5416–5425 *Nucleic Acids Research, 2014, Vol. 42, No. 9* *Published online 13 March 2014*
doi: 10.1093/nar/gku202

miR-Synth: a computational resource for the design of multi-site multi-target synthetic miRNAs

Alessandro Laganà[1,*,†], Mario Acunzo[1,†], Giulia Romano[1], Alfredo Pulvirenti[2], Dario Veneziano[1,2], Luciano Cascione[3], Rosalba Giugno[2], Pierluigi Gasparini[1], Dennis Shasha[4], Alfredo Ferro[2] and Carlo Maria Croce[1]

[1]Department of Molecular Virology, Immunology and Medical Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, OH, 43210 USA, [2]Department of Clinical and Molecular Biomedicine, University of Catania, 95100 Italy, [3]IOR-Institute of Oncology Research, Bellinzona, 6500 Switzerland and [4]Department of Computer Science, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012 USA

Received September 13, 2013; Revised February 20, 2014; Accepted February 25, 2014

ABSTRACT

RNAi is a powerful tool for the regulation of gene expression. It is widely and successfully employed in functional studies and is now emerging as a promising therapeutic approach. Several RNAi-based clinical trials suggest encouraging results in the treatment of a variety of diseases, including cancer. Here we present miR-Synth, a computational resource for the design of synthetic microRNAs able to target multiple genes in multiple sites. The proposed strategy constitutes a valid alternative to the use of siRNA, allowing the employment of a fewer number of molecules for the inhibition of multiple targets. This may represent a great advantage in designing therapies for diseases caused by crucial cellular pathways altered by multiple dysregulated genes. The system has been successfully validated on two of the most prominent genes associated to lung cancer, c-MET and Epidermal Growth Factor Receptor (EGFR). (See http://microrna.osumc.edu/mir-synth).

INTRODUCTION

Many diseases, such as cancer and neurological pathologies, occur as the result of multiple alterations in genes which are part of crucial cellular pathways.

Up to the present day, drug development has generally been focused on therapeutical targeting of individual genes or gene products. This strategy, however, has proven to be limited because the inhibition of single molecules may not be sufficient to effectively counteract disease progression and often leads to drug resistance with consequent relapse. In light of this evidence, the focus of drug therapy may need to shift from single- to multi-target approaches (1).

This approach is further justified by the fact that most cancers reflect a dysfunctionality in multiple pathways and an accumulation of new oncogenic mutations as the disease progresses. Thus, a valid strategy can come from targeting multiple genes involved in altered pathways rather than single genes, potentially assuring greater and more durable therapeutic benefits (2).

RNAi is now emerging as a promising therapeutic approach (3,4). Selective gene silencing through small interfering RNAs is widely and successfully employed in functional studies and is currently being investigated as a potential tool for the treatment of various diseases, including cancer, skin diseases and viral infections. siRNA, shRNA and their optimized chemical modifications are the active silencing agents and are intended to target single mRNAs in a specific way (5).

Several ongoing and already completed RNAi-based clinical trials suggest encouraging results (6). siRNA-mediated cleavage of a target mRNA, with a consequent reduction of protein expression level, was obtained in the first in-human phase I clinical trial in which siRNA were administered systemically to solid cancer patients (4).

The goal of targeting multiple genes and disrupting complex signaling pathways can be reached by co-expression of multiple siRNA or shRNA which enable multiple target inhibition, along with the targeting of multiple sites on a specific gene (7).

An important experiment in antiviral therapy research has shown that stable expression of a single shRNA targeting the HIV-1 Nef gene strongly inhibits viral replication, but the shRNA does not maintain such inhibition due to mutation or deletion of the nef target sequence which allows the virus to escape. A delay in virus escape is observed

*To whom correspondence should be addressed. Tel: +1 614 292 7278; Fax: +1 614 292 3558; Email: alessandro.lagana@osumc.edu
Correspondence may also be addressed to Carlo Maria Croce. Tel: +1 614 292 4930; Fax: +1 614 292 3558; Email: carlo.croce@osumc.edu
†These authors equally contributed to the work.

