRNA.

Downloaded from rnajournal.cshlp.org on July 24, 2018 - Published by Cold Spring Harbor Laboratory Press

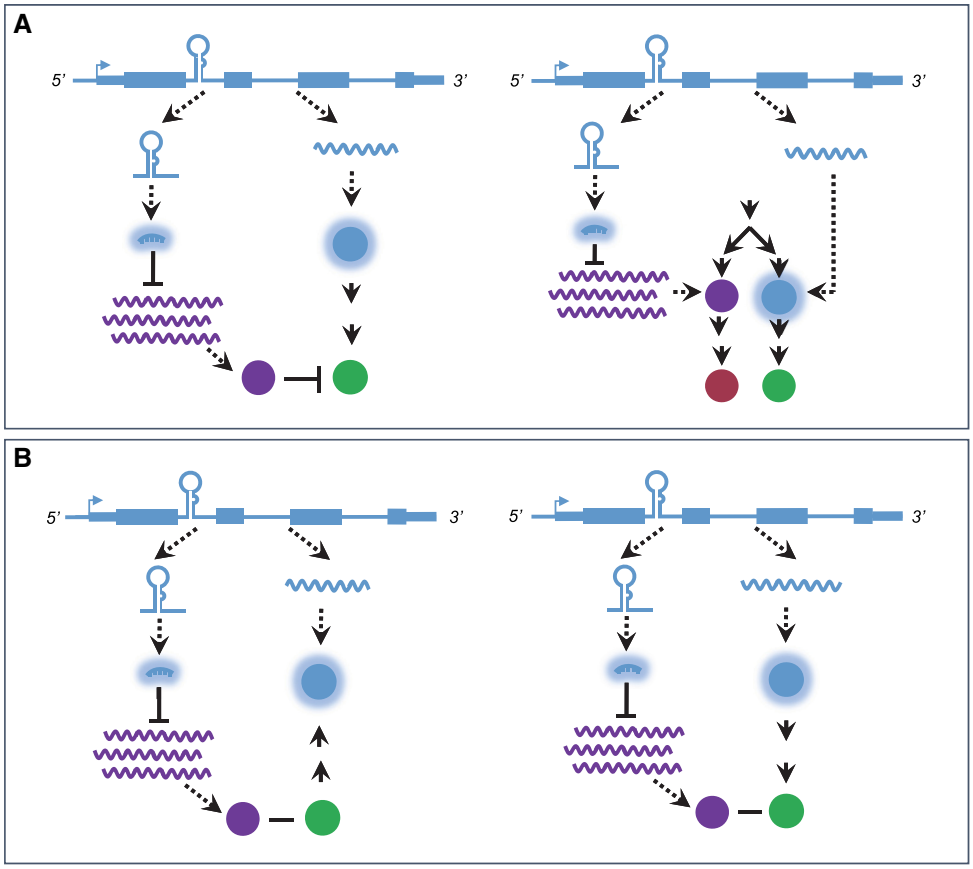

FIGURE 5. Intronic miRNAs may target transcripts of proteins that participate in the same pathway as their hosts. (*A*) Synergistic effects. (*Left*) An intronic miRNA may repress the transcripts of a protein that negatively affects the downstream product of the host. (*Right*) In the case of a host protein participating in a pathway that branches out into subpathways, an intronic miRNA can aid its host by repressing the transcripts of a protein that is part of an alternate subpathway. (*B*) Antagonistic effects. (*Left*) An intronic miRNA can be antagonistic to its host's function by repressing the transcript of a protein that lies in a pathway upstream of the host protein. (*Right*) An intronic miRNA may repress the transcript of a protein that interacts with a protein in a pathway downstream from the host protein.

To further survey the span of the host-miRNA parallel functional output, we analyzed pathways from the KEGG database in which both targets of TF host genes (targets taken from TRRUST database) and targets of their intronic miRNAs were involved. We identified 32 statistically significant unique pathways shared between the targets of five intronic miRNAs and the targets of their respective transcription-factor-encoding hosting genes (Table 2). Statistical significance was computed for each pathway through Z test by normal distribution approximation of the binomial distribution (Supplemental Fig. S2). Here too, antagonistic or synergistic interactions were observed between intronic miRNAs and their hosts, including: *DNMT3A* and hsa-miR-1301-3p (Fig. 6); and *NFYC* and hsa-miR-30c-5p (Fig. 7). One

TABLE 2. Intronic miRNAs hosted in genes encoding TFs that participate in pathways with targets of the host protein

Intronic miRNA	Host gene	Number of pathways with statistically significant shared targets	PhastCons100ways mean conservation score
hsa-miR-1301-3p	*DNMT3A*	2	0.98[a]/1.0[b]
hsa-miR-30c-5p	*NFYC*	19	0.98[a]/1.0[b]
hsa-miR-30e-5p	*NFYC*	6	0.98[a]/1.0[b]
hsa-miR-30e-3p	*NFYC*	1	0.98[a]/1.0[b]
hsa-miR-33b-5p	*SREBF1*	4	0.79[a]/0.98[b]

[a]Score calculated for pre-miRNA.
[b]Score calculated for mature miRNA.

