HHS Public Access

Author manuscript

Nat Chem Biol. Author manuscript; available in PMC 2014 October 01.

Published in final edited form as:

Nat Chem Biol. 2014 April ; 10(4): 291–297. doi:10.1038/nchembio.1452.

Sequence-based design of bioactive small molecules that target precursor microRNAs

Sai Pradeep Velagapudi[1,2], **Steven M. Gallo**[3], and **Matthew D. Disney**[1,*]

[1]Department of Chemistry, The Scripps Research Institute, Scripps Florida, 130 Scripps Way #3A1, Jupiter, Florida, 33458

[2]The Department of Chemistry, The University at Buffalo, Buffalo, NY 14260

[3]The New York State Center of Excellence in Bioinformatics and Life Sciences, The University at Buffalo, 701 Ellicott St., Buffalo, NY 14203

Abstract

Oligonucleotides are designed to target RNA using base pairing rules, however, they are hampered by poor cellular delivery and non-specific stimulation of the immune system. Small molecules are preferred as lead drugs or probes, but cannot be designed from sequence. Herein, we describe an approach termed Inforna that designs lead small molecules for RNA from solely sequence. Inforna was applied to all human microRNA precursors and identified bioactive small molecules that inhibit biogenesis by binding to nuclease processing sites (41% hit rate). Amongst 29 lead interactions, the most avid interaction is between a benzimidazole (**1**) and precursor microRNA-96. Compound **1** selectively inhibits biogenesis of microRNA-96, upregulating a protein target (FOXO1) and inducing apoptosis in cancer cells. Apoptosis is ablated when *FOXO1* mRNA expression is knocked down by an siRNA, validating compound selectivity. Importantly, microRNA profiling shows that **1** only significantly effects microRNA-96 biogenesis and is more selective than an oligonucleotide.

INTRODUCTION

RNA has essential cellular functions and is a highly desirable target for small molecule modulators of function. Indeed, recent studies have identified small molecules that bind to and modulate the function of repeating RNAs[1–4], microRNAs (miRNAs)[5], and viral elements[6–8]. Developing bioactive compounds targeting RNA is very challenging, however, compounded by the perception that RNA is "undruggable"[9, 10]. The lack of success in this area can be traced to a fundamental lack of understanding about the RNA secondary

*Correspondence to: Matthew D. Disney; Disney@scripps.edu.

Supplementary Information: This article contains Supplementary Information online.

AUTHOR CONTRIBUTIONS: S. P. V. designed and completed all experiments and contributed to writing of the manuscript; S. M. G. programmed Inforna; and M. D. D. conceived Inforna, designed experiments, and wrote the manuscript.

COMPETING FINANCIAL INTERESTS. Various aspects of this technology have copyright protection or are part of a provisional patent application.

structural motifs that are the preferred binding sites of small molecules and about the types of small molecules that bind RNA motifs with high affinity and specificity. If small molecules could be reliably designed to target RNA, it could enable investigations of RNA function and the development of therapeutics, much like studies of antibacterials binding to the ribosome were critical to elucidate the intricacies of the translational machinery[11].

Despite the large number of potential RNA targets in the transcriptome, few small molecules elicit their effects by modulating RNA function. Targeting RNA, however, could be advantageous because structural information, such as the composite and orientation of RNA motifs, can be accurately modeled from sequence[12, 13]. Moreover, many folded RNAs contribute to cellular function and dysfunction[14, 15], establishing RNA as a promising therapeutic target. Yet, non-ribosomal RNAs are rarely exploited with small molecules. Merging the ability to convert RNA sequence into secondary structure accurately with information about RNA secondary structural elements that bind to small molecules could provide an expedited route to design small molecule modulators of RNA function. Such studies have the potential to be completed quickly on a transcriptome-wide scale because the only input required is the target RNA's sequence.

Herein, we report a lead identification strategy called Inforna that integrates advances in RNA structure determination and prediction, identification of RNA motif-small molecule interactions, and the scoring of those interactions via statistical analysis (Fig. 1a). Specifically, we employed Inforna to design lead small molecules that bind to Dicer or Drosha processing sites in human miRNA precursors that are up-regulated in disease. Amongst 29 lead interactions, 12 small molecules inhibit biogenesis of the desired miRNA in primary cell lines, affording a hit rate of 41%. The most avid of these interactions targets the desired miRNA at least as selectively as an oligonucleotide antagomir and induces apoptosis in cancer cells. Collectively, these studies suggest that lead small molecules that target RNA can be designed from solely sequence in a target agnostic fashion; that is, the small molecule dictates the RNA target.

RESULTS

Development of Inforna, our lead identification strategy

Our research group previously developed a selection-based strategy (Two-Dimensional Combinatorial Screening; 2DCS)[16] and a method to statistically analyze selection data (Structure-Activity Relationships Through Sequencing; StARTS) to identify and annotate (score) RNA motif-small molecule interactions[17, 18]. In 2DCS, a small molecule library is conjugated onto an agarose microarray surface. The microarray is then probed for binding to a library of small RNA motifs likely to be found as components of larger cellular RNAs (3×3 ILL, for example; Supplementary Results, Supplementary Fig. 1). Incubation is completed in the presence of competitor oligonucleotides (C2-C8; Supplementary Fig. 1) to ensure that small molecules bind to the randomized region (N's; Supplementary Fig. 1) and not regions common to all library members[16]. During the selection process, the small molecule binds precise tertiary structures that accommodate binding. Although the exact nature of the tertiary structure and its interaction with the small molecule is unknown, this information is stored implicitly in the selected RNA's sequence. We have shown previously

that a selected motif can be extracted and placed into a different context without affecting binding affinity for the corresponding small molecule[16].

The sequences of selected RNAs are analyzed via StARTS[17, 18]. StARTS is a statistical approach that identifies features in RNA motifs that positively and negatively contribute to binding[17, 18]. StARTS predicts the affinities and selectivities of RNA-small molecule interactions by comparing the rate of occurrence of a feature in selected RNA motifs (a guanine adjacent to an adenine, for example) to its rate of occurrence in the entire RNA library. The confidence that a selected feature did not occur randomly is assigned a Z-score and a corresponding two-tailed *p*-value[17, 18]. Only features that are statistically significant ($p \leq 0.05$ or ≥95% confidence) are considered. This analysis identifies features that contribute positively (positive Z-score) and negatively (negative Z-score) to binding, allowing prediction of RNAs that bind and RNAs that do not. Each RNA motif has many statistically significant features. Therefore, the Z-scores for each feature are summed to afford a ΣZ-score (Supplementary Fig. 2)[17, 18]. By combining statistical parameters from StARTS with experimentally determined binding affinities, the affinity and selectivity of every RNA motif displayed in a library can be predicted[17, 18]. The selectivity of an RNA for different small molecules can be predicted by comparing its ΣZ-score for one small molecule to its ΣZ-score for another. For example, an RNA that has a large ΣZ-score for small molecule A and a small ΣZ-score for small molecule B is selective for small molecule A. Fitness Scores are calculated by normalizing ΣZ-scores to the RNA motif with the highest ΣZ-score. Thus, a Fitness Score of 100 indicates that a given RNA motif-small molecule interaction is the highest affinity RNA motif(s) derived from an RNA library that binds to a given small molecule. Likewise, a Fitness Score of 80 means that an RNA-small molecule interaction is 80% as fit as the optimal interaction (Fitness Score = 100).

Informa, our lead compound identification strategy, integrates RNA motif-small molecule interactions identified by 2DCS, the fitness of those interactions as determined by StARTS, and structural information about the RNA target of interest, whether determined by experiment, phylogeny, or computation (Supplementary Note). Specifically, Informa generates lead compounds for a target of interest by comparing the motifs found in its structure to the motifs in our database of annotated RNA motif-small molecule interactions (Fig. 1a). The output is the targetable motifs, the corresponding lead small molecules for the RNA of interest, and the fitness of the predicted RNA-small molecule interaction. Lead compounds are then tested for modulating biological function.

Informa identifies lead compounds that inhibit miRNAs—We validated Informa by using it to design lead compounds that target miRNA precursors that have been identified through sequencing and functional studies (Fig. 1)[19, 20]. MiRNAs are 21–25 nucleotide RNAs that negatively regulate gene expression through translational repression or cleavage of a target mRNA. They are transcribed as precursors that are processed by the nucleases Drosha and Dicer (Fig. 1a)[21]. The goal of our studies was to use *informa* to identify a small molecule that binds a precursor miRNA and inhibits its maturation (Fig. 1a).

The sequences and secondary structures of all known precursor miRNAs in the human transcriptome were downloaded from miRBase[13] and their structures modeled by free

energy minimzation[12]. Precursor miRNAs fold into small hairpin secondary structures (Fig. 1a) that can be predicted from sequence[22]. The entire set of secondary structures was then parsed by Inforna, which output the targetable motifs in each RNA and the corresponding lead small molecules that bind them. By mining all precursor miRNAs in the human transcriptome for overlap with the RNA motif-small molecule database, Inforna probed >5,400,000 potential interactions (the motifs contained in 1,048 miRNA precursors (~6,850) with 792 RNA motif-small molecule interactions housed in our database). In this study, we required that the targetable motif and its closing base pairs were exact matches for motifs in the database. Previously, it has been shown that the identity of internal loop closing pairs can dramatically affect loop structure and thus recognition by a small molecule[23, 24].

