A noncanonical microRNA derived from the snaR-A noncoding RNA targets a metastasis inhibitor

DANIEL STRIBLING1,2,3,7 YI LEI4,5,7,8 CASEY M. GUARDIA4,5,7,9 LU LI4,5,7,9 CHRISTOPHER J. FIELDS4,5 PAWEL NOWIALIS5,6 RENE OPAVSKY5,6 ROLF RENNE1,2,3,5 and MINGYI XIE2,4,5

1Department of Molecular Genetics and Microbiology, University of Florida, Gainesville, Florida 32610, USA
2UF Genetics Institute, University of Florida, Gainesville, Florida 32610, USA
3UF Informatics Institute, University of Florida, Gainesville, Florida 32611, USA
4Department of Biochemistry and Molecular Biology, University of Florida, Gainesville, Florida 32610, USA
5UF Health Cancer Center, University of Florida, Gainesville, Florida 32610, USA
6Department of Anatomy and Cell Biology, University of Florida, Gainesville, Florida 32610, USA

ABSTRACT

MicroRNAs (miRNAs) are small noncoding RNAs that function as critical posttranscriptional regulators in various biological processes. While most miRNAs are generated from processing of long primary transcripts via sequential Drosha and Dicer cleavage, some miRNAs that bypass Drosha cleavage can be transcribed as part of another small noncoding RNA. Here, we develop the target-oriented miRNA discovery (TOMiD) bioinformatic analysis method to identify Drosha-independent miRNAs from ArgoNaut crosslinking and sequencing of hybrids (Ago-CLASH) data sets. Using this technique, we discovered a novel miRNA derived from a primate specific noncoding RNA, the small NF90 associated RNA A (snaR-A). The miRNA derived from snaR-A (miR-snaR) arises independently of Drosha processing but requires Exportin-5 and Dicer for biogenesis. We identify that miR-snaR is concurrently up-regulated with the full snaR-A transcript in cancer cells. Functionally, miR-snaR associates with Ago proteins and targets NME1, a key metastasis inhibitor, contributing to snaR-A’s role in promoting cancer cell migration. Our findings suggest a functional link between a novel miRNA and its precursor noncoding RNA.

Keywords: microRNA biogenesis; snaR-A; NME1; ribonomics; qCLASH

INTRODUCTION

MicroRNAs (miRNAs) are ~22 nucleotide (nt) RNAs that serve an essential regulatory role via targeted silencing and degradation of messenger RNAs (mRNAs) (Bartel 2018). MiRNA targeting is primarily mediated by Watson–Crick base-pairing between the miRNA seed region (nts 2–7) and a complementary target sequence often located within the mRNA 3′ untranslated region (UTR). The pervasive influence of miRNA-driven regulation is observed in nearly all physiological processes, including cell growth, stress responses, metabolism, and apoptosis, with a predicted ≥70% of human mRNAs regulated by miRNAs (Miranda et al. 2006). Given the major role played by miRNAs in normal cellular function, disruption of miRNA expression can lead to significant developmental and physiological defects and miRNA dysregulation has been shown to be a significant driver of carcinogenesis (Peng and Croce 2016; Bartel 2018). As such, both uncovering the mechanisms of miRNA biogenesis and identifying miRNA targets are crucial for the understanding of intrinsic cellular regulatory processes and for designing effective therapeutics of diseases associated with miRNA dysregulation.

Since the initial discovery of miRNAs in C. elegans, the canonical miRNA biogenesis pathway has been well-characterized (Lee et al. 1993; Ha and Kim 2014; Bartel 2018). Canonical miRNA processing begins with transcription of primary miRNA transcripts (pri-miRNAs), most often by RNA polymerase II (Pol II) (Lee et al. 2002), followed by nuclear cleavage via the Microprocessor complex (composed

© 2021 Stribling et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see http://rnajournal.cshlp.org/site/misc/terms.xhtml). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.
of Drosha and DGCR8) (Nguyen et al. 2015). The resulting precursor miRNA (pre-miRNA) hairpins are then exported to the cytoplasm by Exportin 5 (XPO5) in a Ran-GTP-dependent manner (Lund et al. 2004). In the cytoplasm, Dicer cleaves pre-miRNA hairpins producing ~22 nt mature miRNA molecules that tightly associate with a member of the Argonaute (Ago) family of proteins (Bernstein et al. 2001). This Ago-miRNA complex is known as the RNA-induced silencing complex (RISC) and acts to mediate mRNA stability and translation as a part of the posttranscriptional regulatory network (Kawamata and Tomari 2010).

In addition to the canonical pathway, alternative miRNA biogenesis mechanisms bypassing Drosha or Dicer cleavage have been identified (Xie and Steitz 2014). The first identified Drosha-independent miRNAs are mirtrons, which are pre-miRNAs that arise as products of splicing instead of via Drosha cleavage (Ruby et al. 2007; Okamura et al. 2008). Another group of Drosha-independent miRNAs originate from short transcripts derived immediately downstream from Pol II transcription start sites (TSS) (Babiarz et al. 2008; Zamudio et al. 2014). These pre-miRNAs contain a 7-methylguanosine cap at the 5’ terminus and are therefore formed independently of Drosha (Xie et al. 2013; Sheng et al. 2018). Additionally, Drosha-independent miRNAs can be processed from other small noncoding RNAs (ncRNAs), including small nucleolar RNAs (snoRNAs) and transfer RNAs (tRNAs) (Babiarz et al. 2008; Ender et al. 2008; Lemus-Diaz et al. 2020). In the absence of Drosha, these noncanonical miRNAs are up-regulated in cells, presumably because they are free of competition from canonical miRNAs for access to the limited cellular miRNA machinery (Kim et al. 2016; Sheng et al. 2018). Recent studies have demonstrated the increasingly emerging functions of these noncanonical miRNAs in several diseases, including cancer (Stavast and Erkeland 2019).

Although miRNA biogenesis pathways have been well characterized, miRNA target identification remains a major challenge. Recent methodological developments in miRNA sequencing analysis have significantly improved the accuracy of miRNA target prediction on the genomewide scale (Lewis et al. 2005; Agarwal et al. 2015; McGeary et al. 2019; Nazarov and Kreis 2021). However, these predictions rely heavily on conserved base-pairing of the miRNA seed region, which can lead to a high error rate given the high frequency of 6-nt seed-matches in the transcriptome (Pinzon et al. 2017) and the existence of other functional nonseed interactions between miRNAs and their targets (Broughton et al. 2016; Zhang et al. 2018; Chipman and Pasquinelli 2019; McGeary et al. 2019). Biochemical approaches such as high-throughput sequencing of RNAs isolated by crosslinking immunoprecipitation (HITS-CLIP) provide a powerful platform to identify RNA–protein interactions in vivo (Licatalosi et al. 2008; Darnell 2010). In this method, ultraviolet (UV) irradiation is used to covalently crosslink interacting RNA and protein molecules within cells. As such, Ago HITS-CLIP allows identification of bound miRNAs and their associated mRNAs simultaneously but relies on assumption-driven analysis methods for determination of miRNA/target pairs (Chi et al. 2009). HITS-CLIP has inspired a number of related methods for direct investigation of miRNA–mRNA targeting, including: crosslinking, ligation, and sequencing of hybrids (CLASH) (Helwak et al. 2013); quick CLASH (qCLASH) (Gay et al. 2018); and covalent ligation of endogenous Argonaute-bound RNAs with crosslinking and immunoprecipitation (CLEAR-CLIP) (Moore et al. 2015). These methods induce intermolecular miRNA–mRNA ligation on the Ago protein, thus producing miRNA–mRNA hybrid molecules for subsequent high-throughput sequencing. As opposed to traditional CLIP methods in which sequenced miRNAs and targets must then be analytically assigned, the hybrid reads generated by CLASH allow unambiguous pairing between miRNAs and their in vivo targets. While these methods have been effectively used for miRNA target identification in several studies (Helwak et al. 2013; Helwak and Tolleryve 2014; Gay et al. 2018; Sethuraman et al. 2018; Bullard et al. 2019; Ungerleider et al. 2020; Kozar et al. 2021; CJ Fields and M Xie, in prep.), we were interested in determining whether CLASH data could be used for identification of previously unknown miRNA sequences, and specifically for identification of new miRNAs arising via noncanonical biogenesis pathways.