Downloaded from rnajournal.cshlp.org on July 24, 2018 - Published by Cold Spring Harbor Laboratory Press

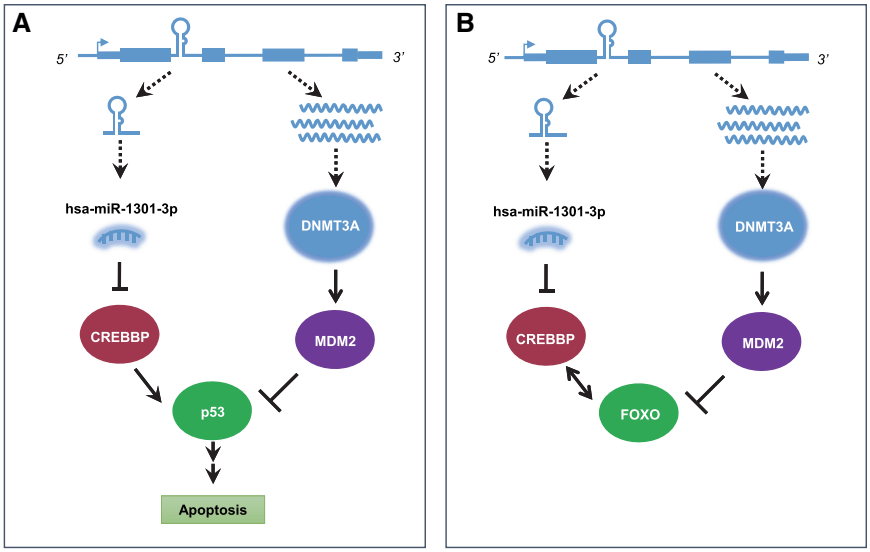

FIGURE 6. MiRNA and host targets are involved in common pathways. (*A*) *DNMT3A*-hsa-miR-1301-3p cancer promoting coherent regulatory circuit. DNMT3A activates the transcription of *MDM2* proto-oncogene, which represses p53 and thus prevents apoptosis. hsa-miR-1301-3p synergistically supports *DNMT3A* by repressing CREBBP, a cofactor of p53. A number of other pathways were also found to be statistically significantly enriched in targets of the host DNMT3A and targets of its hosted miRNA, hsa-miR-1301-3p (Table 2). (*B*) In the FOXO signaling pathway, DNMT3A activates *MDM2*, which leads to the ubiquitination and eventual proteolysis of the FOXO protein (Fu et al. 2009). hsa-miR-1301-3p collaborates with its host function by repressing CREBBP, a cofactor of FOXO (Daitoku et al. 2011).

pathway in which hsa-miR-1301-3p was found to share targets with its host *DNMT3A* was the prostate cancer pathway (Fig. 6A), in accord with a recent study showing that overexpression of hsa-mir-1301 promoted prostate cancer proliferation (Bi et al. 2016).

miRNAs that are functionally associated with their host genes are evolutionarily conserved

We classified the intronic miRNAs into two groups, miRNAs that showed functional association with their hosts (a total of seven out of 193 intragenic miRNAs in the CLASH network, Tables 1, 2) and those that did not (the rest). Comparison of the distributions of conservation scores of the two groups showed that the miRNAs functionally associated with their hosts are statistically significantly more conserved than the rest of the intronic miRNAs (Fig. 8, Wilcoxon rank sum test $P \leq 4.42 \times 10^{-4}$), supporting