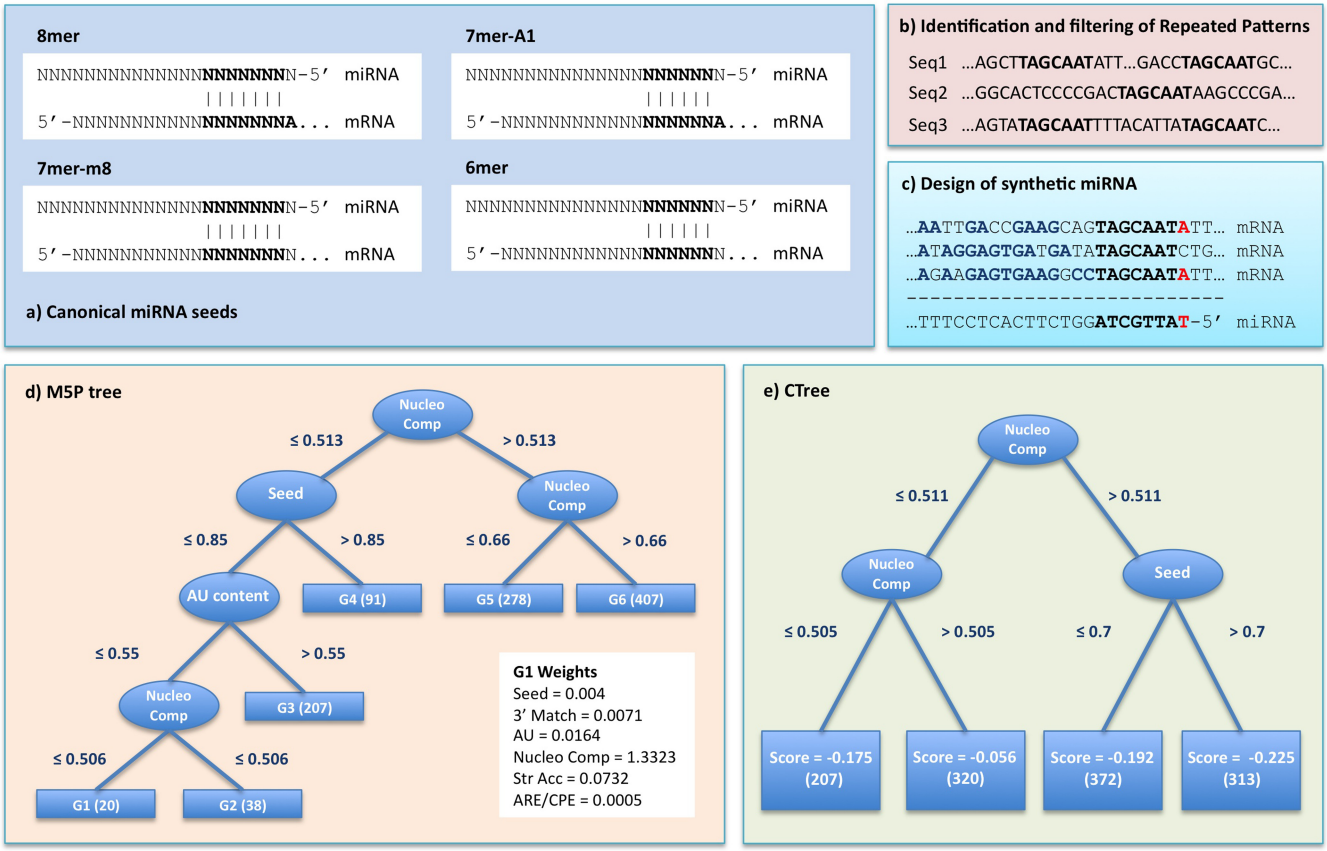

Figure 1. (**a**) The four different kinds of canonical miRNA seeds are depicted. They all share a 6mer core (bases 2–7). 7mer-A1 sites feature an A opposite of the first base of the miRNA, 7mer-m8 sites are full 7mers (bases 2–8) and 8mer sites are 7mer-m8 with an A opposite of the first base of the miRNA. (**b**) Input sequences are screened for repeated 6mer/7mer subsequences that will constitute the binding sites for the synthetic miRNA seeds. (**c**) Repeated patterns are used as anchors for the alignment of the binding sites of synthetic miRNAs. miRNAs are designed by maximizing complementarity to the consensus target sequence (see also Supplementary Figure S1). Target bases complementary to miRNA bases are indicated in blue and the seed match is indicated in red. (**d**) The tree generated by the learning system M5P. Six different sets of weights for the six considered features are calculated based on the values of the three discriminant features 'seed type', 'nucleotide composition' and 'AU content'. The white box contains the set of weight G1. See Supplementary Table S1 for the complete list of weight's sets. (**e**) The tree generated by the learning system CTree. The system assigns each miRNA to one of four different score classes, based on the values of the discriminant features 'seed type' and 'nucleotide composition'.

instead in HIV-1 infected cells that were previously transduced with a double shRNA viral vector (8).

Optimizations for co-expression of siRNA have also been proposed. In a recent work, dual-targeting siRNA with two active strands were specifically designed to target distinct mRNA transcripts with complete complementarity. This resulted in easier RISC entry since only two strands, instead of four, were competing for it (9).

An alternative approach for targeting multiple genes is suggested by the endogenous microRNA (miRNA) way of action (10). miRNAs, indeed, are naturally intended to target multiple genes, often in multiple sites, due to the partial complementarity they exhibit to their targets (11). This strategy would also enjoy the advantage that comes from involving fewer number of molecules.

In light of these considerations, we have developed miR-Synth, a bioinformatics tool available through a web interface for the design of synthetic miRNAs able to target multiple genes in multiple sites. We have validated our system by designing and testing single- and double-target miRNAs for two of the most prominent genes associated with lung cancer, c-MET and EGFR. A scoring function ranks the designed miRNAs according to their predicted repression efficiency.