Previous methods developed for analyzing CLASH hybrid sequence data require an annotated database of possible miRNAs and therefore do not permit identification of novel miRNA sequences (Travis et al. 2014). To enable novel miRNA discovery from CLASH data, we developed the target-oriented miRNA discovery (TOMiD) analysis method to identify miRNA sequences within reads generated from CLASH experiments by a stepwise process of hybrid read identification independent of previous miRNA annotation. We applied the TOMiD method to search for Drosha-independent miRNAs in Ago-qCLASH data sets obtained from wild-type (WT), Drosha-knockout (KO), and Dicer-KO HCT116 colorectal cancer cell lines (CJ Fields and M Xie, in prep.). We successfully identified a novel Drosha-independent miRNA: miR-snaR, which arises from the small NF90-associated RNA A (snA-R) family of transcripts (Parrott and Mathews 2007; Kozomara et al. 2019). We found that miR-snaR biogenesis is Drosha-independent, but dependent on XPO5-mediated transport and Dicer processing. In addition, we detected Ago-bound miR-snaR in cells expressing high levels of snA-R, which is up-regulated in multiple cancer-derived cell lines and tumors, including breast, colorectal, liver, and ovarian cancers (Parrott and Mathews 2007; Lee et al. 2016, 2017; Huang et al. 2018; Shi et al. 2019). To
investigate the potential function of this miRNA, we identified miR-snRNA/target hybrids in multiple Ago-CLASH data sets and discovered that miR-snRNA represses expression of nonmetastatic cells 1 (NME1), which correspondingly increases mobility of cancer cells.

RESULTS

TOMiD identifies Drosha-independent miRNAs in Ago-CLASH data sets

Existing methods for Ago-CLASH data analysis rely on previous sequence annotations to identify miRNA/target hybrids (Travis et al. 2014). To identify novel miRNAs within CLASH data, we developed the TOMiD analysis method (Fig. 1). In this approach, sequence reads generated from a CLASH experiment are first aligned to a reference transcript database. Reads that do not align completely to a single reference transcript are then assigned as potential chimeric sequences (hybrids). Potential hybrids are then analyzed via several successive criteria to select for sequences matching characteristics of miRNA/target RNA hybrids specifically. This includes selection based on a specific length of the unaligned portion of the hybrid and on a minimum predicted folding energy of the hybrid via intra-read base-pairing. The unaligned region of each remaining read is then assigned as a candidate miRNA (cmiRNA). In this annotation-naive approach, a cmiRNA can represent either a previously known miRNA, a novel miRNA, or a non-miRNA sequence. To maximize the ratio of identified miRNA sequences to nonspecific sequences, several processing steps are then utilized to increase detection of miRNA-containing hybrids.

We evaluated the sensitivity of this approach in detecting hybrid reads by analyzing a previously published qCLASH data set (Gay et al. 2018) via TOMiD and comparing predicted hybrids to those identified by Hyb, the primary chimeric read caller utilized for analyzing hybrid CLASH data (Travis et al. 2014). This evaluation showed that the TOMiD approach recaptures a large proportion (91.7%) of the hybrids identified by Hyb in this data set (Supplemental Fig. S1), demonstrating a high sensitivity of TOMiD in hybrid detection.

We applied the TOMiD method to detect Drosha-independent miRNAs in Ago-qCLASH data obtained from WT, Drosha-KO, and Dicer-KO HCT116 cells (CJ Fields and MXie, in prep.). Comparison between these data sets provides a unique platform to identify previously uncharacterized Drosha-independent miRNAs, which should be enriched in Drosha-KO cells but depleted in Dicer-KO cells (Kim et al. 2016; Sheng et al. 2018). In total, we analyzed 13 Ago-qCLASH data sets, including five from WT cells, four from Drosha-KO cells, and four from Dicer-KO cells. To reduce identification of results arising due to sequencing error, each predicted unique cmiRNA sequence was required to exist in at least two nonidentical hybrids per sample and be detected in two or more samples. To avoid the potential addition of bias to our analyses by clustering of predicted sequences, each cmiRNA was analyzed independently. This resulted in prediction of many highly similar cmiRNAs that map to the same RefSeq transcript, but which vary by a few nucleotides in length or identity. As the final step in cmiRNA selection, we filtered cmiRNA that did not fall into a length range of 18-23 nt and which did not meet a predicted intrahybrid folding energy \( \Delta G_{fold} \leq -7 \) kcal/mol to specifically select for hybrids with highly miRNA-like characteristics based on miRNA/target sequence complementarity.

Over the 13 evaluated data sets, TOMiD produced 106,021 unique cmiRNA sequences that aligned to 12,600 RefSeq transcripts, with many instances where several nearly identical cmiRNAs were assigned to the same
reference transcript (O’Leary et al. 2016). After initial alignment, several instances were identified in which cmiRNAs aligned to different RefSeq transcript variants arising from the same gene. For accurate quantification of results, in each of these cases a single reference transcript was selected as representative of that respective gene. A per-transcript combined count was then calculated by addition of all unique cmiRNA counts aligning to each respective RefSeq transcript. To account for potential differences in counts per transcript per sample). To encompass Drosha-KO samples, and depleted in the Dicer-KO samples (Kim et al. 2016; Sheng et al. 2018). To encompass all of these characteristics, we developed a scoring metric (scoreDrosha/Dicer) that combines cmiRNA enrichment in Drosha-KO cells and depletion in Dicer-KO cells compared to the average counts in the WT cells (details available in the Supplemental Methods). After hybrid prediction, filtration, and alignment of cmiRNAs to transcripts, we applied this scoring metric to the per-transcript combined cmiRNA counts to identify Drosha-independent cmiRNAs based on a high scoreDrosha/Dicer (range 0 to 501).

To select for high-confidence Drosha-independent miRNA candidates, we chose a threshold of scoreDrosha/Dicer ≥ 20 which identified six high-confidence cmiRNAs for further analysis (Table 1). Of these, five cmiRNA were identified as known Drosha-independent miRNAs: miR-886 (vtRNA-2-1), miR-3615, miR-877, miR-484, and miR-320a (Berezikov et al. 2007; Babiarz et al. 2008; Persson et al. 2009; Xie et al. 2013; Kim et al. 2016). The other top-scoring cmiRNA was not identified as a previously known miRNA, and instead was identified as arising from the small NF90-Associated RNA A (snaR-A) family of miRNAs (Fig. 2A,B). The snaR-A transcripts are ~120 nt, highly structured RNAs that interact with NF90 via a double-stranded RNA binding motif. The human genome contains 14 nearly identical copies of snaR-A located on chromosome 19 (Parrott et al. 2011), which are transcribed by RNA polymerase III (Pol III) by an intragenic promoter, with sequences of different members of the snaR-A family varying at only two nucleotide positions (Fig. 2A; Parrott and Mathews 2007; Parrott et al. 2011). While a previous bioinformatic study suggested that a Dicer-dependent small RNA may derive from snaR-A, it did not identify any relationship to Ago that might indicate miRNA function (Langenberger et al. 2013).

Characterization of the snaR-A sequence

The potential miRNA derived from snaR-A, which we describe as miR-snaR, is located at the base of the snaR-A hairpin, followed by a ~15 nt 3′ tail (Fig. 2A). To elucidate miR-snaR processing, we examined the endogenous snaR-A sequences as both 5′ and 3′ termini of a pre-miRNA are crucial elements for mature miRNA production (Ma et al. 2004; Park et al. 2011). We adapted rapid amplification of cDNA ends (RACE) to determine the 5′ and 3′ ends of the snaR-A endogenously expressed in human embryonic kidney (HEK) 293T cells. Sanger sequencing results reveal a stretch of at least five uridines (U) at the 3′ end of snaR-A, which acts as an efficient termination signal for Pol III transcription (Supplemental Fig. S2A). However, mixed nucleotide signals were detected from the 5′ end of the snaR-A (Supplemental Fig. S2B). We then cloned individual sequences from 5′ RACE products and found that snaR-A contain heterogeneous sequences in the first 2 nt at the 5′ end (Fig. 2A; Supplemental Fig. S2C). These findings are mostly consistent with previous reports, which also found heterogeneous 5′ di-nucleotide, but indicated that as many as 8 Us could be present at the 3′ end of snaR-A (Parrott and Mathews 2007).