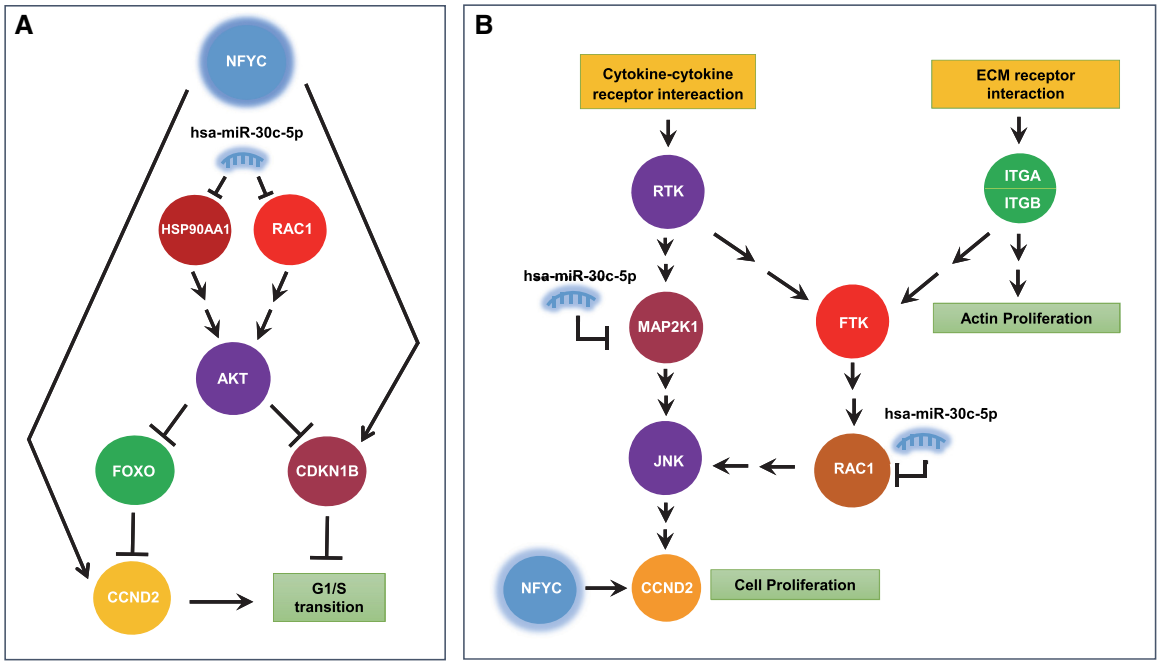

FIGURE 7. MiRNA and host targets are involved in common pathways. (*A*) *NFYC*-hsa-miR-30c-5p circuit in the PI3K-AKT signaling pathway. The PI3K-AKT signaling pathway is responsible for regulating the cell cycle and thus leads to different downstream responses, including metabolism, protein synthesis, glycolysis, apoptosis, and cell cycle progression among others. NFYC activates transcription of both *CCND2* and *CDKN1B*, promoting G1/S transition. hsa-miR-30c-5p antagonizes *NFYC* by targeting *HSP90AA1* and *RAC1*, thus leading to the repression of *AKT*, which results in the promotion of FOXO, which then represses *CCND2*, thus preventing G1 progression. In addition, AKT also represses *CDKN1B*, thus the repression of *AKT* promotes CDKN1B, once again, preventing G1 transition (Rebhan et al. 1998). (*B*) *NFYC*-hsa-miR-30c-5p circuit in the focal adhesion pathway. NFYC activates transcription of *CCND2*, which promotes cell proliferation in the focal adhesion pathway. hsa-miR30c-5p antagonizes its host by targeting *MAP2K1* and *RAC1*, which transcribe proteins found upstream of CCND2 in the focal adhesion pathway.

Downloaded from rnajournal.cshlp.org on July 24, 2018 - Published by Cold Spring Harbor Laboratory Press

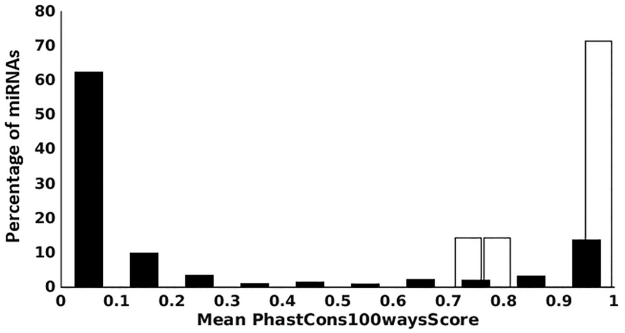

FIGURE 8. MiRNAs that are functionally associated with their hosts are evolutionarily conserved. MiRNAs functionally associated with their hosts (white bars) show higher sequence conservation than the other intronic miRNA sequences (black bars; Wilcoxon rank sum test, $P \leq 4.42 \times 10^{-4}$). Conservation analysis was conducted on the hairpin pre-miRNA sequences.

their functional importance. Analysis of the other studied characteristics (cotranscription and coexpression) resulted in no statistically significant difference between the two miRNA groups.

DISCUSSION

Intronic miRNAs are highly abundant in various organisms (Meunier et al. 2013). By surveying several genomes it was shown that 36%–65% of the miRNA genes are encoded within introns, while introns constitute ~22%–45% of a genome, thus alluding to a possible evolutionary advantage of this genetic architecture (Meunier et al. 2013). It is possible that the prevalence of miRNAs in introns may simply be a result of the expediency of miRNAs to be derived from already transcribed noncoding regions, such as the intron. On the other hand, it is possible that host genes might offer an initial evolutionary advantage to miRNAs by providing a suitable environment for their integration into transcriptional networks, as suggested by França et al. (2016). A conceivable conjecture would have been that the intronic miRNAs and hosting genes are functionally related and cooperate in the same cellular mechanisms. We aimed to assess this conjecture and the extent of the functional association between the hosting genes and hosted miRNAs using experimental expression data of miRNAs as well as experimentally validated targets.

Theoretically, miRNAs can be functionally associated with their hosts in a variety of ways. Yet, a prerequisite for all the various types of possible functional associations between intronic miRNAs and hosts is their coexpression. In principle, intronic miRNAs can be transcribed in association with the host, or independently, as observed for prokaryotic sRNAs embedded in 3′UTRs of protein-coding genes (Miyakoshi et al. 2015). Intronic miRNAs can be transcribed and spliced out of the mRNA, generating pri-miRNA that is subsequently processed by Drosha and DGCR8 to produce the pre-miRNA, or they can be transcribed from promoters that are independent of the promoters of the hosting genes. The first scenario guarantees coexpression of the hosting gene and hosted miRNA, while coexpression is less probable but still possible by the second scenario. Examples of independent transcription of intronic miRNAs have been reported in previous studies of various organisms. In *Drosophila melanogaster,* Xiong et al. (2009) demonstrated transcription of an intronic miRNA, miR-281, independently of its host gene ODA. In mouse, Polster et al. (2010) revealed discordant expression between miR-103 and miR-107 and their pantothenate kinase host genes. In addition, recent studies using sequencing technologies, such as rapid amplification of cDNA ends (RACE), showed that intronic miRNAs can be transcribed independently of their host transcripts (Ramalingam et al. 2014). Moreover, identification of DNA features of promoters using CHIP-seq has also supported the understanding that intronic miRNAs may have promoters independent of their hosts (Ozsolak et al. 2008; Wang et al. 2009). Here, we assessed the coexpression of miRNA and hosting gene in human by analyzing both experimental data of TSSs and cell line specific expression data of miRNAs and protein-coding genes. Our examination of cell specific intronic miRNA TSSs based on the analysis of Hua et al. (2016), as well as findings in other papers (Monteys et al. 2010; Marsico et al. 2013; Budach et al. 2016), suggest that the majority of the intronic miRNAs are independently transcribed in many cell lines. Consistently, using the available expression data we also found that many intronic miRNAs are not coexpressed in the same tissues as their hosting genes. These findings by themselves, without turning yet to target data, might suggest that functional association between miRNAs and their hosting genes is narrower than expected. These findings also emphasize the importance of using experimental data in the next phase of the analysis rather than predictions, as predicted targets indicate potential miRNA–mRNA interactions without taking into account the cotranscription of the miRNA and its target in a cellular context.