Our experimental validations of the scoring function show that a downregulation of up to 70% was obtained by top-ranking miRNAs. miR-Synth is available at http://microrna.osumc.edu/mir-synth.

MATERIALS AND METHODS

Cell culture, transfection and chemicals

HeLa and HEK-293A cells were seeded and grown in Roswell Park Memorial Institute medium (RPMI) (HEK-293A) or Dulbecco's modified Eagle's medium (DMEM) (HeLa) with 10% fetal bovine serum (FBS), L-glutamine and antibiotics (Invitrogen, Carlsbad, CA, USA). All the transfections were performed by using Lipofectamine 2000 (Invitrogen) as suggested by the manufacturer. HEK-293A cells transfection for luciferase assay is described below. HeLa cells were cultured to 80% confluence in p60 plates with a serum-free medium without antibiotics, trans-

fected with 100 nmol of artificial miRNA (a-miR) oligonucleotides or negative control and harvested after 48 h.

Western-blot analysis

HeLa cells were seeded and grown in DMEM with 10% FBS in six-well plates for 24 h before the transfection. 48 h after transfection, cells were washed with cold phosphate-buffered saline and subjected to lysis in lysis buffer (50 mM Tris-HCl, 1 mM EDTA, 20 g/l SDS, 5 mM dithiothreitol, 10 mM phenylmethylsulfonyl fluoride). Equal amounts of protein lysates (50 mg each) and rainbow molecular weight marker (Bio-Rad Laboratories, Hercules, CA, USA) were separated by 4–20% SDS–PAGE and then electrotransferred to nitrocellulose membranes. The membranes were blocked with a buffer containing 5% non-fat dry milk in Tris-buffered saline with 0.1% Tween-20 for 2 h and incubated overnight with antibodies at 4°C. After a second wash with Tris-buffered saline with 0.1% Tween 20, the membranes were incubated with peroxidase-conjugated secondary antibodies (GE Healthcare, Amersham, Pittsburg, PA, USA) and developed with an enhanced chemiluminescence detection kit (Pierce, Rockford, IL, USA).

Antibody used for western-blot analysis

β-Actin (Sigma) was used as a loading control. MET and EGFR antibodies were from Cell Signaling Technologies.

RNA extraction

Total RNA was extracted with TRIzol solution (Invitrogen, Carlsbad, CA, USA), according to the manufacturer's instructions.