To further investigate the length of endogenous snaR-A transcripts, we in vitro transcribed three different versions

| # | RefSeq transcript | Consensus cmiRNA nts. | Gene/miRNA name | CPM WT | CPM Drosha-KO | CPM Dicer-KO | ScoreDrosha/Dicer |
|---|------------------|------------------------|-----------------|--------|--------------|--------------|-----------------|
| 1 | NR_030583.3      | 1–23                   | vtRNA-2-1/miR-886* | 24.1   | 97.9         | 0.0          | 501.4           |
| 2 | NR_037409.1      | 51–73                  | miR-3615        | 10.4   | 36.5         | 0.0          | 318.7           |
| 3 | NR_030615.1      | 1–22                   | miR-877         | 57.6   | 294.8        | 1.0          | 281.2           |
| 4 | NR_030159.1      | 8–28                   | miR-484        | 50.5   | 149.8        | 0.9          | 181.4           |
| 5 | NR_004351.5      | 84–105                 | snaR-A         | 37.7   | 738.7        | 5.9          | 132.2           |
| 6 | NR_029714.1      | 48–69                  | miR-320a       | 190.7   | 3803.9       | 86.8         | 46.5            |

*miR-snaR processing, we examined the endogenous snaR-A sequences as both 5′ and 3′ termini of a pre-miRNA are crucial elements for mature miRNA production (Ma et al. 2004; Park et al. 2011). We adapted rapid amplification of cDNA ends (RACE) to determine the 5′ and 3′ ends of the snaR-A endogenously expressed in human embryonic kidney (HEK) 293T cells. Sanger sequencing results reveal a stretch of at least five uridines (U) at the 3′ end of snaR-A, which acts as an efficient termination signal for Pol III transcription (Supplemental Fig. S2A). However, mixed nucleotide signals were detected from the 5′ end of the snaR-A (Supplemental Fig. S2B). We then cloned individual sequences from 5′ RACE products and found that snaR-A contain heterogeneous sequences in the first 2 nt at the 5′ end (Fig. 2A; Supplemental Fig. S2C). These findings are mostly consistent with previous reports, which also found heterogeneous 5′ di-nucleotide, but indicated that as many as 8 Us could be present at the 3′ end of snaR-A (Parrott and Mathews 2007).
of sna-R-A, including either five or eight consecutive Us as the 3’ terminal nucleotides (5U and 8U), as well as a truncated sna-R-A (Δtail) whose 3’ end coincides with the potential miR-snaR. We reasoned that a truncated sna-R-A is more likely to be the direct substrate for Dicer, because a 3’ tailed hairpin is not a typical Dicer substrate (Ma et al. 2004). We analyzed these three transcripts alongside with total RNA extracted from 293T cells on a near-infrared northern blot (Fig. 2C; Miller et al. 2018; Fields et al. 2019). The major population of sna-R-A from 293T cells migrate at the same position as the sna-R-A containing 5Us (Fig. 2C, compare lanes 1 and 3). Therefore, both our RACE and northern blot data suggest that the majority of endogenous sna-R-A transcripts have 5 Us at the 3’ end.

miR-snaR biogenesis is independent of Drosha, but depends on XPO5 and Dicer

As mentioned before, the full length sna-R-A is unlikely to be processed directly by Dicer due to the long 3’ tail, while sna-R-A Δtail resembles a traditional pre-miRNA with a short 3’ overhang. Indeed, when we subject both sna-R-A full length and Δtail to FLAG-tagged Dicer purified from 293T cells, only the sna-R-A Δtail was successfully processed into miR-snaR (Fig. 2D, compare lanes 3 and 5). However, no such sna-R-A Δtail intermediate is abundantly detected in the total RNA sample (Fig. 2C, compare lanes 1 and 2; Fig. 2D, compare lanes 1 and 2). It is also unclear how the tail region is removed from the full length sna-R-A for subsequent Dicer processing. Nonetheless, a hairpin without the 3’ tail would be compatible with both XPO5-mediated export and subsequent Dicer cleavage, and our Dicer processing assays validate that the potential miR-snaR can indeed originate from sna-R-A.

Next, we explored the biogenesis of miR-snaR in HCT116 cells, in which sna-R-A is not highly expressed (Supplemental Fig. S3A). We cloned the endogenous sna-R-A gene into the pBlueScript plasmid and transfected the plasmid into HCT116 cells to examine miR-snaR expression. In three additional HCT116 cell lines we used for transfection, essential genes for canonical miRNA biogenesis were knocked out by CRISPR-Cas9, including Drosha, Dicer, and XPO5 (Kim et al. 2016). Expression of miR-snaR in these HCT116 cells was analyzed by northern blot after Ago-immunoprecipitation (IP). MiR-snaR is clearly detectable in the WT HCT116 cells (Fig. 3A, lanes 1–3), again confirming that miR-snaR can be generated from the full-length sna-R-A transcript. Compared to WT HCT116 cells, knockout of Drosha resulted in an obvious increase of miR-snaR (Fig. 3A, compare lanes 3 and 6). In contrast, miR-snaR could not be detected in Dicer-KO cells, and is detectable at a lower level in XPO5-KO cells (Fig. 3A, lanes 9 and 12). In the small RNA-seq data from these HCT116 cell lines, we also confirmed that the expression levels of endogenous miR-snaR are the highest in Drosha-KO cells, but diminished in Dicer-KO and XPO5-KO cells (Supplemental Fig. S3B; Kim et al. 2016). Combined with results from the Dicer processing assay, we conclude that miR-snaR is a noncanonical miRNA dependent on Dicer and XPO5 but independent of Drosha.

Detection of miR-snaR and its repressive function in cells

To examine if functional miR-snaR is expressed endogenously in cells, we performed Ago IP in 293T and breast cancer cell lines that are known to express sna-R-A (Parrott and Mathews 2007; Lee et al. 2016, 2017), which may lead to a higher endogenous level of miR-snaR. RNAs coimmunoprecipitated with Ago were analyzed by northern blot (Fig. 3B). While miR-snaR was not readily detectable in 293T cells, knocking out Drosha dramatically enriched miR-snaR (Fig. 3B, compare lanes 3 and 6), consistent with the results of our previous analysis (Table 1; Fig. 3A). In HER2 positive breast cancer MCF-7 cells, endogenous miR-snaR are detectable after Ago-IP (Fig. 3B, lane 9). Interestingly, the levels of miR-snaR and sna-R-A in MCF-7 and 293T cells do not correlate, indicating that cell-specific factors contribute to the efficiency of miR-snaR biogenesis (Fig. 3B, compare lanes 1–3 to lanes 7–9). In triple negative breast cancer MDA-MB-231 cells, neither sna-R-A nor miR-snaR was detected by northern blot (Fig. 3B, lanes 10–12), contradicting a previous report of high sna-R-A expression in this cell line (Lee et al. 2017). In conclusion, we confirm that miR-snaR is endogenously expressed in cell lines that express sna-R-A.

To determine if mature miR-snaR can function in RNA interference, we performed luciferase reporter assays in both WT and Drosha-KO 293T cells. Cells were transfected with either a plasmid expressing full-length sna-R-A or an empty vector. Additionally, cells were cotransfected with a luciferase reporter plasmid with a miR-snaR binding site consisting of two repeats of a sequence that is fully complementary to miR-snaR in the 3’UTR of the luciferase gene. Luciferase signal was then measured 48 h after transfection. In both WT and Drosha-KO cells the luciferase signal in groups expressing sna-R-A was reduced ~50%–60% compared to their respective controls (Fig. 3C). Of note, the luciferase signal in the Drosha-KO control was lower than that in the WT control. As a Drosha-independent miRNA, endogenous miR-snaR, as well as other Drosha-independent miRNAs, is expected to be elevated in Drosha-KO cells (Table 1; Fig. 3A), which accounts for the lower luciferase signal. The result of this assay confirms that miR-snaR is processed from the full-length sna-R-A and that the resulting miR-snaR is biologically active. These results fully complement and support the results of previous experiments examining the biogenesis and Ago-interaction of miR-snaR (Fig. 3A,B).
Identification of miR-snaR targets in Ago-CLASH data sets

A number of previous studies have revealed snaR-A's involvement in metabolic dysregulation of breast, liver, and ovarian cancers (Lee et al. 2016, 2017; Huang et al. 2018; Shi et al. 2019) and in drug resistance in human colon cancer cells (Lee et al. 2014). We questioned whether miR-snaR can contribute to the phenotypes associated with snaR-A. To investigate this, mRNA targets of miR-snaR were identified from multiple Ago-CLASH and qCLASH data sets obtained from HCT116 WT and Drosha-KO cells, as well as from HEK293 cells (Helwak et al. 2013), by the Hyb software package using a custom reference database containing miR-snaR.