Initial large scale studies of the functional association between intronic miRNAs and host genes in eukaryotes were based on prediction of miRNA targets (Hinske et al. 2010; Gao et al. 2012) and argued that intronic miRNAs widely affect their host function. In the current study, using a model system based on the experimentally determined network of miRNA-target interactions in HEK293 cells (Helwak et al. 2013), we found that a small number of intronic miRNAs out of the 193 intronic miRNAs included in the CLASH network are functionally associated with their hosts. This small number is further supported by the CLEAR-CLIP miRNA target data obtained in a hepatoma cell line (Moore et al. 2015) and by the data from Grosswendt et al. (2014). On the one hand, since in all three studies the intronic miRNAs constitute ~50% of the miRNAs in the network, it can be argued that they provide representative model systems from which it might be extrapolated that functional association between a hosting gene and an embedded miRNA is

Downloaded from rnajournal.cshlp.org on July 24, 2018 - Published by Cold Spring Harbor Laboratory Press

not widespread. On the other hand, it can be argued that these results were obtained from data in specific cellular contexts, and additional experimental target data obtained in the future in other cellular contexts may expand this repertoire. Interestingly, we found that intronic miRNAs that show a functional association with their hosts are more evolutionarily conserved than other intronic miRNAs (Fig. 8). This may suggest that other conserved intronic miRNAs for which data were not available in the current experimentally determined networks might be identified in future data as functionally associated with their hosts, as suggested by the previous studies based on predictions.

Our analysis shows that intronic miRNAs that do affect their host's function may do so in a variety of ways: through autoregulation of the host, through the regulation of targets shared with their host gene, or by targeting genes that participate in the same pathways as their host. These intronic miRNAs were found in key pathways such as PI3K-AKT signaling pathway, responsible for the regulation of the cell cycle, and in replication pathways, cell signaling pathways, as well as cancer pathways, all of which are tightly regulated. In these pathways we found that the function of the intronic miRNA could be either antagonistic or synergistic to the hosting gene function. Furthermore, the specificity of miRNA expression in different tissues and conditions, as well as the expression of different targets in these tissues and conditions, enables the intronic miRNAs to influence the host-involved pathways differently in various situations. Thus, the current data provide variable examples that span a wide range of regulatory scenarios, but is too limited to enable the inference of general design principles for the involvement of the intronic miRNAs in the cellular regulatory circuitry. Additional experimental data sets of miRNA targets should enable further assessment of the extent of such functional modules and of their repertoire.

MATERIALS AND METHODS

Intronic miRNAs were identified from the chromosomal coordinates of human miRNAs taken from miRBase v21, based on the GRCh38 genome build (Griffiths-Jones 2004, 2006; Griffiths-Jones et al. 2008; Kozomara and Griffiths-Jones 2011, 2014), and mapped into genes annotated in the RefSeq Human Genome Annotations, build hg38, using an in-house script. Only miRNAs and genes annotated on the same strand were considered. 866 precursor miRNAs (pre-miRNAs) were found in 786 protein-coding genes (Supplemental File S1). Mouse and *C. elegans* intronic miRNAs were extracted from build GRCm38 (mouse) and build WBcel235 (*C. elegans*).

Experimental data used in the study were downloaded from the following respective sources:

1. In the main analysis of the study we included miRNA targets based on 14,684 high confidence interactions determined by the CLASH methodology (Helwak et al. 2013) in the Human HEK293 cell line (361 mature miRNAs, 6256 target genes, Supplemental File S2). For comparison, we also used data determined by the CLEAR-CLIP method (Moore et al. 2015) in the Human HUH7.5 cell line (436 mature miRNAs, 6746 target genes). From both data sets we included interactions of miRNAs found to be bound to 3′UTR, 5′UTR, or CDS with binding energy ≤ -13.4 kcal/mol. Another data set used was of endogenously ligated miRNA-mRNA pairs in human cell lines from the study of Grosswendt et al. (2014) (273 miRNAs, 4573 target genes).
2. Transcriptional regulatory interactions of the host genes were downloaded from the TRRUST reference database of human transcriptional regulatory interactions (748 TFs, 714 of which are found to be expressed in HEK293 according to the RNA-seq expression data of the Human Protein Atlas [Supplemental File S3; Uhlén et al. 2015]). Forty-five of the TFs host intronic miRNAs and are involved in 475 transcriptional regulatory interactions (Han et al. 2015). When investigating common targets between the host TFs and intronic miRNAs, only TFs found to be expressed in HEK293 according to the RNA-seq data of the Human Protein Atlas (Uhlén et al. 2015) were used.
3. Protein–protein interaction data were downloaded from the Biogrid Database Release 3.2.115 (17997 proteins involved in 144729 direct interactions) (Chatr-Aryamontri et al. 2017). 13780 of these proteins are shown to be expressed in HEK293 (based on RNA-seq data from the Human Protein Atlas [Uhlén et al. 2015]), and are involved in 140405 interactions (Supplemental File S4). The encoding genes of 682 proteins in this database host intronic miRNAs. These proteins are involved in 16,674 interactions.
4. Human pathway data were taken from KEGG database (Kanehisa et al. 2016) and included 293 human pathways (Supplemental File S5)
5. Transcription start site (TSS) data of intronic miRNAs were taken from Hua et al. (2016). The study identified cell-specific miRNA TSSs in 54 cell lines, by integrating data of H3K4me3 and DNase I hypersensitivity sites from the ENCODE project (The Encode Project Consortium 2012), as well as by considering conservation and sequence features. Hua et al. (2016) defined the TSSs as host dependent or independent relying on the distance of the intronic miRNA's candidate TSS from its host TSS.
6. Expression data of miRNAs were taken from either Landgraf and coworkers' mammalian miRNA expression atlas, using the pre-miRNA expression data (Landgraf et al. 2007), or from miRmine, a recent resource of miRNA expression profiles in numerous cell lines and tissues from publicly available human miRNA-seq data (Panwar et al. 2017).
7. RNA-seq data of the hosting genes were taken from the Human Protein Atlas (www.proteinatlas.org [Uhlén et al. 2015]).

MiRNA sequence conservation was calculated by finding the mean PhastCons100way score for each hairpin precursor or mature miRNA sequence. Comparison of the miRNA conservation scores between two sets of miRNAs was done by Wilcoxon Rank Sum Test with R (R Core Team 2016). MATLAB R2014b (MathWorks) was used for generating the graphs.

The analysis of coexpression of intronic miRNAs with their host genes was done by computing a Jaccard index, as follows: For each intronic miRNA-host pair, let A be the set of tissues/cell lines in which the miRNA is expressed and B the set of tissues/cell lines in which the host is expressed. $A \cap B$ is the number of tissues/cell

Downloaded from rnajournal.cshlp.org on July 24, 2018 - Published by Cold Spring Harbor Laboratory Press

lines where the miRNA and its host were coexpressed. $A \cup B$ is the total number of cell lines in which either the miRNA or its host were expressed. The Jaccard index of coexpression is obtained by $(A \cap B)/(A \cup B)$. It ranges between 0 and 1, where 0 indicates no coexpression in any tissue and 1 indicates that the miRNA and host gene were coexpressed in all tissues.

To analyze intronic miRNAs dependency on their hosts' promoters for transcription initiation, a Jaccard index was calculated for the TSS data from Hua et al. (2016): For each intronic miRNA we divided the number of cell lines where the miRNA TSS shares a promoter with the host and is therefore dependent on the host's promoter for transcription initiation, by the total number of cell lines in which we had TSS data for the intronic miRNA (either the miRNA was independently transcribed or its transcription was dependent on host promoter). Intronic miRNAs that were dependent on their host gene's promoter for transcription in all cell lines in the data obtained a Jaccard index of 1, whereas those miRNAs that had independent TSSs in all cell lines in the data obtained a Jaccard index of 0.

The functional association between a hosted miRNA and its hosting gene was assessed in several ways.

1. When shared targets were considered we computed the probability to obtain at random at least k overlapping targets between a hosted miRNA and a hosting gene (encoding either a TF or a protein involved in protein–protein interaction, Table 1; Supplemental Table S1). This was computed by the cumulative hypergeometric distribution, given the number of targets of each intronic miRNA and host, in reference to the total number of genes found in a union of the Biogrid and TRRUST databases, as well as genes found in the experimental miRNA-target network being assessed (CLASH/CLEAR-CLIP/Grosswendt et al. data, 2014). For the CLASH data, the Biogrid and TRRUST databases were filtered to include only those TFs/genes found to be expressed in HEK293.
2. When involvement in common pathways was considered, we computed the statistical significance of the host and miRNA to share pathways in two ways: (i) The statistical significance of an intronic miRNA's targets (based on CLASH network) participating in at least k pathways together with the host gene (Supplemental Table S2) was calculated by the cumulative hypergeometric distribution, given the number of pathways that the host participates in, as well as the number of pathways the intronic miRNA's targets participate in, in reference to the total number of pathways found in the KEGG database (Kanehisa et al. 2016) for human. (ii) To determine the significance of an intronic miRNA's targets participating in a pathway that its host targets participated in (Table 2), we used the Z test using normal distribution approximation of the binomial distribution. For each shared pathway of the host gene and its intronic miRNA, we calculated the probability to obtain at least the observed number of miRNA and host targets given the number of the miRNA targets (based on CLASH network) and the host targets expected at random to participate in the pathway (Supplemental Fig. S2).