Next, we refined the lead interactions based on the following criteria: (i) the targetable motif must be in a Drosha or Dicer processing site, which are cleaved to produce pre-miRNAs and mature miRNAs, respectively. Processing is required for the mature, active miRNA to be produced[21]. (Processing sites have been determined by comparing the sequences of precursor miRNAs to mature miRNAs as determined by deep sequencing (RNA-seq)[25]. The information in miRBase is based on these experimental results[25].); and (ii) the miRNA must be causative of disease. In order to be validated as a miRNA up-regulated in disease, phenotype reversal with oligonucleotides (i.e., antagomirs) must be previously established[26–28]. For miRNAs under-expressed in a disease, supplementation of the miRNA to cells must improve disease-associated defects. In summary, lead small molecules were identified for 22 different miRNA precursors that are causative of diseases including prostate, breast, ovarian, and pancreatic cancers, Parkinson's disease, and Alzheimer's disease (Figs. 1b & c, Supplementary Fig. 3, and Supplementary Table 1). The confidence in the accuracy of the predicted structures for the miRNA hits was investigated by calculating the probabilities of all pairings using a partition function[29]. Importantly, ~77% (17/22) of the putative binding sites for our lead small molecules are predicted with >90% probability (Supplementary Fig. 4). Of the remaining six target sites, the loop nucleotides and 5' closing base pair of three are predicted with >90% confidence. (Supplementary Fig. 4).

StARTS measures the fitness of lead interactions—StARTS analysis of hit compounds was used to assess the fitness of each small molecule lead for binding the corresponding RNA motif in the precursor miRNA target (Fig. 1b, Supplementary Fig. 2, and **Online Methods**)[17, 18]. StARTS analysis determined that the fittest interactions were: **1** with the miR-96 precursor (Fitness Score = 100), **2** with the miR-210 precursor (Fitness Score = 95), and **3** with miR-182 precursor (Fitness Score = 75) (Figs. 1b, 1c, and 2A). The affinities and selectivities of these RNA motif-small molecule interactions and their Fitness Scores were measured in previous studies[18, 30].

The activities of all lead small molecules in primary cell lines were assessed by measuring mature miRNA expression by quantitative real time RT-PCR. As shown in Figures 2a & b and Supplementary Figure 5, the expression of 12 miRNAs was significantly decreased by the lead small molecule predicted by Inforna, corresponding to a hit rate of 41%. Inactive lead compounds could be optimized by modular assembly approaches[2, 31] or chemical similarity searching[4, 32–34]. Importantly, these data show that bioactive small molecules targeting RNA can be designed in a transcriptome-wide manner without target bias (or target

agnostic). Below, we describe additional characterization of the miRNA:small molecule combinations that were predicted to be the most optimal by Inforna: **1**-miR-96; **2**-miR-210; and **3**-miR-182.

Bioactivities and selectivities of compounds 1, 2, & 3—Compounds **1, 2**, and **3** inhibited biogenesis of its target precursor miRNA to varying extents (Fig. 2b): **1** reduced the expression level of miR-96 by 90% at 40 μM; **2** reduced the formation of miR-210 by 60% at 500 nM; and **3** reduced the production of miR-182 by 40% at 200 μM. (Treatment of primary cells with higher concentrations of **2** led to reduced potency, perhaps due to a lack of selectivity at these elevated concentrations.) These differences in bioactivity could be due to differences in affinity, selectivity, permeability, and/or cellular localization. Interestingly, small molecules targeting either Dicer (**2**) or Drosha (**1** and **3**) sites in precursor miRNAs can inhibit biogenesis in cell culture.

MiR-96, −182, and −183 precursors (Fig. 2a) are transcribed as a single transcript[35]; thus, biogenesis of one miRNA can serve as an internal control for the others. Therefore, **1** (targets precursor miR-96) and **3** (targets precursor miR-182) were studied for decreasing production of the other two miRNAs in the MCF7 cell line by qRT-PCR. Although **3** decreased production of the desired target, miR-182, by ~40% when cells are dosed with 200 μM compound, a less but significant effect was observed on miR-96 expression that indicates sub-optimal selectivity (Fig. 2b). In contrast, **1** efficiently and selectively silenced production of miR-96 at 40 μM while not affecting miR-182 or −183 (Fig. 2b). Thus, **1** knocked down expression of the target miRNA to a great extent than **2** and **3** and was more selective than **3**.

To confirm that the secondary structure of the miR-96 precursor was accurately modeled in miRBase and that the binding site of **1** was accurately predicted by Inforna, we completed enzymatic mapping assays. Indeed, the results of our enzymatic mapping studies were consistent with the secondary structure reported in miRBase and predicted by RNAstructure based on the probability of base pair formation (Supplementary Fig. 6). Moreover, the cleavage pattern of the putative binding site for **1** did not change when placed in a different context (Supplementary Figs. 6b & 6c). These studies also confirm that **1** bound the Drosha processing site as indicated by protection from nuclease cleavage (Supplementary Fig. 6a). Moreover, **1** inhibited Drosha cleavage of pri-miR-96 *in vitro* (Supplementary Fig. 7a) and *in vivo*, as evidenced by an increase in the levels of pri-miR-96 and a reduction in levels of pre- and mature miR-96 in treated cells, as expected if **1** binds to the Drosha site (Supplementary Fig. 7b).

Next, we compared the selectivity of **1** to a locked nucleic acid (LNA) oligonucleotide that is complementary to nucleotides 2–9 in miR-96's seed region (Fig. 2c). The LNA directed at miR-96 was studied for silencing miR-96, miR-182, and miR-183; miR-182:LNA and miR-183:LNA complexes each contain a single mismatch. Interestingly, the LNA only modestly discriminated between miR-96 and miR-183 at all concentrations tested (1 – 200 nM) (Fig. 2c). At 50 nM concentration, the LNA silenced ~90% of miR-96 expression and ~50% of miR-183 expression. (Non-selective effects of oligonucleotides designed to silence specific miRNAs have been previously observed[36].) In contrast, at a concentration of **1** that

silenced 90% of miR-96 expression, miR-182 was affected by only 15% and miR-183 was not affected (Fig. 2c). Taken together, small molecules can be more selective modulators of function than oligonucleotides in some instances.

Assessing the downstream effects of 1—Next, the effect of **1** on the downstream targets of miR-96 was assessed. MiR-96 is upregulated in cancer and is linked to oncogenic transformation by silencing of *FOXO1* (Forkhead box protein O1) through translational repression[37, 38]. FOXO transcription factors function as regulators of cell cycle progression, including apoptosis[39]. We tested **1** for its ability to increase expression of FOXO1 protein using a luciferase model system; that is, a target sequence fully complementary to miR-96 was inserted downstream of the Renilla luciferase gene[38]. A small molecule that inhibits maturation of miR-96 will increase luciferase expression. MCF7 cells were transfected with a plasmid that expresses the luciferase model system, followed by treatment with **1**. In agreement with decreased miR-96 production observed in qRT-PCR experiments (Fig. 2b), **1** stimulated production of luciferase, indicating inhibition of miR-96 maturation (Fig. 3c). Specifically, **1** increased production of luciferase by ~1.7-fold when cells were treated with 40 μM of compound. Importantly, **1** did not affect production of luciferase when a plasmid encoding a *FOXO1* 3' UTR that is unresponsive to miR-96 was used (Fig. 3a), confirming that miR-96, and not the *FOXO1* UTR, is **1**'s target.

Next, we tested **1** for its ability to increase endogenous levels of FOXO1 protein by Western blotting (Figs. 3b & Supplementary Fig. 8). When MCF7 cells were treated with 40 μM of **1**, an ~2.5-fold increase in FOXO1 protein levels was observed, consistent with luciferase experiments. No effect was observed on the expression of a GAPDH control (Fig. 3b & Supplementary Fig. 8). Since FOXO1 up-regulation stimulates apoptosis[40], we determined if **1** stimulates apoptosis via a TUNEL assay. When MCF7 cells were treated with 40 μM of **1**, approximately 40% of the cells were TUNEL positive (Fig. 3c). As a secondary test of apoptosis, Annexin V/PI assays were employed as they can distinguish necrosis from early apoptosis. As expected, these studies showed that **1** stimulated early apoptosis, not global cell death via necrosis (Fig. 3c & Supplementary Fig. 9).

1 stimulates apoptosis via the miR-96-*FOXO1* pathway—To confirm that **1** induces apoptosis by modulation of the miR-96-*FOXO1* regulation pathway, *FOXO1* expression was knocked down via an siRNA (Fig. 3d). If **1** induces apoptosis by selectively targeting miR-96, then removal of *FOXO1* should affect apoptosis. Indeed, when *FOXO1* siRNA was applied to cells that were then treated with **1**, a 70% reduction in apoptosis was observed compared to cells transfected with an siRNA against *GAPDH* (control) and treated with **1** (Fig. 3d). It is not surprising that using an siRNA to ablate *FOXO1* mRNA from cells does not completely eliminate the apoptotic effect of **1** since it is well known that miRNAs can target many different mRNAs for silencing simultaneously[41]. These studies, however, validated that the miR-96-*FOXO1* mRNA pathway is regulated by **1** and that **1** is a specific inducer of apoptosis.