We first used UNAFold to analyze the pattern of miR-snaR/target interaction (Markham and Zuker 2008; Travis et al. 2014). The hybkit software package (version 0.2.1a) developed by the Renne laboratory was utilized to select hybrids containing mRNA and mRNA-pseudogenes and the frequency of pairing interaction at each nucleotide position of miR-snaR was calculated (Fig. 4A, left graph;
Previous studies utilizing Ago-CLASH have observed hybrids with a higher contribution of significant or exclusive nonseed miRNA/target interactions than expected based on canonical understanding of seed-focused miRNA targeting (Helwak et al. 2013; Helwak and Tollervey 2014; Gay et al. 2018; McGeary et al. 2019). The observed targeting interactions of miR-snaR conversely display primarily seed-focused pairing. Within miR-snaR hybrids, an average of 84.0% of miR-snaR interactions include base paring of the seed region (nts 2–8). However, a region of significant supplemental base-pairing is also observed toward the 3′ end of miR-snaR (nts 15–17, average 88.2%) (Fig. 4A, left graph). When considering the entire miRNA/target data set (excluding miR-snaR), a relatively even distribution of target interaction across the complete miRNA length was observed (Fig. 4A, right graph), consistent with previous observations (Gay et al. 2018). The seed-focused interactions of miR-snaR suggest a potentially high potency in mediating target gene repression that is observed with canonical seed-focused miRNA/target interactions (Bartel 2018). We next examined specific miR-snaR targets identified in Ago-CLASH data sets in detail. Based on miR-snaR’s predominantly seed-focused binding, targets were filtered to only mRNA targets in which a full seed match was present and in which two nonidentical hybrids were identified. This resulted in 63 identified targets across the three data sets. Targets were then selected where they were identified in at least two data sets, which provided four high-confidence targets of miR-snaR for detailed study (Fig. 4B). Three of these targets, KRTCAP2, MIDN, and TNFRSF12A, appear in two data sets. The target site in KRTCAP2 falls...
within the coding sequence, an atypical location for miRNA binding (Fig. 4C). For MIDN and TNFRSF12A, the target sites are complementary with the miR-snaR seed, but MIDN target site contains G–U wobble base-pairing (Fig. 4C). Among the four high-confidence targets, nucleoside diphosphate kinase 1/nonmetastatic cells 1 (NME1) was the only target to appear in all three data sets (Fig. 4B). Additionally, the putative miR-snaR binding site located within the 3′-UTR of NME1 has the characteristics of a canonical miRNA binding site. This site is located just ∼50 nt downstream from the stop codon. In addition, NME1 includes a target sequence that is fully complementary to the miR-snaR seed region and which is also conserved among the higher-order primate species in which the snaR-A gene exists (Fig. 4C; Supplemental Fig. S4; Parrott et al. 2011). Together, these details strongly suggest NME1 as a genuine target of miR-snaR.

Interestingly, NME1 was the first human gene to be identified as a suppressor of cancer metastasis (Steeg et al. 1988). In breast cancer, an inverse correlation between NME1 and metastatic potential has been well documented (Leone et al. 1993; Horak et al. 2007). Studies have also shown that breast cancer cells secrete NME1, and this extracellular NME1 supports primary tumor growth and metastasis (Yokdang et al. 2015). Given that snaR-A promotes proliferation and migration of breast cancer cells (Lee et al. 2017), miR-snaR may cooperate with snaR-A by repressing NME1 expression.

**miR-snaR reduces NME1 mRNA levels via a specific binding site in the 3′-UTR**

To examine whether miR-snaR can negatively impact NME1 mRNA levels, we performed reverse transcription followed by real time PCR (RT-qPCR) after overexpressing miR-snaR in 293T cells where miR-snaR is not highly expressed (Fig. 3B). Introduction of miR-snaR was performed either by plasmid transfection or by transfection of a synthetic miRNA mimic (Fig. 5A,B). This experimental setup allows identification miR-snaR function independent of the full length snaR-A transcript, which may influence gene expression through a different mechanism. The plasmid was designed to express a U6 promoter-driven short hairpin RNA (shRNA) to be subsequently cleaved by Dicer and produce miR-snaR (Fig. 5A). As a control, a plasmid expressing a shRNA with a scramble sequence of miR-snaR hairpin was constructed.

Northern blot confirmed the expression of miR-snaR from the utilized plasmid (Fig. 5A, northern blot). In both miR-snaR mimic and plasmid transfection experiments, NME1 mRNA is significantly reduced in cells concurrent to elevated levels of miR-snaR (Fig. 5C). The results of this experiment provide evidence that the NME1 mRNA can indeed be regulated by miR-snaR in cells.

Binding site specificity is critical to the regulatory function of miRNAs. To investigate whether the miR-snaR binding site on NME1 is the sequence captured by Ago-qCLASH (Fig. 4B), we performed a dual luciferase reporter assay, in which the firefly luciferase reporter contained the 3′-UTR of NME1 (Fig. 5D). As a control, we also constructed a luciferase reporter plasmid with the miR-snaR binding site in NME1 3′-UTR mutated to base-pair with miR-HSUR4 (mut). This miR-HSUR4 control was chosen due to its viral origin (Herpesvirus saimiri) and its lack of endogenous expression in 293T cells. A luciferase assay was conducted using both the previously described plasmid and mimic transfection methods for miR-snaR expression (Fig. 5A,B). In both scenarios, miR-snaR significantly decreased luciferase expression when coexpressed with WT NME1 3′-UTR reporter, but not when coexpressed with a mut NME1 3′-UTR reporter (Fig. 5E). In contrast, luciferase signal in cells transfected with mut NME1 3′-UTR reporter was
miR-snaR down-regulates NME1 protein to promote cancer cell migration

After confirming that miR-snaR reduces NME1 mRNA within cells, we then analyzed NME1 protein expression. We transfected two cell lines—293T and MDA-MB-231—with miR-snaR mimic, as these cells have low levels of endogenous miR-snaR expression (Fig. 3B, lanes 3 and 12). Additionally, MCF-7 cells, in which endogenous miR-snaR is detectable, were transfected with miR-snaR inhibitor to reverse potential miR-snaR-mediated repression on NME1. In each case, total protein lysate was collected 48 h after transfection and protein levels were determined by western blot. In both 293T and MDA-MB-231 cells, miR-snaR transfection resulted in a significant decrease in NME1 protein levels (Fig. 6A, lanes 1–4; Fig. 6B). On the other hand, inhibition of miR-snaR is observed to increase NME1 level in MCF-7 cells, albeit by a rather moderate amount (Fig. 6A, compare lanes 5 and 6; Fig. 6B). We also investigated NME1 protein levels in HCT116 cells, in which miR-snaR is up-regulated in Drosha-KO cells but diminishes in Dicer-KO cells (Supplemental Fig. S3B). Supporting targeting of NME1 by miR-snaR, we observed an inverse correlation between NME1 and miR-snaR levels, with low levels of NME1 in Drosha-KO cells and high levels of NME1 in Dicer-KO cells, respectively (Fig. 6C, lanes 2 and 3). Importantly, introduction of miR-snaR inhibitor in Drosha-KO cells elevated NME1 (Fig. 6C, compare lanes 4 and 5), and introduction of miR-snaR mimic in Dicer-
KO cells repressed NME1 (Fig. 6C, compare lanes 6 and 7). These results suggest that NME1 protein levels are directly influenced by miR-snaR.

Previous studies have shown that knockdown of snaR-A could inhibit proliferation and migration of breast cancer cells (Lee et al. 2016, 2017). Given this information, we looked further into the effect of miR-snaR in breast cancer cell lines. We utilized the MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay to assess the effects of miR-snaR on cell proliferation. Considering that snaR-transcript knockdown inhibits proliferation (Lee et al. 2016, 2017), we speculated that increasing miR-snaR expression in MDA-MB-231 or MCF-7 cells would result in increased proliferative activity. To this end, MDA-MB-231 and MCF-7 cells were transfected with miR-snaR mimic. Additionally, 293T cells were also transfected. MTT assays were performed on cells 24 h post transfection and immunoblotting was used to show reduction in NME1 protein as a control mimic. White lines outline the edge of the scratched wound. Scale bar: 100 µm. (E) Quantification of three experiments as shown in D. (Student's t-test, [*] P ≤ 0.05). (F) Model of snaR-A’s role in enhancing tumor growth and metastasis. Derived from snaR-A, miR-snaR is inhibiting NME1 to promote metastasis.
healing assay using MCF-7 cells transfected with either a control or a miR-snaR mimic. Consistent with our initial result, miR-snaR transfection in MCF-7 cells significantly reduced NME1 protein levels (Supplemental Fig. S5B). We also observed a significant increase in wound closure for miR-snaR-transfected cells over the course of 48 h (Fig. 6D,E). Therefore, our data suggest that miR-snaR can promote cell migration by repression of NME1. Given the role of NME1 as a metastasis suppressor, our findings suggest an oncogenic role for miR-snaR in cancer.