All *P*-values were adjusted for multiple testing by FDR (Benjamini and Hochberg 1995). Cumulative hypergeometric distribution and *P*-value adjustment for multiple testing were carried out using the R Stats library (R Core Team 2016).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank Yael Altuvia for her useful comments. This study was supported by the Israeli Centers for Research Excellence (I-CORE) program of the Planning and Budgeting Committee and the Israeli Science Foundation (grant no. 41/1), the European Research Council (grant no. 322920), and the Israeli Science Foundation administered by the Israeli Academy for Sciences and Humanities (grant no. 1411/13).

Received October 8, 2017; accepted April 26, 2018.

REFERENCES

Alon U. 2007. Network motifs: theory and experimental approaches. *Nat Rev Genet* **8:** 450–461.

Altuvia Y, Landgraf P, Lithwick G, Elefant N, Pfeffer S, Aravin A, Brownstein MJ, Tuschl T, Margalit H. 2005. Clustering and conservation patterns of human microRNAs. *Nucleic Acids Res* **33:** 2697–2706.

Asli NS, Pitulescu ME, Kessel M. 2008. MicroRNAs in organogenesis and disease. *Curr Mol Med* **8:** 698–710.

Barik S. 2008. An intronic microRNA silences genes that are functionally antagonistic to its host gene. *Nucleic Acids Res* **36:** 5232–5241.

Baskerville S, Bartel DP. 2005. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* **11:** 241–247.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* **57:** 289–300.

Bi D, Ning H, Liu S, Que X, Ding K. 2016. miR-1301 promotes prostate cancer proliferation through directly targeting PPP2R2C. *Biomed Pharmacother* **81:** 25–30.

Boivin V, Deschamps-Francoeur G, Scott MS. 2017. Protein coding genes as hosts for noncoding RNA expression. *Semin Cell Dev Biol* **75:** 3–12.

Budach S, Heinig M, Marsico A. 2016. Principles of microRNA regulation revealed through modeling microRNA expression quantitative trait loci. *Genetics* **203:** 1629–1640.

Chamorro-Jorganes A, Araldi E, Rotllan N, Cirera-Salinas D, Suárez Y. 2014. Autoregulation of glypican-1 by intronic microRNA-149 fine tunes the angiogenic response to FGF2 in human endothelial cells. *J Cell Sci* **127:** 1169–1178.

Chao Y, Vogel J. 2016. A 3′ UTR-derived small RNA provides the regulatory noncoding arm of the inner membrane stress response. *Mol Cell* **61:** 352–363.

Chatr-Aryamontri A, Oughtred R, Boucher L, Rust J, Chang C, Kolas NK, O'Donnell L, Oster S, Theesfeld C, Sellam A, et al. 2017. The BioGRID interaction database: 2017 update. *Nucleic Acids Res* **45:** D369–D379.

Chi SW, Zang JB, Mele A, Darnell RB. 2009. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460:** 479–486.

Chuang TD, Luo X, Panda H, Chegini N. 2012. miR-93/106b and their host gene, *MCM7*, are differentially expressed in leiomyomas and functionally target *F3* and *IL-8*. *Mol Endocrinol* **6:** 1028–1042.

Daitoku H, Sakamaki J, Fukamizu A. 2011. Regulation of FoxO transcription factors by acetylation and protein-protein interactions. *Biochim Biophys Acta* **1813:** 1954–1960.

Dill H, Linder B, Fehr A, Fischer U. 2012. Intronic miR-26b controls neuronal differentiation by repressing its host transcript, ctdsp2. *Genes Dev* **26:** 25–30.

Downloaded from rnajournal.cshlp.org on July 24, 2018 - Published by Cold Spring Harbor Laboratory Press

The Encode Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489:** 57–74.

França GS, Vibranovski MD, Galante PA. 2016. Host gene constraints and genomic context impact the expression and evolution of human microRNAs. *Nat Commun* **7:** 11438.

Fu W, Ma Q, Chen L, Li P, Zhang M, Ramamoorthy S, Nawaz Z, Shimojima T, Wang H, Yang Y, et al. 2009. MDM2 acts downstream of p53 as an E3 ligase to promote FOXO ubiquitination and degradation. *The J Biol Chem* **284:** 13987–14000.

Gao X, Qiao Y, Han D, Zhang Y, Ma N. 2012. Enemy or partner: relationship between intronic microRNAs and their host genes. *IUBMB Life* **64:** 835–840.

Gottwein E, Corcoran DL, Mukherjee N, Skalsky RL, Hafner M, Nusbaum JD, Shamulailatpam P, Love CL, Dave SS, Tuschl T, et al. 2011. Viral microRNA targetome of KSHV-infected primary effusion lymphoma cell lines. *Cell Host Microbe* **10:** 515–526.

Griffiths-Jones S. 2004. The microRNA registry. *Nucleic Acids Res* **32:** D109–D111.

Griffiths-Jones S. 2006. miRBase: the microRNA sequence database. *Methods Mol Biol* **342:** 129–138.

Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Res* **36:** D154–D158.

Grosswendt S, Filipchyk A, Manzano M, Klironomos F, Schilling M, Herzog M, Gottwein E, Rajewsky N. 2014. Unambiguous identification of miRNA:target site interactions by different types of ligation reactions. *Mol Cell* **54:** 1042–1054.

Ha M, Kim VN. 2014. Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* **15:** 509–524.