The selectivity of 1 as revealed by miRNA profiling—Compound **1** was predicted to bind Dicer and Drosha processing sites in other miRNA precursors; however, these

predicted interactions are less fit than that between **1** and the miR-96 precursor (Fig. 4a). Therefore, the selectivity of **1** for modulating expression of 149 disease-associated and highly abundant miRNAs (Supplementary Table 2) was assessed via qRT-PCR. As shown in Figure 4b, only miR-96 was significantly affected. These studies confirm the selectivity of **1** for the desired target on a transcriptome-wide level. For comparison, we profiled the effect of a full-length antagomir directed at miR-96 (100 nM) on the expression of the same 149 miRNAs (Fig. 4b). The antagomir affected expression of 12 miRNAs (including miR-96) by at least 2.5-fold. Thus, it appears that **1** provides an unparalleled level of selectivity for a small molecule that modulates RNA function and that selectivity is greater than that observed with some miRNA-targeting oligonucleotides, as evidenced by the studies in Figs. 2c & Fig. 4b. These results can be rationalized from our StARTS analysis: the Fitness Score for **1** and the target site in miR-96 is 100, indicating the most fit interaction possible. The Fitness Scores for **1** and potential binding sites in other miRNAs are much less fit, ranging from ~10–40 (Fig. 4a). In fact, there are only five other motifs, none of which are found in human miRNAs, that have Fitness Scores >50, indicating that **1** is a highly selective small molecule (Fig. 4a).

Inforna vs. traditional medicinal chemistry approaches—The luciferase model system described above provides a robust assay to test other small molecules for modulating the miR-96-*FOXO1* pathway. We therefore used this system to compare the design of small molecules via Inforna to traditional medicinal chemistry approaches – screening and lead optimization via the synthesis of compound derivatives. First, we tested a previously constructed library of small molecules that are biased for binding RNA (28 total compounds; Supplementary Fig. 10)[42]. None of the compounds stimulated luciferase production at 40 μM (Supplementary Fig. 10), indicating that: (i) the compounds are not bioactive; and (ii) the miRNA-*FOXO1* pathway is not easily druggable.

A commonly employed method for drug optimization is the synthesis of compound derivatives, i.e., chemically similar small molecules. We therefore tested three compounds that are chemically similar to **1**, compounds **2, 4**, and **5** (Fig. 1c), as determined by their shape Tanimoto coefficients[43]. Shape Tanimoto coefficients quantitatively determine the three dimensional similarity of two compounds; values range from 0 (no similarity) to 1 (complete similarity)[43]. The shape Tanimoto coefficients for **2, 4**, and **5** as compared to **1** are 0.94, 0.89 and 0.80, respectively, illustrating quantitative similarity between these compounds. Although all four compounds are based on a benzimidazole scaffold, visual inspection of structure suggests that **2** and **5** might bind RNA with higher affinity than **1** because of the larger surface area and additional hydrogen bond donors and acceptors. Each factor suggests that these compounds could modulate miR-96 maturation. Despite their similarities, **2** and **4** have very different Fitness Plots for binding to the target site of **1** in the miR-96 precursor (Supplementary Fig. 11). (A Fitness Plot could not be generated for **5** as it does not bind RNA motifs in 2DCS selections[18].) In fact, Inforna predicts that none of the compounds should bind avidly. A lack of binding affinity of these compounds for the miR-96 precursor was experimentally confirmed (Supplementary Fig. 12).

The activities of compounds **2, 4**, and **5** were studied in the MCF7 cells. The compounds were inactive at 40 μM concentration in all assays: (i) they did not affect miR-96 expression levels (Supplementary Fig. 13); (ii) they did not stimulate luciferase production in the miR-96-*FOXO1* model system (Fig. 3A); and (iii) they did not induce apoptosis (Fig. 3c). The inactivity of **2, 4**, and **5** was not due to differences in cellular permeability (Supplementary Fig. 14). Thus, our integrated Inforna approach provides reliable prediction of small molecules that are capable and incapable of targeting RNAs, and these predictions are more accurate than standard medicinal chemistry approaches such as screening and chemical similarity searching.

DISCUSSION

Herein, we have described a computational approach, based on experimentally derived parameters, that identifies lead small molecules for RNA. Importantly, this Inforna method allows the small molecule to dictate its ideal RNA target as defined by the fitness of the interaction. That is, this approach can be “target agnostic” in that it can identify the potential preferred RNA targets for a small molecule. Although a target agnostic approach was employed herein, Inforna can potentially be used for any RNA target of interest. Its structure need not be predicted from sequence; rather the structure can be determined from experiment, phylogeny, computation, or combinations thereof and parsed by Inforna to predict lead small molecules.

Indeed, Inforna successfully identified leads for 12 miRNA precursors, corresponding to a hit rate of 41%. In these studies, we refined lead small molecule-miRNA interactions by requiring the small molecule to bind to Dicer or Drosha processing sites as this could be a general manner to inhibit miRNA biogenesis. We chose not to focus on apical (hairpin) loops or targetable motifs within the mature miRNA for various reasons. Previous studies indicate that the apical loop is dispensable[44] and that there is poor phylogenetic conservation of the apical loop of a given miRNA across vertebrate genomes (only ~14% of all miRNAs)[45]. This suggests that, in general, the loop is not essential for processing. It is likely that those that are conserved require special trans-factors for efficient biogenesis[46]. Aptamers have been developed that target apical loops; two inhibit Drosha processing[46, 47] while another stimulates it[47]. Targetable motifs located within the mature miRNA sequence were also not considered as a recent study showed that, in some instances, mutations within these regions are well tolerated; that is, they do not affect processing[48]. Taken together, apical loops or structures within the stem region do not appear to be *general* target sites for inhibition of miRNA biogenesis.

The most fit small molecule-miRNA interactions predicted by Inforna is highly selective *in vivo* as determined from miRNA profiling studies (Fig. 4). This selectivity can be traced to modulation of an oncogene rather than non-specifically affecting cellular function. Most anticancer therapeutics, such as *cis*-platin and chlorambucil, target biomolecules non-specifically, giving rise to side effects[49]. Moreover, compound **1** was more selective than an LNA oligonucleotide that binds miR-96’s seed region and an antagmoir (Figs. 2 & 4). Taken together, our studies suggest that small molecules targeting the secondary structure of precursor miRNAs can be more selective modulators of function than oligonucleotides that

target RNA sequence. This is perhaps due to energetic degeneracy in targeting sequence (even with mismatches) that may not occur when targeting structure. Moreover, StARTS can provide insight into the energetic degeneracy of a small molecule for binding to more than one RNA target. Another important advantage of small molecules is their cellular permeability, a property not innate to oligonucleotides.

Informa's hit rate for providing bioactive compounds is quite good and suggests that our approach could be more efficient than high throughput screening or other computational approaches such as structure-based design and docking. We compared the bioactivity of compound **1** to a library of small molecules that are biased for binding to RNA and three chemically similar compounds. None affected miR-96 biogenesis (Supplementary Figs. 12 – 14). We have demonstrated that Informa enables reliable design of bioactive small molecules from only sequence. Since a myriad of genomic and functional studies are rapidly providing information about disease-associated genes[50], our approach could provide an expedited route to small molecules that target the RNA product of those genes and potentially transform the manner in which RNA-directed chemical probes or therapeutics are identified. It is likely as our dataset expands that Informa can enable targeting of other RNA targets with small molecules, including viral RNAs and structured regions in mRNAs such as UTRs. Such rational, computational approaches could afford lead compounds to study function without requiring the 3D structure of the RNA (structure-based design/docking) or completing a high throughput screen, expeditiously fulfilling the promise of genomics-to-patient therapeutics.

ONLINE METHODS

DNA Templates and PCR Amplification

The RNA motifs (internal loops) used in these studies were embedded into a hairpin cassette, **C1** (Supplementary Fig. 1)[51]. The corresponding DNA templates (Integrated DNA Technologies, Inc.) were PCR amplified in 1× PCR Buffer (10 mM Tris, pH 9.0, 50 mM KCl, and 0.1% (v/v) Triton X-100), 2 μM forward primer (5'-GGCCGGATCCTAATACGACTCACTATAGGGAGAGGGTTTAAT), 2 μM reverse primer (5'-CCTTGCGGATCCAAT), 4.25 mM $MgCl_2$, 330 μM dNTPs, and 1 μL of Taq DNA polymerase in a 50 μL reaction. The cycling conditions used for PCR were 95 °C for 30 s, 50 °C for 30 s, and 72 °C for 1 min.