**DISCUSSION**

In this study, we developed the TOMiD method, an innovative bioinformatic approach for identification of novel miRNAs using the Ago (q)CLASH ribonomics technique (Fig. 1). Utilizing qCLASH data sets obtained from a panel of HCT116 cells lacking essential miRNA biogenesis factors (Drosha or Dicer), we discovered a miRNA (miR-snaR) that originates from snaR-A, a small noncoding RNA implicated in tumor proliferation and metastasis (Fig. 2). Despite being independent of Drosha processing, miR-snaR biogenesis is XPO5- and Dicer-dependent (Figs. 2, 3). We showed that miR-snaR is biologically active due to its association with Ago and its ability to repress luciferase-reporter-containing complementary target sites (Fig. 3). Functionally, miR-snaR inhibits NME1 expression both at the mRNA and the protein levels (Figs. 4–6). Because NME1 is a metastasis inhibitor, we propose that snaR-A exerts its metastasis-promoting function at least in part through miR-snaR targeting of NME1 (Fig. 6F).

**TOMiD analysis for de novo miRNA discovery**

The TOMiD approach expands the scope of Ago-(q) CLASH by enabling discovery of novel miRNAs in addition to the primary use of this technique for unambiguous identification of miRNA targeting interactions (Helwak et al. 2013; Gay et al. 2018). As demonstrated by the identification of miR-snaR, this methodology has a significant potential for use in characterization of previously unidentified noncanonical human miRNAs. In addition, TOMiD expands the ability of (q)CLASH to enable de novo miRNA characterization in nonmodel organisms or miRNA-producing pathogens, as it does not rely on preexisting annotations of miRNAs and requires only a reference transcriptome for miRNA discovery.

The TOMiD method also has a significant potential for further refinement and development to increase specificity in identification of miRNAs. In the current analysis, a low “signal-to-noise” ratio was observed with many non-miRNA hybrids predicted for each valid miRNA interaction, albeit with a characteristically lower count for occurrences in nonspecific predictions. While differentiation of potentially important novel or previously known miRNAs can be made based on hybrid counts, performing a detailed characterization of the cutoff parameters utilized for filtration at different stages in the method and the addition of further data preprocessing steps each have the potential to significantly increase the specificity of miRNA-containing hybrid prediction and thus increase utility in novel miRNA discovery. Additionally, this method has the potential for future modification to utilize genomic, rather than transcriptomic reference sequences, allowing characterization of small sequence fragments that may escape detection as valid transcripts and facilitating the study of nonmodel organisms without a robust reference transcriptome. This method is currently under further development and is being implemented as a publicly available analysis pipeline to be described elsewhere.

**MiR-snaR biogenesis and function**

Previously, a small RNA sequence derived from the snaR-A transcript in a Dicer-dependent fashion was identified via a retrospective data analysis (Langenberger et al. 2013). However, whether this sequence could be classified as a miRNA was debatable. It has been shown that snaR-A is predominantly distributed in the cytoplasm of 293T and HeLa cells (Parrott et al. 2011). Yet nuclear snaR-A is more likely to be the source of miR-snaR, because the truncated snaR-A, which serves as the precursor, is presumably produced in the nucleus and then exported by XPO5 (Figs. 2D, 3A). Our discovery that miR-snaR associates with Ago and the ability of miR-snaR to suppress NME1 argues that this sequence is a bona-fide miRNA. SnaR-A is a noncoding RNA that is most highly expressed in testis, followed by brain and placenta, and is specific to higher-order primates (Parrott and Mathews 2007). Additionally, the full snaR-A transcript is up-regulated in many cancer cell lines (Parrott and Mathews 2007; Lee et al. 2016; Huang et al. 2018; Shi et al. 2019). While it would seem reasonable for miR-snaR expression to follow the expression pattern of snaR-A, we have observed instances of disproportional snaR-A and miR-snaR expression in different cell lines (Fig. 3B), suggesting that cell-specific factors can control the processing from snaR-A to miR-snaR. One such factor could be the enzyme(s) that removes the tail of snaR-A, before Dicer could process the hairpin (Fig. 2A,D). The identity of this enzyme(s) requires further investigation. MiR-snaR can potentially arise by processing of any of the fourteen members of the snaR-A family, as each homolog has an identical sequence in the miR-snaR region (Fig. 2A). This abundant genetic source may explain the large degree of miR-snaR up-regulation in Drosha-KO cells (Table 1; Supplemental Fig. S3B). In light of this, miR-snaR levels may be particularly sensitive to perturbation of Drosha levels. Indeed, in HCT116 Drosha-KO cells, miR-snaR rises to the second most abundant miRNA (Supplemental Fig. S3B).
and data not shown). Therefore, miR-snaR may be most influential in cells with low Drosha expression.

The discovery of miR-snaR adds to the ever-growing list of miRNAs that are derived from other small noncoding RNAs (Babiarz et al. 2008; Ender et al. 2008; Bartel 2018; Lemus-Diaz et al. 2020). Compared to canonical pre-miRNAs that are largely consumed to produce mature miRNAs, small noncoding RNAs are less efficiently processed into miRNAs, presumably because they are shielded by specific RNA-binding proteins. Also, small noncoding RNAs usually deviate from their normal conformation to adopt a hairpin that may not be optimal for Dicer processing (Babiarz et al. 2008; Lemus-Diaz et al. 2020). For example, Dicer cleavage efficiency of snaR-A hairpin is considerably lower compared to the pre-let-7a hairpin (Fig. 2D, compare lanes 3–7). Consequently, miRNAs originating from small noncoding RNAs are generally less abundant compared to most canonical miRNAs. Nonetheless, miR-snaR occurs in greater abundance than its peers as discussed above. These observations suggest that miR-snaR plays a substantial role in gene regulation.

In miR-snaR/target RNA hybrids obtained from Ago-(q) CLASH data sets, we observed seed-focused interactions (Fig. 4A, left) suggesting that miR-snaR is predominantly involved in canonical seed-match target-regulation. Based on identification of four miR-snaR targets found in more than two qCLASH data sets each, we predicted four genes (NME1, KRTCAP2, MIDN, and TNFRSF12A) as high-confidence targets of miR-snaR. Among these, miR-snaR targeting of NME1 is particularly interesting. First, the miR-snaR target site in NME1 is located only 56 nt downstream from the stop codon, a preferred site for miRNA-mediated repression in the 3'-UTR (Grimson et al. 2007), with absolute conservation among the higher-order primate species in which the snaR-A gene exists (Supplemental Fig. 54). The combination of the functional target location with this conservation suggests a functional importance for this targeting. Second, NME1 has been shown both to regulate expression of genes associated with metastasis and to limit cell motility through interactions with proteins such as Gelsolin (Horak et al. 2007; Marino et al. 2013; McCorkle et al. 2014). Combined with the role of NME1 as a well-known metastasis inhibitor and the demonstrated role of snaR-A in the promotion of metastasis in a variety of cancers, our identification of NME1 targeting by miR-snaR suggests that these two ncRNAs may play complementary roles in promoting cancer cell growth and/or migration. Furthermore, other than miR-snaR, two potential PIWI-interacting (pi)RNAs can derive from snaR-A (Parrott and Mathews 2007). Given that the highest expression of snaR-A occurs in the testis, the existence of potential piRNA fragments derived from snaR-A is intriguing. Additionally, one of the piRNAs, the 27 nt long piR-30611, covers the entire miR-snaR sequence (Parrott and Mathews 2007). Whether and how these piRNAs may coordinate with miR-snaR and snaR-A in gene regulation warrants further investigation.