Q-real-time PCR

For the detection of single-target a-miRs, we performed quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR) by using a standard TaqMan PCR Kit protocol on an Applied Biosystem 7900HT Sequence Detection System (Applied Biosystems, Carlsbad, CA, USA). For the TaqMan qRT, the 10 ml PCR reaction included 0.67 ml RT product, 1 ml TaqMan Universal PCR Master Mix (Applied Biosystems, Carlsbad, CA, USA), 0.2 mM TaqMan probe, 1.5 mM forward primer and 0.7 mM reverse primer. The reactions were incubated in a 96-well plate at 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. All reactions ran in triplicate. The threshold cycle (Ct) is defined as the fractional cycle number at which the fluorescence passes the fixed threshold. The comparative Ct method for relative quantization of gene expression (Applied Biosystems, Carlsbad, CA, USA) was used to determine a-miR expression levels. The y-axis represents the relative expression of the different a-miRs (Figures 2c, 3c and 4c). a-miR expression was calculated relative to U44 and U48 rRNA. Experiments were carried out in triplicate for each data point, and data analysis was performed by using software tools (Bio-Rad Laboratories, Hercules, CA, USA).

It was not possible to synthesize TaqMan custom primers for the detection of the a-miRs targeting both EGFR and c-MET so we performed SYBR Green PCR assay. cDNA was obtained from 1 μg of total RNA with miScript II (Qiagen, Venlo, The Netherlands) using random primers. cDNA was then treated with RNase H, in order to remove any residual RNA. For SYBR Green qPCR, amplification of cDNA was performed with SYBR Green PCR kit (Qiagen, Venlo, The Netherlands), and normalized using the $2^{-\Delta Ct}$ method to U6 rRNA. The y-axis represents the relative expression of the different a-miRs.

Luciferase assay

We used the luciferase reporter constructs described in other works (12,13). Mutations in a-miR binding sites in MET and EGFR constructs were introduced by using the QuikChange Mutagenesis Kit (Stratagene, La Jolla, CA, USA). HEK-293A cells were transfected with Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA), 1.2 mg of pGL3control containing EGFR, MET or MET and EGFR mutants, 200 ng of Renilla luciferase expression construct. After 24 h, cells were lysed and assayed with Dual Luciferase Assay (Promega) according to the manufacturer's instructions. Mutagenesis' primers are reported in Supplementary Table S9.

Development of the design tool and web interface

The miR-Synth design tool was written in Ruby v1.9.3. The program uses the external software tools RNAplfold from the Vienna RNA Package v1.8.4 for the computation of the structural accessibility and the statistical package R v3.0.1 for the computation of the conditional inference trees (CTree) score. R is executed from the Ruby script by using the gem rinruby. The M5P weights and scores were computed by using the M5P implementation available in the software tool Weka v3.7.9.

The miR-Synth web interface was developed in Ruby on Rails v2.3.5, a framework based on the MVC (Model-View-Controller) design pattern. All transcript sequence data for the species provided by miR-Synth, along with all user specified data, are collected and maintained in a MySQL database v5.1 running on an Apache server v2.2.15. The queries that the database allows to perform were coded leveraging on the association mechanisms between models that the framework provides. The interface makes use of the jQuery v1.7 technology to improve the usability through a fast and agile client-side update of selections and results.

RESULTS

The miR-Synth algorithm and the design features

miR-Synth is a tool for the design of a-miRs for the repression of single or multiple targets. The problem of designing effective a-miRs is strictly connected to the prediction of miRNA binding sites. The main issue is that target prediction tools yield many false positives (14). Nevertheless, the remarkable progress made in recent years has identified key features to characterize miRNA functional target sites.