The miR-96 precursor used in nuclease protection assays was modified to contain a 5'-GGG overhang to facilitate transcription using T7 RNA polymerase, or GGG-pre-microRNA-96. There was no change in the lowest free energy secondary structure predicted by RNAstructure[12], and nuclease protection assays confirmed that the predicted structure is adopted in solution (Supplementary Fig. 6). The DNA template for GGG-pre-microRNA-96 was PCR amplified as described above except the primers were 5'-GGCCGGATCCTAATACGACTCACTATAGGGTGGCCGATTTTGGC (forward) and 5'-TTTCCCATATTGGCA (reverse) and the cycling conditions were 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 1 min. The DNA templates used for PCR to produce double stranded DNAs suitable for transcription were: **C1**: 5'-GGGAGAGGGTTTAATTACGAAAGTAATTGGATCCGCAAGG **5'UUU/3'AUA**:5'-

GGGAGAGGGTTTAA**TTT**TACGAAAGTA**ATA**TTGGATCCGCAAGG **5'CGAUUU/3'GGUAUA**: 5'-GGGAGAGGGTTTAATC**CGATTT**TACGAAAGTA**ATATG-G**GATTGGATCCGCAAGG GGG-pre-microRNA-96: 5'**GGG**TGGCCGATTTTGGCACTAGCACATTTTTGCTTGTGTC-TCTCCGCTCTGAGCAATCATGTGCAGTGCCAATATGGGAAA

RNA Transcription

RNA oligonucleotides were *in vitro* transcribed by T7 RNA polymerase in 1× Transcription Buffer (40 mM Tris HCl, pH 8.1, 1 mM spermidine, 0.001% (v/v) Triton X-100 and 10 mM DTT)[52] containing 2.25 mM of each rNTP and 5 mM $MgCl_2$ at 37 °C overnight. The transcribed RNAs were purified on a denaturing 15% polyacrylamide gel and isolated as previously described[30]. Concentrations were determined by absorbance at 260 nm and the corresponding extinction coefficient. Extinction coefficients were calculated using the HyTher server[53, 54], which uses parameters based on the extinction coefficients of RNA nearest neighbors[55].

Binding Affinity Measurements

Dissociation constants were determined using an in solution, fluorescence-based assay[16]. The RNA of interest was folded in 1× Assay Buffer (8 mM Na_2HPO_4, pH 7.0, 190 mM NaCl, 1 mM EDTA and 40 μg/mL BSA) by heating at 60 °C for 5 min and slowly cooling to room temperature. Fluorescently labeled small molecule was added to a final concentration of 50 nM for **1-Fl, 4-Fl**, and **5-Fl** or 500 nM for **2**. Serial dilutions (1:2) were then completed in 1× Assay Buffer supplemented with 50 nM of **1-Fl, 4-Fl**, or **5-Fl** or 500 nM of **1**. The solutions were incubated for 30 min at room temperature and then transferred to a 96-well plate and fluorescence intensity measured. The change in fluorescence intensity as a function of RNA concentration was fit to the following equation[56]:

$$I=I_0+0.5\Delta\varepsilon(([FL]_0+[RNA]_0+K_t)-(([FL]_0+[RNA]_0+K_t)^2-4[FL]_0[RNA]_0)^{0.5})$$

where I and I_0 are the observed fluorescence intensity in the presence and absence of RNA respectively, Δε is the difference between the fluorescence intensity in the absence of RNA and in the presence of infinite RNA concentration, $[FL]_0$ and $[RNA]_0$ are the concentrations of small molecule and RNA, respectively, and K_t is the dissociation constant.

The targetable loops (motifs) were embedded into **C1** (Supplementary Fig. 1) such that affinity measurements could be completed. The secondary structures of the RNAs as predicted by RNAstructure[12] and representative binding curves are shown in Supplementary Fig. 12.

Nuclease Protection Assays

GGG-pre-microRNA-96 was 5'end labeled with ^{32}P as previously described[57]. Double stranded RNA-specific endoribonuclease cleavage: The RNA was folded in 1× Reaction Buffer (Ambion) by heating at 60 °C for 5 min and slowly cooling to room temperature. Double stranded-RNA specific endoribonuclease (*Escherichia coli* RNase III; Ambion) was

then added to a final concentration of 0.15 units/μL followed by addition of serially diluted concentrations of **1**. The solution was incubated at 37 °C for 2 h, and the cleavage products were separated on a denaturing 15% polyacrylamide gel. RNase T1 Cleavage: The RNA was folded in either 1× RNA Sequencing Buffer (Ambion; denaturing conditions) or 1× RNA Structure Buffer (Ambion; native conditions) by incubating it at 55 °C for 10 min followed by slowly cooling to room temperature. RNase T1 was added to a final concentration of 0.1 units/μL followed by addition of serially diluted concentrations of **1**. The solution was incubated at room temperature for 15 min, and the cleaved products were separated on a denaturing 15% polyacrylamide gel.

Preparation of Cell Extracts Containing Drosha

HEK 293T cells were maintained in DMEM supplemented with 10% FBS in a T-75 flask. After cells reached 70% confluency, they were transfected with Drosha-cmyc[58] (Addgene plasmid 10828) using Lipofectamine 2000 (Invitrogen) per the manufacturer's protocol. Approximately 48 h post-transfection, the cells were collected by scraping into 1 mL ice-cold 1× DPBS followed by centrifugation at 6000 rpm for 5 min at 4 °C. The cells were resuspended in 500 μL of 1× Lysis Buffer (20 mM Tris HCl, pH 8.0, 100 mM KCl, and 0.2 mM EDTA) and sonicated for 30 s. Cellular debris was pelleted by centrifugation (12000 rpm for 15 min at 4 °C), and the supernatant containing Drosha was transferred to a new tube.

Inhibition of Drosha Cleavage *In Vitro*

The cDNA template for pri-miRNA-96 was PCR amplified from MCF7 genomic DNA using the following primers: forward primer, 5'-GGCCGAATTCTAATACGACTCACTATAGGCACCAGTGCCATCTGCTT; and reverse primer, 5'-CGCAGCTGCGGGTCCT. The forward primer contains a T7 promoter that was employed to produce pri-miR-96 via run-off transcription as described above. Internally labeled pri-miR-96 was transcribed using [α-^{32}P]ATP and purified using a denaturing 10% polyacrylamide gel.

To determine if **1** inhibits Drosha cleavage *in vitro*, 2 μL of ^{32}P-labeled pri-miR-96 (approx. 10,000 counts) and 40 μM of **1** were incubated in 6.4 mM $MgCl_2$ (30 μL total volume) at room temperature for 10 min. (Untreated controls included 0.04% DMSO, the same concentration as in treated samples). Then, 1 μL of the Drosha-cmyc lysate was added, and the samples were incubated at 37 °C for 3 h. The reactions were quenched by phenol-chloroform extraction followed by ethanol precipitation. The resulting pellet was dissolved in 10 μL of 2× Gel Loading Buffer (8 M urea, 50 mM EDTA, 0.05% (w/v) bromophenol blue, 0.05% (w/v) xylene cyanol), and the reaction products were separated on a denaturing 10% polyacrylamide gel.

Cell Culture

MCF7 cells were cultured in Dulbecco's modified eagle medium/F12 (DMEM/F12) supplemented with 10% FBS (complete growth medium) at 37 °C and 5% CO_2.

Plasmids

All luciferase constructs were generous gifts from Dr. Bruce White (University of Connecticut Health Center)[38].

RNA Isolation and Quantitative Real Time Polymerase Chain Reaction (qRT-PCR) of miRNAs

MCF7 cells were cultured in either 6-well or 12-well plates, and total RNA was extracted using TRIzol LS reagent (Ambion) per the manufacturer's protocol. Approximately 200 ng of total RNA was used in reverse transcription reactions, which were completed using a Taqman MicroRNA RT Kit (Applied Biosystems) or a miScript II RT kit (Qiagen) per the manufacturer's protocol. qRT-PCR was performed on a 7900HT Fast Real Time PCR System (Applied Biosystem) using Taqman Universal PCR Master Mix or power SYBR Green Master Mix (Applied Biosystems). All primer sets for mature miRNAs were purchased from Applied Biosystems. The expression level of mature miRNAs was normalized to U6 small nuclear RNA.

Pri-, pre-, and mature miR-96 RT-PCR products were analyzed on a denaturing 10% polyacrylamide gel stained with ethidium bromide. Expression levels were normalized to GAPDH. The primers used for pre- and pri-miR-96 and GAPDH were purchased from IDT (Supplementary Table 2). Statistical significance was calculated using a two-tailed student *t*-test.

Transcriptome-wide Profiling of Mature MiRNAs by qRT-PCR

MCF7 cells were cultured and total RNA extracted as described above. Approximately 1 μg of total RNA was used in reverse transcription reactions, which were completed using miScript II RT kit (Qiagen) per the manufacturer's protocol. qRT-PCR was performed using power SYBR Green Master Mix (Applied Biosystems) on a 7900HT Fast Real Time PCR System. All forward primers were purchased from Life Technologies while the universal reverse primer was purchased from IDT (Supplementary Table 2). Expression levels were normalized using RNU6, SNORD44, SNORD47 and SNORD48. Cells were treated with 40 μM of **1** or 100 nM of full-length antagomir against miR-96 (anti-hsa-miR-96-5p; Qiagen).