**MATERIALS AND METHODS**

**Bioinformatic analysis**

To identify novel miRNAs, the TOMID analysis method was implemented as a series of stepwise analyses (see Supplemental Methods for full implementation details). Sequence reads from Ago-qCLASH experiments (GEO accession number: GSE164634) in HCT116 cells were first analyzed for sequence quality and length utilizing FastQC (Andreas 2010). Sequencing adapters were then removed by Primomatic, and Illumina read mate pairs were merged using PEAR (Bolger et al. 2014; Zhang et al. 2014). PCR-duplicates were then removed using fastx-collapser, followed by removal of individual read barcodes by cutadapt (https://hannonlab.cshl.edu/fastx_toolkit; Martin 2011). Reads were then split into two portions, with the longer alignment selected as the “target” and the shorter portion as the candidate miRNA (cmiRNA). Hybrids were then retained where the cmiRNA was 18 to 23 bases in length and was represented in two or more candidate hybrids occurring in at least two independent Ago-qCLASH sequencing libraries.

For identification of possible Drosha-independent miRNAs, the counts of each cmiRNA were compared in Ago-qCLASH data sets across WT, Drosha-KO, and Dicer-KO HCT116 cells. A scoring metric was applied to stratify cmiRNAs significantly enriched in Drosha-KO cells and depleted in Dicer-KO cells, and cmiRNAs with scoreDrosha/Dicer≥20 were identified as high-confidence Drosha-independent cmiRNAs. For identification of miR-snaR targets, a reference database containing miR-snaR was prepared and the Hyb pipeline was utilized to identify CLASH hybrids containing miR-snaR with its associated targets (Travis et al. 2014). The Vienna hybrid pairing representations produced by the UNAFold portion of the Hyb pipeline were then analyzed to plot the overall and miR-snaR-specific patterns of miRNA positional base-pairing frequency utilizing the Hybkit toolkit (https://github.com/RenneLab/hybkit). A detailed description of analysis methods, as well as the custom analysis scripts utilized, have been supplied in the Supplemental Materials (Harower and Brewer 2003; Chen and Boutros 2011; R Core Team 2012).

**Plasmid construction**

For expression of the snaR-A transcript, 555 bp of genomic sequence encompassing the snaR-A sequence was cloned into
In vitro transcription and Dicer processing assays

Polymerase chain reaction (PCR) DNA templates for snr-A transcription were amplified from pBS-snr-A with a forward primer containing the T7 promoter sequence and various reverse primers to generate different versions of snr-A transcripts. These included a truncated snr-A (Δtail), as well as 5 or 8 consecutive Us (SU and BU) as the 3′ termination nucleotides. The 5′ variable di-nucleotides of the transcripts are set to GC to allow efficient T7 transcription. Utilized primers are listed in Supplemental Table S1. The DNA products were separated on a 3% agarose gel and then extracted using gel extraction kit (ZYMO research). In vitro transcription with T7 RNA polymerase and in vitro Dicer cleavage assays were performed as previously described (Sheng et al. 2018).

Cell culture, transfection, total RNA, and protein extraction

HEK293T, breast cancer MCF-7, and breast cancer MDA-MB-231 cells were cultured in Dulbecco’s modified Eagle medium (DMEM, HyClone SH30243.FS) with 10% fetal bovine serum (FBS, Gibco) and 1% penicillin/streptomycin (P/S) at 37°C with 5% CO₂. HCT116 colorectal cancer cells were cultured in McCoy’s 5A medium (HyClone, SH30200.FS) with 10% FBS, 1% P/S at 37°C, with 5% CO₂.

Plasmid DNA was transfected using polyethylenimine (PEI) (Polysciences Inc, 24765-1) or Lipofectamine 3000 (Life Tech, L3000015) in 293T, MCF-7, and MDA-MB-231 cells in suspension, according to the manufacturer’s instructions.

Total RNA was isolated using TRIzol reagent (Life Tech, 15596018) and treated with RQ1 DNase (Promega, M6101) 48 h after transfection, according to manufacturer’s instructions. After DNase treatment, RNA was purified using phenol-chloroform-isoamyl alcohol (25:24:1) (PCA) followed by ethanol precipitation. To collect total protein, ∼200 µL of NP-40 lysis buffer (50 mM Tris–HCl pH 7.5, 1% NP-40, 10% glycerol, 150 mM NaCl, 5 mM EDTA, and 0.5 mM PMFS) was added per million cells. Cell lysate were mixed with gentle agitation for 30 min at 4°C and supernatant was collected by centrifugation at 21,000g for 10 min at 4°C.

5′ and 3′ RACE of the endogenous snr-A

Total RNA was extracted from 293T WT and Drosha-KO cells. The 3′ terminus of total RNAs were ligated with the irCLIP adaptor by T4 RNA Ligase 1 (NEB, M0437M). After ligation, the RNAs were annealed with a cDNA synthesis primer and cDNAs were generated by SuperScript IV Reverse Transcriptase (Invitrogen, 18090050). The cDNA/RNA duplexes were then purified by streptavidin beads (Invitrogen, 65601) and the RNAs hydrolyzed by NaOH. The cDNA was circularized by CircLigase II ssDNA Ligase (Epícentre, CL925K). Next, cDNA of snr-A was amplified from circular single-stranded DNA templates by PCR using primers (P6 tall, P3 tall) that anneal to the irCLIP adaptor and the gene specific reverse primers (snr-A 5′, snr-A 3′). The PCR amplicons were purified by Zymoclean Gel DNA Recovery Kit (ZYMO RESEARCH, D4008) and phosphorylated at the 5′ termini by T4 Polynucleotide Kinase (NEB, M0201L). The pBluescript plasmid was cut by EcoRV and dephosphorylated by calf intestinal alkaline phosphatase (CIP) (Promega, E1330). Two oligonucleotides were annealed to form two repeats of a sequence that is reverse complementary of the mir-snr-R sequence, which was then inserted into pmirGLO via XbaI and SacI restriction enzyme sites. To confirm the mir-snr-R binding site in the 3′-UTR of NME1 mRNA, pmirGLO-NME1 and pmirGLO-NME1-mut were generated on pmirGLO vector. To construct pmirGLO-NME1, cDNA of the last exon of the NME1 mRNA, pmirGLO-NME1 was constructed by replacing miR-snr-A shRNA with a scrambled sequence. Two repeats of a sequence that is reverse complementary of the pmirGLO-NME1-mut (Fig. S5). All primers utilized in cloning are listed in Supplemental Table S1.

Immunoprecipitation, western blot, and antibodies

Ago-IP was performed using α-Ago antibody (clone 4F9) as previously described (Sheng et al. 2018). Antibodies for NME1 (ABClonal, A0259), Hsc70 (Santa Cruz, sc-7298), and GAPDH (Cell Signaling, 14C10) were used for western blots.

RNA quantification

Northern blot analyses with near infrared dye-labeled probes were performed as previously described (Miller et al. 2018; Fields et al. 2019). The probes utilized are listed in Supplemental Table S1. For RT-qPCR, 1 µg of total RNA was reverse transcribed to generate cDNA using iScript RT Supermix (Bio-Rad, 1708841). The reaction was completed using the following protocol: 5 min at 25°C (priming), 20 min at 46°C (reverse transcription), 1 min at 95°C (RT inactivation). Resultant cDNA was used for quantitative PCR (qPCR) using SYBR GreenER qPCR SuperMix (Bio-Rad, 1725275) using the standard qPCR protocol after initial denaturing step at 95°C for 30 sec, repeating 40 cycles (95°C 15 sec, 60°C 30 sec) followed by melting curve analysis.
NME1 mRNA levels were normalized against β-actin mRNA levels using $2^{-\Delta \Delta CT}$. Primers utilized are listed in Supplemental Table S1.

**Luciferase assay**

HEK293T cells were transfected with different pairs of pmirGLO reporter and miRNA-expressing plasmids or miRNA mimic. Dual-Luciferase reporter assays were carried out 48 h post transfection according to the manufacturer’s instructions. Briefly, cells were lysed with 1 x passive lysis buffer by gently shaking at room temperature for 15 min. Then, 20μL lysate was added to 100μL of Luciferase Assay Buffer II (LARII) and firefly luciferase activity was measured immediately. Finally, 100μL Stop&Go reagent was added and Renilla luciferase activity was measured.

**Cell proliferation and migration assays**

HEK293T, MDA-MB-231, and MCF-7 cells were seeded at a density of 0.3 x 10^5 cells/well and transfected with synthetic miR-snaR or miR-snaR inhibitor or the appropriate control. To measure cell proliferation, the MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium) assay was performed at 24 h post transfection. The culture medium of each cell culture was replaced by 500 ng/µL of Luciferase Assay Buffer II (LARII) and firefly luciferase activity was measured immediately. To measure cell migration, a confluent monolayer of cells was scraped and the wound healing size tool plugin for ImageJ was used to quantify wound healing at each timepoint. Images were captured at 0, 24, and 48 h after scraping. The wound healing size tool plugin for ImageJ was used to quantify wound healing at each timepoint.