Hafner M, Lianoglou S, Tuschl T, Betel D. 2012. Genome-wide identification of miRNA targets by PAR-CLIP. *Methods* **58:** 94–105.

Han J, Denli AM, Gage FH. 2012. The enemy within: intronic miR-26b represses its host gene, ctdsp2, to regulate neurogenesis. *Genes Dev* **26:** 6–10.

Han H, Shim H, Shin D, Shim JE, Ko Y, Shin J, Kim H, Cho A, Kim E, Lee T, et al. 2015. TRRUST: a reference database of human transcriptional regulatory interactions. *Sci Rep* **5:** 11432.

Helwak A, Kudla G, Dudnakova T, Tollervey D. 2013. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153:** 654–665.

Hinske LC, Galante PA, Kuo WP, Ohno-Machado L. 2010. A potential role for intragenic miRNAs on their hosts' interactome. *BMC Genomics* **11:** 533.

Hinske LC, Galante PA, Limbeck E, Möhnle P, Parmigiani RB, Ohno-Machado L, Camargo AA, Kreth S. 2015. Alternative polyadenylation allows differential negative feedback of human miRNA miR-579 on its host gene ZFR. *PLoS One* **10:** e0121507.

Hua X, Chen L, Wang J, Li J, Wingender E. 2016. Identifying cell-specific microRNA transcriptional start sites. *Bioinformatics* **32:** 2403–2410.

Inui M, Martello G, Piccolo S. 2010. MicroRNA control of signal transduction. *Nat Rev Mol Cell Biol* **11:** 252–263.

Jonas S, Izaurralde E. 2015. Towards a molecular understanding of microRNA-mediated gene silencing. *Nat Rev Genet* **16:** 421–433.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44:** D457–D462.

Kefas B, Comeau L, Floyd DH, Seleverstov O, Godlewski J, Schmittgen T, Jiang J, diPierro CG, Li Y, Chiocca EA, et al. 2009. The neuronal microRNA miR-326 acts in a feedback loop with notch and has therapeutic potential against brain tumors. *J Neurosci* **29:** 15161–15168.

Kishore S, Jaskiewicz L, Burger L, Hausser J, Khorshid M, Zavolan M. 2011. A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat Methods* **8:** 559–564.

Kos A, Olde Loohuis NF, Wieczorek ML, Glennon JC, Martens GJ, Kolk SM, Aschrafi A. 2012. A potential regulatory role for intronic microRNA-338–3p for its host gene encoding apoptosis-associated tyrosine kinase. *PLoS One* **7:** e31022.

Kozomara A, Griffiths-Jones S. 2011. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* **39:** D152–D157.

Kozomara A, Griffiths-Jones S. 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42:** D68–D73.

Lai L, Azzam KM, Lin WC, Rai P, Lowe JM, Gabor KA, Madenspacher JH, Aloor JJ, Parks JS, Näär AM, et al. 2016. MicroRNA-33 regulates the innate immune response via ATP binding cassette transporter-mediated remodeling of membrane microdomains. *J Biol Chem* **291:** 19651–19660.

Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, et al. 2007. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129:** 1401–1414.

Lee HM, Nguyen DT, Lu LF. 2014. Progress and challenge of microRNA research in immunity. *Front Genet* **5:** 178.

Li Y, Qiu C, Tu J, Geng B, Yang J, Jiang T, Cui Q. 2014. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res* **42:** D1070–D1074.

Lipchina I, Elkabetz Y, Hafner M, Sheridan R, Mihailovic A, Tuschl T, Sander C, Studer L, Betel D. 2011. Genome-wide identification of microRNA targets in human ES cells reveals a role for miR-302 in modulating BMP response. *Genes Dev* **25:** 2173–2186.

Liu M, Roth A, Yu M, Morris R, Bersani F, Rivera MN, Lu J, Shioda T, Vasudevan S, Ramaswamy S, et al. 2013. The IGF2 intronic miR-483 selectively enhances transcription from IGF2 fetal promoters and enhances tumorigenesis. *Genes Dev* **27:** 2543–2548.

Lutter D, Marr C, Krumsiek J, Lang EW, Theis FJ. 2010. Intronic microRNAs support their host genes by mediating synergistic and antagonistic regulatory effects. *BMC Genomics* **11:** 224.

Marsico A, Huska MR, Lasserre J, Hu H, Vucicevic D, Musahl A, Orom U, Vingron M. 2013. PROmiRNA: a new miRNA promoter recognition method uncovers the complex regulation of intronic miRNAs. *Genome Biol* **14:** R84.

Melamed S, Peer A, Faigenbaum-Romm R, Gatt YE, Reiss N, Bar A, Altuvia Y, Argaman L, Margalit H. 2016. Global mapping of small RNA-target interactions in bacteria. *Mol Cell* **63:** 884–897.

Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M, et al. 2013. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495:** 333–338.

Meunier J, Lemoine F, Soumillon M, Liechti A, Weier M, Guschanski K, Hu H, Khaitovich P, Kaessmann H. 2013. Birth and expression evolution of mammalian microRNA genes. *Genome Res* **23:** 34–45.

Miyakoshi M, Chao Y, Vogel J. 2015. Regulatory small RNAs from the 3′ regions of bacterial mRNAs. *Curr Opin Microbiol* **24:** 132–139.