Dual Luciferase Assay

MCF7 cells were grown in 96-well plates to ~80% confluency in complete growth medium. The cells were transiently transfected with 100 ng of plasmid encoding either the miR-96 3' UTR target or a mutated 3' UTR target using Lipofectamine 2000 per the manufacturer's protocol. Approximately 5 h post transfection, the small molecule of interest was added in complete growth medium, and the cells were incubated for another 20 h. Luciferase activity was then measured using a Dual Glo Luciferase Assay System (Promega) per the manufacturer's protocol. The values reported are the average of at least three measurements, and errors are the corresponding standard deviations. Statistical significance was calculated using a two-tailed student *t*-test.

Western Blotting

MCF7 cells were grown in 6-well plates to ~80% confluency in complete growth medium. The cells were then incubated with 40 μM of **1** for 20 h. Total protein was extracted using M-PER Mammalian Protein Extraction Reagent (Pierce Biotechnology) using the manufacturer's protocol. Extracted total protein was quantified using a Micro BCA Protein Assay Kit (Pierce Biotechnology). Approximately 40 μg of total protein was resolved on an 8% SDS-polyacrylamide gel, and then transferred to a PVDF membrane. The membrane was briefly washed with 1× Tris-buffered saline (TBS), and then blocked in 5% milk dissolved in 1× TBST (1× TBS containing 0.1% Tween-20) for 1 h at room temperature. The membrane was then incubated in 1:1000 FOXO1 primary antibody (Cell Signaling Technology) in 1× TBST containing 3% (w/v) BSA overnight at 4 °C. The membrane was washed with 1× TBST and incubated with 1:2000 anti-rabbit IgG horseradish-peroxidase conjugate (Cell Signaling Technology) in 1× TBS for 1 h at room temperature. After washing with 1× TBST, protein expression was quantified using SuperSignal West Pico Chemiluminescent Substrate (Pierce Biotechnology) per the manufacturer's protocol. The membrane was then stripped using 1× Stripping Buffer (200 mM glycine, pH 2.2 and 0.1% SDS) followed by washing in 1× TBS. The membrane was blocked and probed for GAPDH following the same procedure described above using 1:2000 GAPDH primary antibody (Abcam). ImageJ software (National Institutes of Health) was used to quantify band intensities. Statistical significance was calculated using a two-tailed student *t*-test.

siRNA Treatment

ON-TARGETplus SMARTpool FOXO1 siRNA and ON-TARGETplus GAPD Control Pool (Dharmacon; Thermo Scientific) were used at a final concentration of 100 nM. MCF7 cells were reverse transfected in a 6-well plate using Lipofectamine RNAiMAX reagent (Invitrogen) per the manufacturer's protocol.

APO BrdU TUNEL Assay

MCF7 cells were grown in 6-well plates to 70–80% confluency. For experiments in which an siRNA was used to knock down *FOXO1* or *GAPDH* expression, the cells were first transfected as described above. Cells were treated with small molecules for 20 h followed by completion of an APO BrdU TUNEL assay (Molecular Probes) per the manufacturer's protocol. Flow cytometry was performed using a BD LSRII instrument (BD Biosciences). At least 10,000 events were used for analysis. Statistical significance was calculated using a two-tailed student *t*-test.

Annexin V/PI Assays

MCF7 cells were grown in 6-well plates to 70–80% confluency. The cells were incubated with 40 μM of **1** for 20 h and then detached from the surface using accutase. They were washed twice with ice cold 1× DPBS and then twice with 1× Annexin Binding Buffer (50 mM Hepes (pH 7.4), 700 mM NaCl and 12.5 mM $CaCl_2$). The cells were resuspended in 100 μL 1× Annexin Binding Buffer, and then 5 μL Annexin V-APC (eBioscience) were added. The solution was incubated for 10 min at room temperature followed by washing with 1× Annexin Binding Buffer. The cells were then stained with 1 μg/mL propidium

iodide (PI) in 300 μL of 1× Annexin Binding Buffer for 15 min at room temperature. Flow cytometry was performed using a BD LSRII instrument (BD Biosciences). At least 10,000 events were used for analysis. Statistical significance was calculated using a two-tailed student *t*-test.

Cellular Permeability

MCF7 cells were grown in 6-well plates to ~80% confluency in complete growth medium. Then, 10 μM of the indicated small molecule was added, and the cells were incubated for 20 h. The cells were trypsinized from the plate, washed twice with 1× DPBS, and stained with 1 μg/mL PI for 30 min on ice. Cellular permeability was quantified by flow cytometry using a BD LSRII flow cytometer (BD Biosciences) and Hoechst filters. At least 10,000 events were used for analyses (Supplementary Fig. 14).

Assessment of Cellular Morphology

MCF7 cells were grown in 6-well plates to ~80% confluency in complete growth medium. Then, 40 μM of the indicated small molecule was added to the well, and the cells were incubated for 20 h. The cells were imaged using an Olympus IX71 microscope (Supplementary Fig. 14).

Synthesis of 1, 2, 4, 5 and their fluorescein conjugates

Compounds **1, 2, 4** and **5** and the fluorescein conjugates of **1** and **4 (1-Fl** and **4-Fl**, respectively; Supplementary Fig. 15) were synthesized as previously described[22]. Compounds **1-Fl, 4-Fl**, and **5-Fl** were used to measure binding affinities; **2** was used directly to measure affinities (without fluorescein conjugation).

Synthesis of 5-Fl (Supplementary Figs. 15 & 16)

A 3.4 μmol sample of *N*-(2-propynyl) 5-fluoresceincarboxamide in methanol was added to a solution containing 10 μmol **5**, 34 μmol $CuSO_4$, 72 μmol freshly dissolved ascorbic acid, and 0.3 μmol TBTA. The final volume was brought to 1.5 mL with methanol. The reaction mixture was transferred to a microwave reaction vessel, a magnetic stir bar was added, and the flask was sealed with a Teflon septum and aluminum crimp top. The reaction vessel was placed in an Emrys™ Optimizer system (Biotage), and the reaction was maintained at 110 °C for 4 h with stirring. The crude reaction mixture was purified by reverse phase HPLC using a linear gradient of 20% to 100% B in A over 60 min (A is 0.1% (v/v) TFA in water while B is 0.1% (v/v) TFA in methanol). The purity of the product was evaluated on a Waters Symmetry® C18 5 μm 4.6×150 mm column using a Waters 1525 binary HPLC pump equipped with a Waters 2487 dual λ absorbance detector system. Separations were completed at room temperature using a 1 mL/min flow rate and a linear gradient of 0% to 100% B in A over 50 min. Absorbance was monitored at 220 and 254 nm. t_R = 35 min; isolated yield = 70% (determined by absorbance at 496 nm in 1X PBS, pH 7.4 using an extinction coefficient of 45000 M^{-1} cm^{-1}). High-resolution mass spectrometry analysis: calculated mass: 890.3580 ($M+H^+$); observed mass: 890.3590 ($M+H^+$). ^1H NMR: (CD_3OD, 400 MHz) δ 8.46 (s, 1H), 8.23 (2H,d, J=9), 8.00 (3H, m), 7.68 (1H,d, J=9), 7.40 (1H, J=9), 7.34 (1H, dd, J=2,34) 7.25 (2H, d, J=9), 6.71 (s, 2H), 6.57 (4H, m), 5.00 (s, 1H), 4.70 (s,

2H), 4.44 (2H, t, J=14), 4.18 (2H, t, J=14),3.95 (2H, m), 3.70 (3H,m), 3.24 (5H, m), 3.02 (3H, s), 2.43 (2H, t, 14), 2.13 (4H, m); ^{13}C NMR: (CD_3OD, 175 MHz) δ 173.9, 163.3, 149.2, 149.1, 132.1, 129.2, 126.2, 117.8, 115.6, 114.6, 113.9, 102.2, 99.6, 67.5, 53.2, 46.5, 42.1, 36.0, 31.8, 29.6, 24.7;

Infornа Software

Informa is freely available for academic use and may be accessed at http://RNA-informa.florida.scripps.edu. Please visit http://www.scripps.edu/disney/software.html for access. The authors request citation of the use of Informa in all published works that use this resource.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Prof. Bruce White (Department of Cell Biology, University of Connecticut Health Center) for the luciferase plasmids containing FOXO1 3’ UTRs; Dr. Christopher Haga for assistance designing qRT-PCR primers; Steve Seedhouse for preliminary studies with Informa; Biao Liu, Jesse Sokolow, and Tuan Tran for compiling the secondary structures of the miRNA precursors and the RNA motif-small molecule database; and Dr. Jessica Childs-Disney, Prof. Tom Kodadek, Prof. John Cleveland, Prof. Bill Roush, Prof. Jerry Joyce, Prof. Mark Burkard, and Prof. Min Guo for critical review of the manuscript. This work was funded by the National Institutes of Health (R01GM097455). M.D.D. is a Camille and Henry Dreyfus Teacher-Scholar.