**DATA DEPOSITION**

Ago-qCLASH data from Illumina high-throughput sequencing has been deposited in the Gene Expression Omnibus under accession number GSE164634.

**SUPPLEMENTAL MATERIAL**

Supplemental material is available for this article.

**ACKNOWLEDGMENTS**

We thank Drs. Susan Frost, Shuo Gu, Narry V. Kim, Brian Law, and Jianrong Lu for sharing reagents; J. Bert Flanegan for critical reading of the manuscript; UF Interdisciplinary Center for Biotechnology Research and the UF Health Cancer Center Sequencing Core for sequencing services. Figure 1 was created with Biorender.com. This work was supported by grants from National Institutes of Health (R00-CA190886 and R35-GM128753 to M.X.; P01-CA214091 to R.R.; R01-CA188561 to R.O.), University of Florida Research Office (Research Opportunity Seed Fund to M.X.), and the University of Florida Informatics Institute (UFII Graduate Fellowship to D.S.).

**Author contributions:** D.S., R.R., and M.X. conceived the project. D.S. developed TOMiD, analyzed miR-snaR targeting, and small RNA-Seq data. Y.L. characterized miR-snaR biogenesis. C. M.G. investigated miR-snaR targeting of NME1. L.L., C.J.F., and Y.L. predicted miR-snaR targets. C.M.G. and P.N. performed NME1 western blots. P.N. and R.O. measured miR-snaR's influence on cell proliferation. D.S., Y.L., C.M.G., R.R., and M.X. wrote the manuscript.

**REFERENCES**

Agarwal V, Bell GW, Nam JW, Bartel DP. 2015. Predicting effective microRNA target sites in mammalian miRNAs. Elife 4: e05005. doi:10.7554/eLife.05005

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics. Babraham Institute, Cambridge, United Kingdom.

Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R. 2008. Mouse ES cells express endogenous shRNAs, siRNAs, and other microprocessor-independent, dicer-dependent small RNAs. Genes Dev 22: 2773–2785. doi:10.1101/gad.1705308

Bartel DP. 2018. Metazoan microRNAs. Cell 173: 20–51. doi:10.1016/j.cell.2018.03.006

Bemis LT, Schedin P. 2000. Reproductive state of rat mammary gland stroma modulates human breast cancer cell migration and invasion. Cancer Res 60: 3414–3418.

Berezikov E, Chung WJ, Willis J, Cuppen E, Lai EC. 2007. Mammalian mirtron mRNAs. Mol Cell 28: 328–336. doi:10.1016/j.molcel.2007.09.028

Bernstein E, Caudy A, Hammond S, Hannon G. 2001. Role for a Dicer-dependent RNase in the initiation step of RNA interference. Nature 409: 363–366. doi:10.1038/35053110

Boissan M, De Wever O, Lizarraga F, Wendum D, Poincloux R, Chignard N, Desbois-Mouthon C, Dufour S, Nawrocki-Raby B, Birembaut P, et al. 2010. Implication of metastasis suppressor NM23-H1 in maintaining adherens junctions and limiting the invasive potential of human cancer cells. Cancer Res 70: 7710–7722. doi:10.1158/0008-5472.CAN-10-1887

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114–2120. doi:10.1093/bioinformatics/btu170

Broughton JP, Lovci MT, Huang JL, Yeo GW, Pasquinelli AE. 2016. Pairing beyond the seed supports microRNA target conservation among gammaherpesviruses including host translation and protein modification machinery. PLoS Pathog 15: e1007843. doi:10.1371/journal.ppat.1007843

Chen H, Boutros PC. 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics 12: 35. doi:10.1186/1471-2105-12-35

Chi S, Zang J, Mele A, Damell R. 2009. Argonaute HTS-CLIP decodes microRNA-mRNA interaction maps. Nature 460: 479–486. doi:10.1038/nature08170

Chipman LB, Pasquinelli AE. 2019. miRNA targeting: growing beyond the seed. Trends Genet 35: 215–222. doi:10.1016/j.tig.2018.12.005

www.rnajournal.org 707
Lee Y, Jeon K, Lee JT, Kim S, Kim VN. 2002. MicroRNA maturation: Kozomara A, Birgaoanu M, Griffiths-Jones S. 2019. miRBase: from Harrower M, Brewer CA. 2003. ColorBrewer.org: an online tool for selecting colour schemes for maps. Harwell A, Tollervey D. 2014. Mapping the miRNA interactome by Helwak A, Duda-Kovacs T, Tollervey D. 2013. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Horak CE, Mendoza A, Vega-Valle E, Albaugh M, Graff-Cherry C, Pfeffer S, Rajewsky N, Meister G. 2008. A human snoRNA with Ender C, Krek A, Friedlander MR, Beitzinger M, Weinmann L, Chen W, Pfeffer S, Rajewsky N, Merer G. 2008. A human snRNA with microRNA-like functions. Mol Cell 32: 519–528. doi:10.1016/j.molcel.2008.10.017 Fields C, Sheng P, Miller B, Wei T, Xie M. 2019. Northern blot with IR fluorescent probes: strategies for probe preparation. Bio Protoc 9: e3219. doi:10.21769/BioProtoc.3219 Gay L, Sethuraman S, Thomas M, Turner P, Renne R. 2018. Modified cross-linking, ligation, and sequencing of hybrids (qCLASH). Nat Protoc 9: 711–728. doi:10.1038/nprot.2014.043 Helwak A, Kudla G, Duda-Kovacs T, Tollervey D. 2013. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. J Cell 153: 654–665. doi:10.1016/j.cell.2013.03.043 Horak CE, Mendoza A, Vega-Valle E, Albaugh M, Graff-Cherry C, McDermott WG, Hua E, Merino MJ, Steinberg SM, Khanna C, et al. 2007. Nm23-H1 suppresses metastasis by inhibiting expression of the lysophosphatidic acid receptor EDG2. Cancer Res 67: 11751–11759. doi:10.1158/0008-5472.CAN-07-3175 Huang Y, Hu Y, Jin Z, Shen Z. 2018. LncRNA snRNA upregulates GRB2-associated binding protein 2 and promotes proliferation of ovarian carcinoma cells. Biochem Biophys Res Commun 503: 2028–2032. doi:10.1016/j.bbrc.2018.07.152 Kawamata T, Tomari Y. 2010. Making RISC. Trends Biochem Sci 35: 368–376. doi:10.1016/j.tibs.2010.03.009 Kim Y, Kim B, Kim V. 2016. Re-evaluation of the roles of DROSHA, Exportin 5, and Dicer in microRNA biogenesis. Proc Natl Acad Sci 113: E1881–E1889. doi:10.1073/pnas.1602532113 Kozar I, Philippidou D, Margue C, Gay LA, Renne R, Kreis S. 2021. Cross-linking ligation and sequencing of hybrids (qCLASH) reveals an unpredicted miRNA targetome in melanoma cells. Cancers (Basel) 13: 1096. doi:10.3390/cancers13051096 Kozomara A, Birgaonanu M, Griffiths-Jones S. 2019. miBase: from microRNA sequences to function. Nucleic Acids Res 47: D155–D162. doi:10.1093/nar/gky1141 Langhenberger D, Cakir MV, Hoffmann S, Stadler PF. 2013. Dicer-processed small RNAs: rules and exceptions. J Exp Zool B Mol Dev Evol 320: 35–46. doi:10.1002/jez.b.22481 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357–359. doi:10.1038/nmeth.1923 Lee RC, Feinbaum RL, Ambros V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75: 843–854. doi:10.1016/0092-8674(93)90529-Y Lee Y, Jeon K, Lee JT, Kim S, Kim VN. 2002. MicroRNA maturation: stepwise processing and subcellular localization. EMBO J 21: 4663–4670. doi:10.1093/emboj/cdf476 Lee H, Kim C, Ku J, Kim W, Yoon S, Kuh H, Lee J, Nam S, Lee E. 2014. A long non-coding RNA snRNA contributes to S-fluorouracil resistance in human colon cancer cells. Mol Cells 37: 540–546. doi:10.14348/molcells.2014.0151 Lee J, Jung JH, Chae YS, Park HY, Kim WW, Lee SJ, Jeong JH, Kang SH. 2016. Long noncoding RNA snRNA regulates proliferation, migration and invasion of triple-negative breast cancer cells. Anticancer Res 36: 6289–6295. doi:10.21873/anticancerres.11224 Lee J, Park H, Kim W, Lee S, Jeong J, Kang S, Jung J, Chae Y. 2017. Biological function of long noncoding RNA snRNA in HER2-positive breast cancer cells. Tumor Biol 39: 1–8. Lemus-Díaz N, Ferreira RR, Bohnsack KE, Gruber J, Bohnsack MT. 2020. The human box C/D snoRNA U3 is a miRNA source and miR-U3 regulates expression of sortin nexin 27. Nucleic Acids Res 48: 8074–8089. doi:10.1093/nar/gkq5149 Leone A, Flato U, VanHoutte K, Steeg PS. 1993. Transfection of human mm-23-H1 into the human MDA-MB-435 breast carcinoma cell line: effects on tumor metastatic potential, colonization and enzymatic activity. Oncogene 8: 2325–2333. Lewis B, Burge C, Bartel D. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120: 15–20. doi:10.1016/j.cell.2004.12.035 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25: 2078–2079. doi:10.1093/bioinformatics/btp352 Licatalosi D, Mele A, Fak J, Ule J, Kayikci M, Chi S, Clark T, Schweitzer A, Blume J, Wang X, et al. 2008. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature 456: 464–469. doi:10.1038/nature07488 Lund E, Guttlinger S, Calado A, Dahlberg J, Kutay U. 2004. Nuclear export of microRNA precursors. Science 303: 95–98. doi:10.1126/science.1090599 Ma JB, Ye K, Patel DJ. 2004. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature 429: 318–322. doi:10.1038/nature02519 Marino N, Marshall JC, Collins JW, Zhou M, Qian Y, Veenstra T, Steeg PS. 2013. Nm23-H1 binds to gelsolin and inactivates its actin-severing capacity to promote tumor cell motility and metastasis. Cancer Res 73: 5949–5962. doi:10.1158/0008-5472.CAN-13-0368 Markham NR, Zuker M. 2008. UNAFold: software for nucleic acid folding and hybridization. Methods Mol Biol 453: 3–31. doi:10.1007/978-1-60327-429-6_1 Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal 17: 10–12. doi:10.1.14806/ej.17.1.200 McCorkle JR, Leonard MK, Kranner SD, Blalock EM, Ma D, Zimmer SG, Kaetzel DM. 2014. The metastasis suppressor NME1 regulates expression of genes linked to metastasis and patient outcome in melanoma and breast carcinoma. Cancer Genomics Proteomics 11: 175–194. McGeary SE, Lin KS, Shi CY, Pham TM, Bisaria N, Kelley GM, Bartel DP. 2019. The biochemical basis of miRNA targeting efficacy. Science 366: eaav1741. doi:10.1126/science.aav1741 Miller B, Wei T, Fields C, Sheng P, Xie M. 2018. Near-infrared fluorescent northern blot. RNA 24: 1871–1877. doi:10.1261/rna.068213.118 Miranda K, Huynt T, Tay Y, Ang Y, Tam W, Thomson A, Lim B, Rigoutsos I. 2006. A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. Cell 126: 1203–1217. doi:10.1016/j.cell.2006.07.031 Moore M, Scheel T, Luna J, Park C, Fak J, Nishuichi E, Rice C, Damell R. 2015. miRNA-target chimeras reveal miRNA 3′-end pairing as a major determinant of Argonaute target specificity. Nat Commun 6: 5864. doi:10.1038/ncomms9864
miR-snaR represses NME-1