Monteys AM, Spengler RM, Wan J, Tecedor L, Lennox KA, Xing Y, Davidson BL. 2010. Structure and activity of putative intronic miRNA promoters. *RNA* **16:** 495–505.

Moore MJ, Scheel TK, Luna JM, Park CY, Fak JJ, Nishiuchi E, Rice CM, Darnell RB. 2015. miRNA-target chimeras reveal miRNA 3′-end pairing as a major determinant of Argonaute target specificity. *Nat Commun* **6:** 8864.

Nitzan M, Rehani R, Margalit H. 2017. Integration of bacterial small RNAs in regulatory networks. *Annu Rev Biophys* **46:** 131–148.

Ozsolak F, Poling LL, Wang Z, Liu H, Liu XS, Roeder RG, Zhang X, Song JS, Fisher DE. 2008. Chromatin structure analyses identify miRNA promoters. *Genes Dev* **22:** 3172–3183.

Panwar B, Omenn GS, Guan Y. 2017. miRmine: a database of human miRNA expression profiles. *Bioinformatics* **33:** 1554–1560.

Pickl JM, Tichy D, Kuryshev VY, Tolstov Y, Falkenstein M, Schüler J, Reidenbach D, Hotz-Wagenblatt A, Kristiansen G, Roth W, et al. 2016. Ago-RIP-Seq identifies Polycomb repressive complex I member CBX7 as a major target of miR-375 in prostate cancer progression. *Oncotarget* **7:** 59589–59603.

Polster BJ, Westaway SK, Nguyen TM, Yoon MY, Hayflick SJ. 2010. *Mol Genet Metab* **101:** 292–295.

Downloaded from rnajournal.cshlp.org on July 24, 2018 - Published by Cold Spring Harbor Laboratory Press

Pruitt KD, Tatusova T, Maglott DR. 2005. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **33:** D501–D504.

Qian J, Tu R, Yuan L, Xie W. 2016. Intronic miR-932 targets the coding region of its host gene, Drosophila neuroligin2. *Exp Cell Res* **344:** 183–193.

Quah S, Holland PW. 2105. The Hox cluster microRNA miR-615: a case study of intronic microRNA evolution. *Evodevo* **6:** 31.

R Core Team. 2016. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramalingam P, Palanichamy JK, Singh A, Das P, Bhagat M, Kassab MA, Sinha S, Chattopadhyay P. 2014. Biogenesis of intronic miRNAs located in clusters by independent transcription and alternative splicing. *RNA* **20:** 76–87.

Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. 1998. GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support. *Bioinformatics* **14:** 656– 664.

Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. 2004. Identification of mammalian microRNA host genes and transcription units. *Genome Res* **14:** 1902–1910.

Ruby JG, Jan CH, Bartel DP. 2007. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448:** 83–86.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15:** 1034–1050.

Skalsky RL, Corcoran DL, Gottwein E, Frank CL, Kang D, Hafner M, Nusbaum JD, Feederle R, Delecluse HJ, Luftig MA, et al. 2012. The viral and cellular microRNA targetome in lymphoblastoid cell lines. *PLoS Pathog* **8:** e1002484.

Tan SM, Kirchner R, Jin J, Hofmann O, McReynolds L, Hide W, Lieberman J. 2014. Sequencing of captive target transcripts identifies the network of regulated genes and functions of primate-specific miR-522. *Cell Rep* **8:** 1225–1239.

Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjöstedt E, Asplund A, et al. 2015. Tissue-based map of the human proteome. *Science* **347:** 1260419.

Wang X, Xuan Z, Zhao X, Li Y, Zhang MQ. 2009. High-resolution human core-promoter prediction with CoreBoost_HM. *Genome Res* **19:** 266–275.

Wen J, Ladewig E, Shenker S, Mohammed J, Lai EC. 2015. Analysis of nearly one thousand mammalian mirtrons reveals novel features of dicer substrates. *PLoS Comput Biol* **11:** e1004441.

Westholm JO, Lai EC. 2011. Mirtrons: microRNA biogenesis via splicing. *Biochimie* **93:** 1897–1904.

Xiong H, Qian J, He T, Li F. 2009. Independent transcription of miR-281 in the intron of ODA in *Drosophila melanogaster*. *Biochem Biophys Res Commun* **378:** 883–889.

Downloaded from rnajournal.cshlp.org on July 24, 2018 - Published by Cold Spring Harbor Laboratory Press

RNA
A PUBLICATION OF THE RNA SOCIETY

Assessing the functional association of intronic miRNAs with their host genes

Avital Steiman-Shimony, Orr Shtrikman and Hanah Margalit

RNA 2018 24: 991-1004 originally published online May 11, 2018
Access the most recent version at doi:10.1261/rna.064386.117

Supplemental Material http://rnajournal.cshlp.org/content/suppl/2018/05/11/rna.064386.117.DC1

References This article cites 75 articles, 18 of which can be accessed free at: http://rnajournal.cshlp.org/content/24/8/991.full.html#ref-list-1

Open Access Freely available online through the *RNA* Open Access option.

Creative Commons License This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

To subscribe to *RNA* go to:
http://rnajournal.cshlp.org/subscriptions

© 2018 Steiman-Shimony et al.; Published by Cold Spring Harbor Laboratory Press for the RNA Society