REFERENCES

1. Arambula JF, Ramisetty SR, Baranger AM, Zimmerman SC. A simple ligand that selectively targets CUG trinucleotide repeats and inhibits MBNL protein binding. Proc. Natl. Acad. Sci. U. S. A. 2009; 106:16068–16073. [PubMed: 19805260]
2. Childs-Disney JL, Hoskins J, Rzuczek SG, Thornton CA, Disney MD. Rationally designed small molecules targeting the RNA that causes myotonic dystrophy type 1 are potently bioactive. ACS Chem. Biol. 2012; 7:856–862. [PubMed: 22332923]
3. Ofori LO, Hoskins J, Nakamori M, Thornton CA, Miller BL. From dynamic combinatorial 'hit' to lead: in vitro and in vivo activity of compounds targeting the pathogenic RNAs that cause myotonic dystrophy. Nucleic Acids Res. 2012; 40:6380–6390. [PubMed: 22492623]
4. Kumar A, et al. Chemical correction of pre-mRNA splicing defects associated with sequestration of muscleblind-like 1 protein by expanded r(CAG)-containing transcripts. ACS Chem. Biol. 2012; 7:496–505. [PubMed: 22252896]
5. Bose D, et al. The tuberculosis drug streptomycin as a potential cancer therapeutic: inhibition of miR-21 function by directly targeting its precursor. Angew. Chem. Int. Ed. Engl. 2012; 51:1019–1023. [PubMed: 22173871]
6. Stelzer AC, et al. Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. Nat. Chem. Biol. 2011; 7:553–559. [PubMed: 21706033]
7. Parsons J, et al. Conformational inhibition of the hepatitis C virus internal ribosome entry site RNA. Nat. Chem. Biol. 2009; 5:823–825. [PubMed: 19767736]
8. Davidson A, et al. Simultaneous recognition of HIV-1 TAR RNA bulge and loop sequences by cyclic peptide mimics of Tat protein. Proc. Natl. Acad. Sci. U. S. A. 2009; 106:11931–11936. [PubMed: 19584251]
9. Guan L, Disney MD. Recent advances in developing small molecules targeting RNA. ACS Chem. Biol. 2012; 7:73–86. [PubMed: 22185671]

10. Thomas JR, Hergenrother PJ. Targeting RNA with small molecules. Chem. Rev. 2008; 108:1171–1224. [PubMed: 18361529]
11. Yonath A, Bashan A. Ribosomal crystallography: initiation, peptide bond formation, and amino acid polymerization are hampered by antibiotics. Ann. Rev. Microbiol. 2004; 58:233–251. [PubMed: 15487937]
12. Mathews DH, et al. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc. Natl. Acad. Sci. U. S. A. 2004; 101:7287–7292. [PubMed: 15123812]
13. Woese, C.; Pace, N. Probing RNA structure, function, and history by comparative analysis. In: Gesteland, RF.; Cech, TR.; Atkins, JF., editors. The RNA World. 2nd edn.. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1993. p. 91-117.
14. Batey RT, Rambo RP, Doudna JA. Tertiary motifs in RNA structure and folding. Angew. Chem. Int. Ed. Engl. 1999; 38:2326–2343. [PubMed: 10458781]
15. Spahn CM, et al. Hepatitis C virus IRES RNA-induced changes in the conformation of the 40s ribosomal subunit. Science. 2001; 291:1959–1962. [PubMed: 11239155]
16. Disney MD, et al. Two-dimensional combinatorial screening identifies specific aminoglycoside-RNA internal loop partners. J. Am. Chem. Soc. 2008; 130:11185–11194. [PubMed: 18652457]
17. Velagapudi SP, Seedhouse SJ, Disney MD. Structure-activity relationships through sequencing (StARTS) defines optimal and suboptimal RNA motif targets for small molecules. Angew. Chem. Int. Ed. Engl. 2010; 49:3816–3818. [PubMed: 20397174]
18. Velagapudi SP, Seedhouse SJ, French J, Disney MD. Defining the RNA internal loops preferred by benzimidazole derivatives via 2D combinatorial screening and computational analysis. J. Am. Chem. Soc. 2011; 133:10111–10118. [PubMed: 21604752]
19. Jiang Q, et al. miR2Disease: a manually curated database for microRNA deregulation in human disease. Nucleic Acids Res. 2009; 37:D98–104. [PubMed: 18927107]
20. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. Nucleic Acids Res. 2008; 36:D154–158. [PubMed: 17991681]
21. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004; 116:281–297. [PubMed: 14744438]
22. Ambros V, et al. A uniform system for microRNA annotation. RNA. 2003; 9:277–279. [PubMed: 12592000]
23. Wu M, Turner DH. Solution structure of $(rGCGGACGC)_2$ by two-dimensional NMR and the iterative relaxation matrix approach. Biochemistry. 1996; 35:9677–9689. [PubMed: 8703939]
24. SantaLucia J Jr. Turner DH. Structure of $(rGGCGAGCC)_2$ in solution from NMR and restrained molecular dynamics. Biochemistry. 1993; 32:12612–12623. [PubMed: 8251479]
25. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 2011; 39:D152–157. [PubMed: 21037258]
26. Krutzfeldt J, et al. Silencing of microRNAs in vivo with 'antagomirs'. Nature. 2005; 438:685–689. [PubMed: 16258535]
27. Ebert MS, Sharp PA. MicroRNA sponges: progress and possibilities. RNA. 2010; 16:2043–2050. [PubMed: 20855538]
28. Obad S, et al. Silencing of microRNA families by seed-targeting tiny LNAs. Nat. Genet. 2011; 43:371–378. [PubMed: 21423181]
29. Mathews DH. Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. RNA. 2004; 10:1178–1190. [PubMed: 15272118]
30. Childs-Disney JL, Wu M, Pushechnikov A, Aminova O, Disney MD. A small molecule microarray platform to select RNA internal loop-ligand interactions. ACS Chem. Biol. 2007; 2:745–754. [PubMed: 17975888]
31. Pilch DS, et al. Binding of a hairpin polyamide in the minor groove of DNA: sequence-specific enthalpic discrimination. Proc. Natl. Acad. Sci. U. S. A. 1996; 93:8306–8311. [PubMed: 8710866]

32. Pinto IG, Guilbert C, Ulyanov NB, Stearns J, James TL. Discovery of ligands for a novel target, the human telomerase RNA, based on flexible-target virtual screening and NMR. J. Med. Chem. 2008; 51:7205–7215. [PubMed: 18950148]
33. Disney MD, et al. A small molecule that targets r(CGG)^{exp} and improves defects in fragile X-associated tremor ataxia syndrome. ACS Chem. Biol. 2012; 7:1711–1718. [PubMed: 22948243]
34. Luzhkov VB, et al. Virtual screening and bioassay study of novel inhibitors for dengue virus mRNA cap (nucleoside-2'O)-methyltransferase. Bioorg. Med. Chem. 2007; 15:7795–7802. [PubMed: 17888664]
35. Xu S, Witmer PD, Lumayag S, Kovacs B, Valle D. MicroRNA (miRNA) transcriptome of mouse retina and identification of a sensory organ-specific miRNA cluster. J. Biol. Chem. 2007; 282:25053–25066. Epub 22007 Jun 25027. [PubMed: 17597072]
36. Stenvang J, Petri A, Lindow M, Obad S, Kauppinen S. Inhibition of microRNA function by antimiR oligonucleotides. Silence. 2012; 3:1. [PubMed: 22230293]
37. Xie L, et al. FOXO1 is a tumor suppressor in classical Hodgkin lymphoma. Blood. 2012; 119:3503–3511. [PubMed: 22343918]
38. Guttilla IK, White BA. Coordinate regulation of *FOXO1* by miR-27a, miR-96, and miR-182 in breast cancer cells. J. Biol. Chem. 2009; 284:23204–23216. [PubMed: 19574223]
39. Dansen TB, Burgering BM. Unravelling the tumor-suppressive functions of FOXO proteins. Trends Cell. Biol. 2008; 18:421–429. [PubMed: 18715783]
40. Huang H, Tindall DJ. FOXO factors: a matter of life and death. Future Oncol. 2006; 2:83–89. [PubMed: 16556075]
41. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005; 120:15–20. [PubMed: 15652477]
42. Tran T, Disney MD. Identifying the preferred RNA motifs and chemotypes that interact by probing millions of combinations. Nat. Commun. 2012; 3:1125. [PubMed: 23047683]
43. Hawkins PC, Skillman AG, Nicholls A. Comparison of shape-matching and docking as virtual screening tools. J. Med. Chem. 2007; 50:74–82. [PubMed: 17201411]
44. Han J, et al. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell. 2006; 125:887–901. [PubMed: 16751099]
45. Berezikov E, et al. Phylogenetic shadowing and computational identification of human microRNA genes. Cell. 2005; 120:21–24. [PubMed: 15652478]
46. Michlewski G, Guil S, Semple CA, Caceres JF. Posttranscriptional regulation of miRNAs harboring conserved terminal loops. Mol. Cell. 2008; 32:383–393. [PubMed: 18995836]
47. Lunse CE, et al. An aptamer targeting the apical-loop domain modulates pri-miRNA processing. Angew. Chem. Int. Ed. Engl. 2010; 49:4674–4677. [PubMed: 20533473]
48. Zeng Y, Cullen BR. Sequence requirements for micro RNA processing and function in human cells. RNA. 2003; 9:112–123. [PubMed: 12554881]
49. Wolkenberg SE, Boger DL. Mechanisms of in situ activation for DNA-targeting antitumor agents. Chem. Rev. 2002; 102:2477–2495. [PubMed: 12105933]
50. Kramer R, Cohen D. Functional genomics to new drug targets. Nat. Rev. Drug. Discov. 2004; 3:965–972. [PubMed: 15520818]
51. Bevilacqua JM, Bevilacqua PC. Thermodynamic analysis of an RNA combinatorial library contained in a short hairpin. Biochemistry. 1998; 37:15877–15884. [PubMed: 9843393]
52. McKenna SA, et al. Purification and characterization of transcribed RNAs using gel filtration chromatography. Nat. Protoc. 2007; 2:3270–3277. [PubMed: 18079727]
53. Peyret N, Seneviratne PA, Allawi HT, SantaLucia J Jr. Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches. Biochemistry. 1999; 38:3468–3477. [PubMed: 10090733]
54. SantaLucia J Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc. Natl. Acad. Sci. U. S. A. 1998; 95:1460–1465. [PubMed: 9465037]
55. Puglisi JD, Tinoco I Jr. Absorbance melting curves of RNA. Methods Enzymol. 1989; 180:304–325. [PubMed: 2482421]