Nazarov PV, Kreis S. 2021. Integrative approaches for analysis of mRNA and microRNA high-throughput data. Comput Struct Biotechnol J 19: 1154–1162. doi:10.1016/j.csbj.2021.01.029

Nguyen TA, Jo MH, Choi YG, Park J, Kwon SC, Hohng S, Kim VN, Woo JS. 2015. Functional anatomy of the human microprocessor. Cell 161: 1374–1387. doi:10.1016/j.cell.2015.05.010

Okamura K, Chung WJ, Ruby JG, Guo H, Bartel DP, Lai EC. 2008. The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature 453: 803–806. doi:10.1038/nature07015

O’Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44: D733–D745. doi:10.1093/nar/gkv1189

Park JE, Heo I, Tian Y, Simanshu DK, Chang H, Jee D, Patel DJ, Kim VN. 2011. Dicer recognizes the 5’ end of RNA for efficient and accurate processing. Nature 475: 201–205. doi:10.1038/nature10198

Parrott AM, Mathews MB. 2007. Novel rapidly evolving hominid RNAs bind nuclear factor 90 and display tissue-restricted distribution. Nucleic Acids Res 35: 6249–6258. doi:10.1093/nar/gkm668

Parrott A, Tsi M, Batchu P, Ryan K, Ozer H, Tian B, Mathews M. 2011. The evolution and expression of the snaR family of small non-coding RNAs. Nucleic Acids Res 39: 1485–1500. doi:10.1093/nar/gkq856

Peng Y, Croce CM. 2016. The role of microRNAs in human cancer. Signal Transduct Target Ther 1: 15004. doi:10.1038/sigtrans.2015.4

Persson H, Kvist A, Vallon-Christersson J, Medstrand P, Borg A, Rovira C. 2009. The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small interfering RNAs. Nature 461: 1268–1271. doi:10.1038/nature08372

Pinzon N, Li B, Martinez L, Sergeeva A, Presumey J, Apparailly F, Seitz H. 2017. microRNA target prediction programs predict many false positives. Genome Res 27: 234–245. doi:10.1101/gr.205146.116

R Core Team. 2012. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ruby JG, Jan CH, Bartel DP. 2007. Intronic microRNA precursors that bypass Drosha processing. Nature 448: 83–86. doi:10.1038/nature05989

Sethuraman S, Thomas M, Gay LA, Renne R. 2018. Computational analysis of ribonomics datasets identifies long non-coding RNA targets of γ-herpesviral miRNAs. Nucleic Acids Res 46: 8574–8589. doi:10.1093/nar/gky459

Sheng P, Fields C, Aadland K, Wei T, Kolaczkowski O, Gu T, Kolaczkowski B, Xie M. 2018. Dicer cleaves 5’-extended microRNA precursors originating from RNA polymerase II transcription start sites. Nucleic Acids Res 46: 5737–5752. doi:10.1093/nar/gky306

Shi Z, Wei D, Wu H, Ge J, Lei X, Guo Z, Zou R, Xiao S, Wu T, Ma R, et al. 2019. Long non-coding RNA snaR is involved in the metastasis of liver cancer possibly through TGF-β1. Oncol Lett 17: 5565–5571.

Stavas C, Erkeland S. 2019. The non-canonical aspects of microRNAs: many roads to gene regulation. Cells 8: 1465. doi:10.3390/cells8111465

Steeg PS, Bevilacqua G, Pozzatti R, Liotta LA, Sobel ME. 1988. Altered expression of NM23, a gene associated with low tumor metastatic potential, during adenovirus 2 E1a inhibition of experimental metastasis. Cancer Res 48: 6550–6554.

Travis AJ, Moody J, Helwak A, Tollevsey D, Kudla G. 2014. Hyb: a bioinformatics pipeline for the analysis of CLASH (crosslinking, ligatation and sequencing of hybrids) data. Methods 65: 263–273. doi:10.1016/j.ymeth.2013.10.015

Ungerleider N, Bullard W, Kara M, Wang X, Roberts C, Renne R, Tibbetts S, Flemington EK. 2020. EBV miRNAs are potent effectors of tumor cell transcriptome remodeling in promoting immune escape. bioRxiv doi:10.1101/2020.12.21.423766.

Van Rossum G, Python Development Team. 2018. Python tutorial. Python Software Foundation, Amsterdam.

Xie M, Steitz J. 2014. Versatile microRNA biogenesis in animals and their viruses. RNA Biol 11: 673–681. doi:10.4161/rna.28985

Yokdang N, Nordmeier S, Speirs K, Burkin HR, Buxton IL. 2015. VBAL miR-snaR represses NME-1 expression in breast cancer growth and metastasis. Int J Cancer 137: 1421–1427. doi:10.1002/ijc.29372

Zamudio J, Kelly T, Sharp P. 2014. Argonauta-bound small RNAs from promoter-proximal RNA polymerase II. Cell 156: 920–934. doi:10.1016/j.cell.2014.01.041