56. Wang Y, Rando RR. Specific binding of aminoglycoside antibiotics to RNA. Chem. Biol. 1995; 2:281–290. [PubMed: 9383430]

57. Disney MD, Gryaznov SM, Turner DH. Contributions of individual nucleotides to tertiary binding of substrate by a *Pneumocystis carinii* group I intron. Biochemistry. 2000; 39:14269–14278. [PubMed: 11087376]

58. Landthaler M, Yalcin A, Tuschl T. The human DiGeorge syndrome critical region gene 8 and its *D. melanogaster* homolog are required for miRNA biogenesis. Curr. Biol. 2004; 14:2162–2167. [PubMed: 15589161]

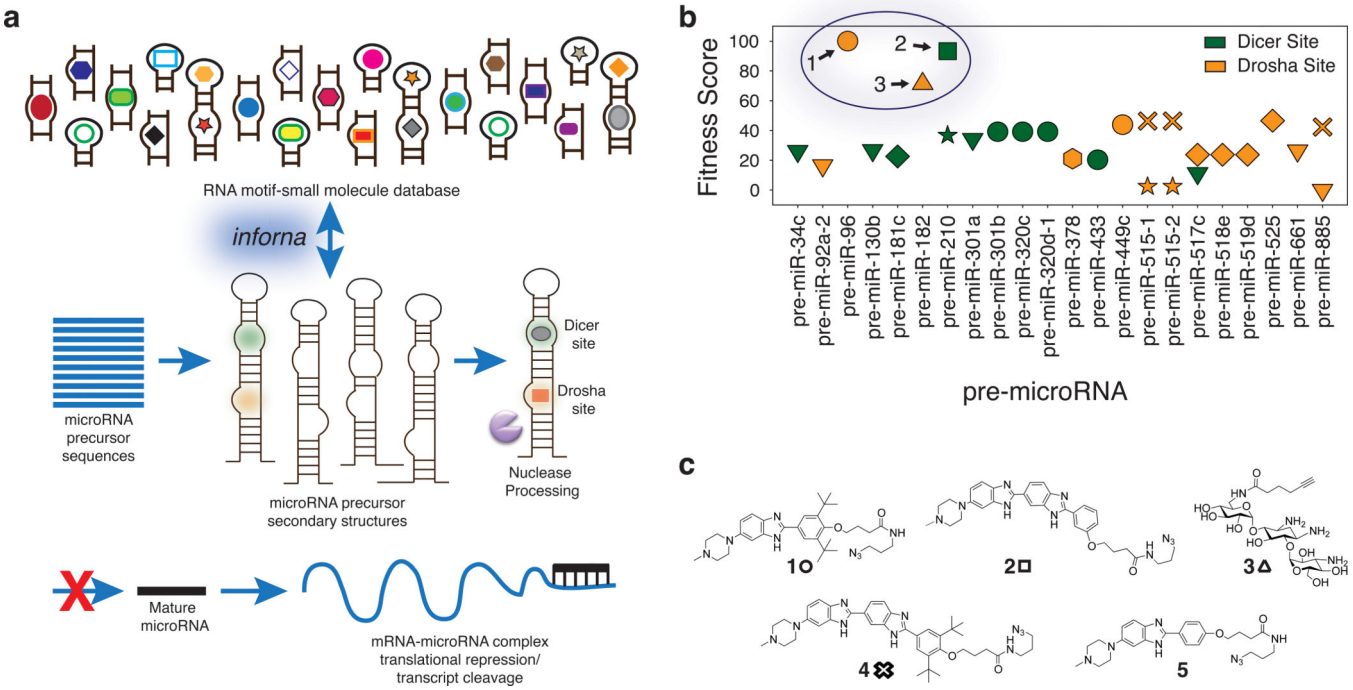

Figure 1.

The Inforna approach to design small molecules that target RNA from sequence applied to human miRNA precursors. **(a)**, sequences of all miRNA precursors in the human transcriptome were downloaded from miRBase[20] and their secondary structures were predicted via RNAstructure[12]. Inforna then extracted the secondary structural elements (motifs) from each RNA and compared them to a database of RNA motif-small molecule interactions identified by 2DCS. **(b)** and **(c)**, the results of Inforna mining of miRNA precursor-small molecule interactions. (b)a plot of the miRNA precursors for which a small molecule was predicted to bind Dicer or Drosha processing sites as a function of Fitness Score. Higher Fitness Scores indicate a higher affinity interaction. (c)structures of small molecules **1, 2, 3**, and **4** that bind to processing sites in miRNA precursors. Compound **5** is structurally similar to compound **1** and was employed in various assays. Structures of the other small molecules indicated in panel B can be found in Supplementary Fig. 3: **6** – 6'-*N*-5-hexynoate neamine (upside down triangle); **7** – 5-O-(2-azidoethyl)-neamine (star); **8** – 6"-azido-tobramycin (hexagon); **9** – 5"-azido-neomycin B (diamond).

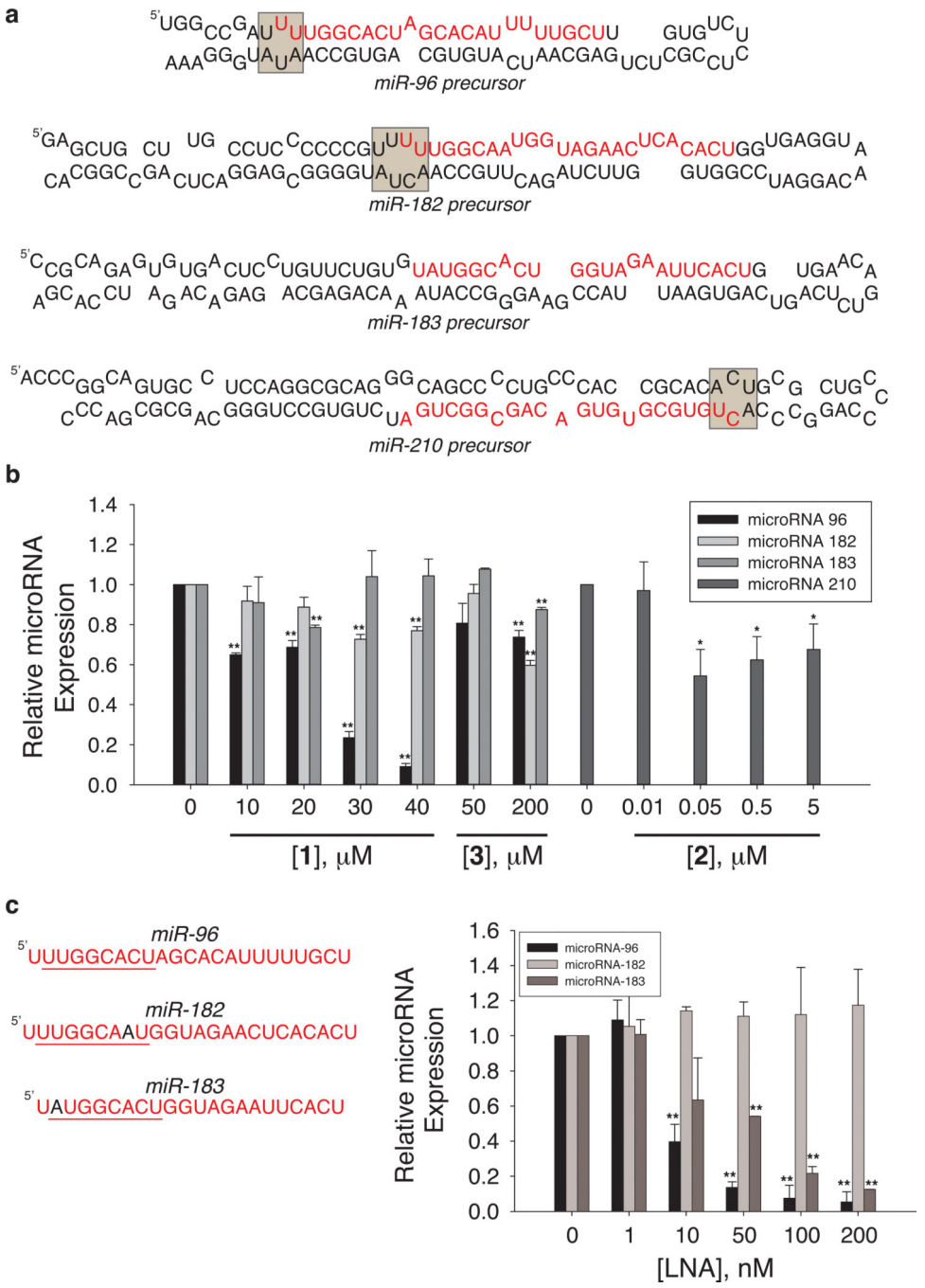

Figure 2.

Designer small molecules targeting miRNA precursors identified via Inforna are bioactive. (a) secondary structures of the miRNA precursors studied. Compounds **1, 2**, and **3** bind the regions indicated with gray boxes in miR-96, miR-182, and miR-210 precursors, respectively. MiR-96, −182, and −183 are transcribed as a cluster. (b)results of qRT-PCR analysis for production of mature miRNAs showing that the designer compounds modulate biogenesis to varying extents and with varying selectivities. **(c)** Left, sequence of the mature miRNAs that were studied. The underlined regions indicate designed oligonucleotide

binding sites; mismatches are shown in black. Right, qRT-PRC data for miR-96, miR-182, and miR-183 after treatment with an LNA oligonucleotide targeting miR-96. The LNA is unable to discriminate between miR-96 and miR-182 at any concentration tested. In contrast, **1** selectively silences miR-96 (panel B). "*" indicates $p<0.05$ while "**" indicates $p<0.01$ as determined by a two-tailed student *t*-test ($n\geq3$). Error bars are the standard deviations in the measurements.

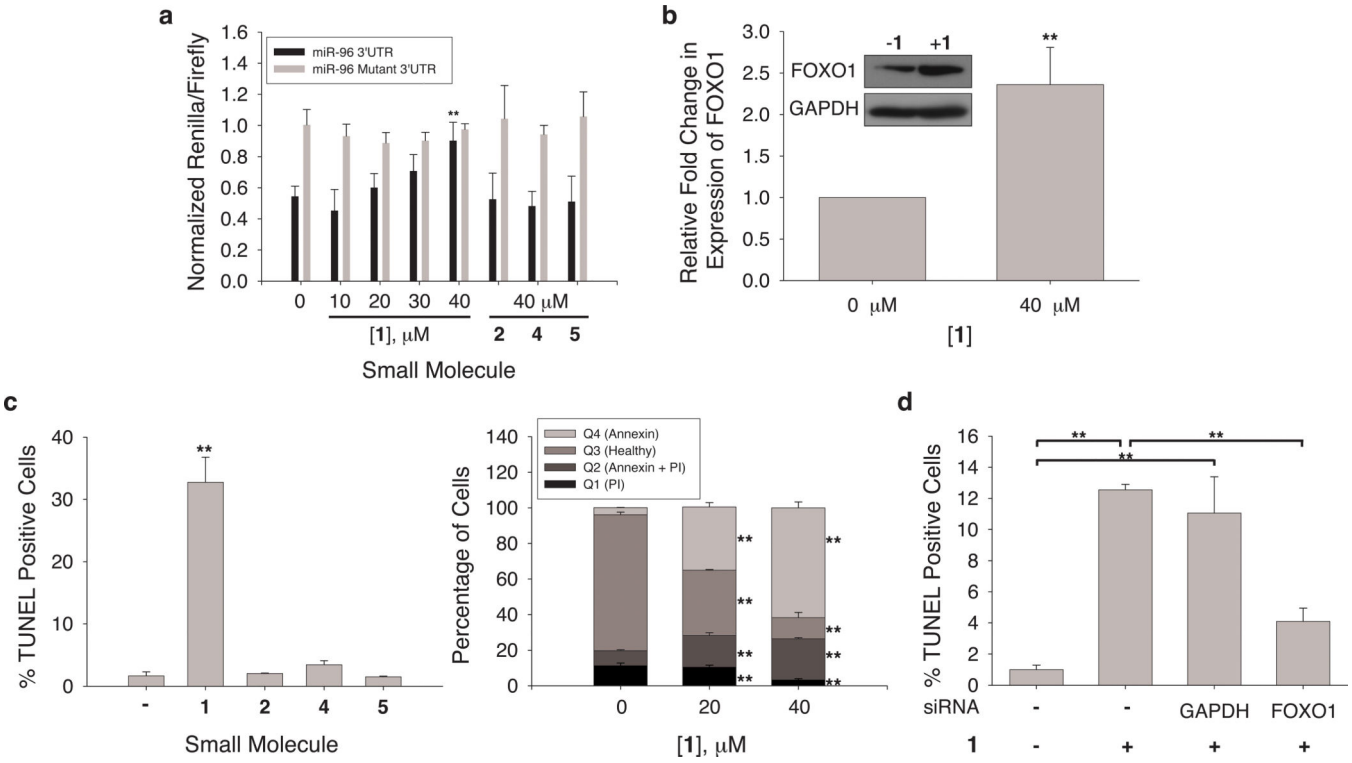

Fig. 3.

Effects of small molecules on the downstream targets of miR-96. (a) a luciferase model system was used to monitor the effect of **1** on FOXO1 expression, which is negatively regulated by miR-96. Small molecules that inhibit miR-96 maturation increase luciferase activity. Compound **1** increases luciferase activity by ~1.7-fold when MCF7 cells are treated with 40 μM compound. No effect on luciferase expression is observed when the miR-96 seed region is mutated such that it is unresponsive to miR-96. Compounds **2, 4**, and **5** have no effect on luciferase expression. (b)**1** increases endogenous FOXO1 protein expression as determined by Western blot. (c) **1** induces apoptosis as determined by TUNEL and Annexin V/PI assays (see also Supplementary Fig. 9). Since FOXO1 expression induces apoptosis, a small molecule that inhibits miR-96 maturation (thus increasing FOXO1 expression) should induce apoptosis. As expected, **1** stimulates apoptosis when MCF7 cells are treated with 40 μM compound while **2, 4**, and **5** do not (also at 40 μM). (d)to confirm that **1** induces apoptosis via increasing expression of FOXO1 protein, RNAi was used to knock down expression of *FOXO1* mRNA. If the expression of another protein is affected by **1**, then treatment should still induce apoptosis. Compound **1**'s ability to induce apoptosis was reduced by ~70% upon *FOXO1* siRNA treatment. Thus, its mode of action is via inhibition of pre-miR-96 maturation and concomitant induction of FOXO1 expression. "**" indicates $p<0.01$ as determined by a two-tailed student *t*-test ($n^3 3$). Error bars are the standard deviations in the measurements.

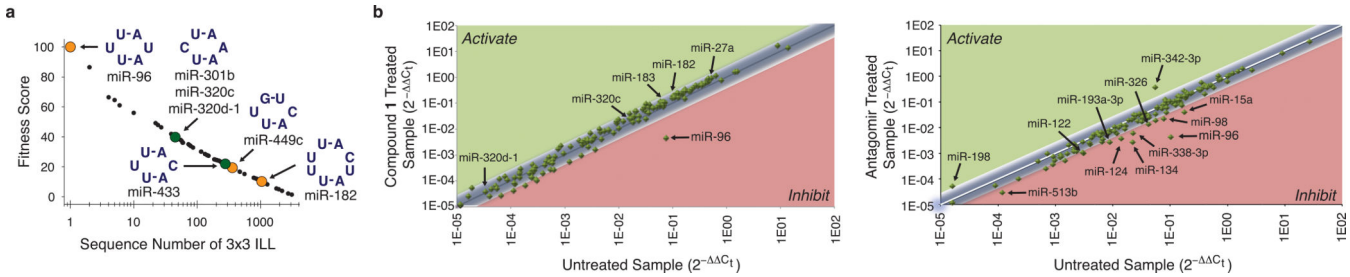

Figure 4.

Profiling of disease-associated miRNAs that are affected upon addition of 40 μM of compound **1** or a miR-96 antagomir. (a)Fitness Plot for compound **1** in which potential off-targets are indicated. Orange circles indicate Drosha sites; green circles indicate Dicer sites. (b)plot of the change in expression of 149 miRNAs after treatment with **1**. Amongst 149 miRNAs, only the production of mature miR-96 is significantly affected. (c)plot of the change in expression of 149 miRNAs after treatment with a miR-96 antaogmir. The expression of 12 miRNAs significantly change (≥2.5-fold) upon tranfection of the antagomir. These studies demonstrate the selectivity of the small molecule for the intended target and pathway.