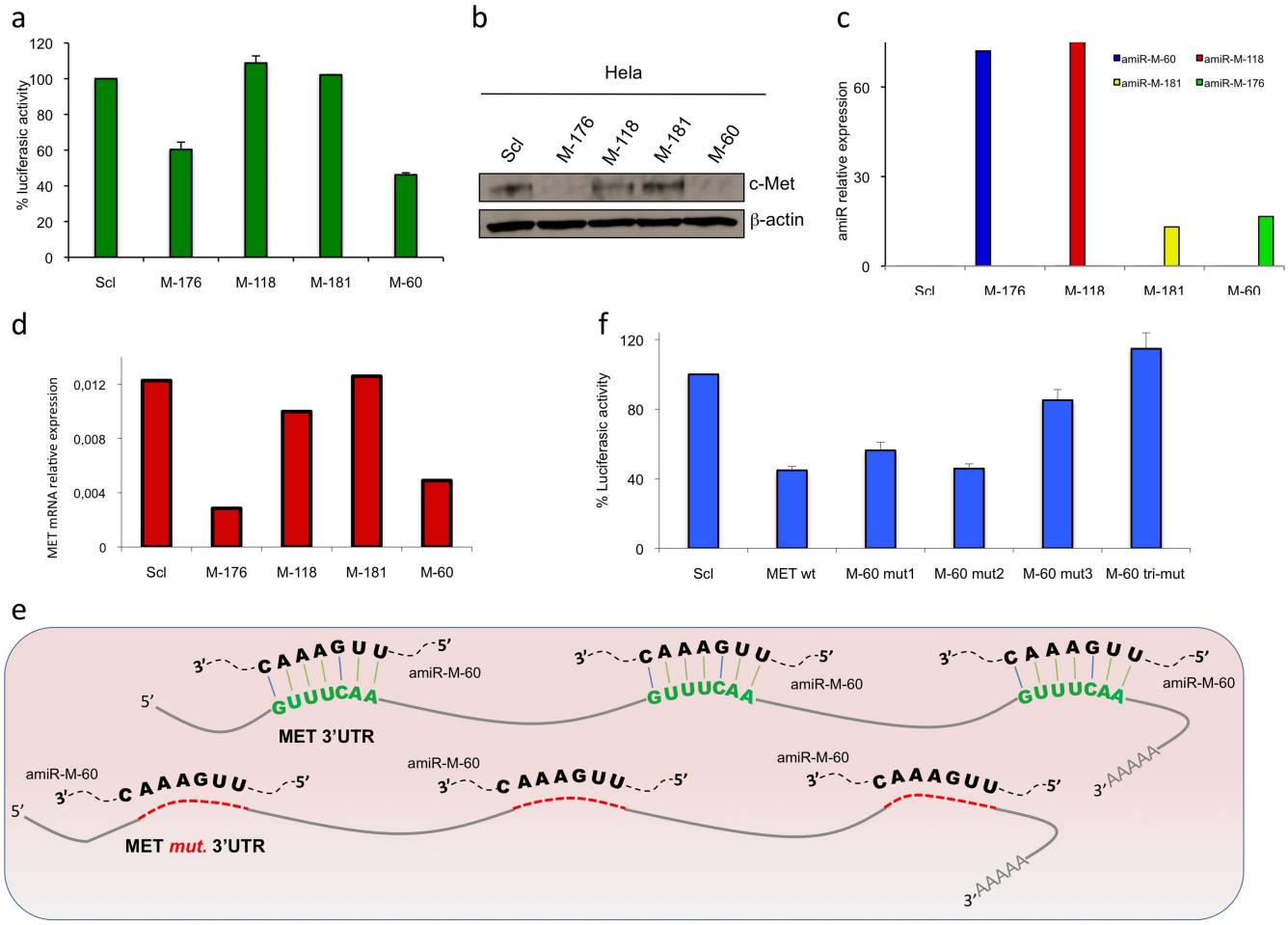

Figure 2. (**a**) pGL3-MET 3′UTR construct was co-transfected with a-miRs or negative control in HEK-293A cells and luciferase assay was performed (error bars: ± SEM, $P < 0.05$). (**b**) c-MET expression was assessed by western blot in HeLa cells transfected with a-miRs or negative control and harvested after 72 h. a-miR-M-60 and a-miR-M-176 enforced expression decreases endogenous levels of the c-MET protein. Loading control was obtained by using anti-β-actin antibody. (**c**) qRT-PCR of the transfected a-miRs in HeLa cells. (**d**) qRT-PCR of the c-MET mRNA after a-miRs enforced expression in HeLa cells. (**e**) Representation of the c-MET 3′UTRs binding sites for a-miR-M-60. In the figure, pairing of the seed region of a-miR-M-60 with the three c-MET binding sites is shown. The deleted binding sites are indicated in red. (**f**) c-MET 3′UTR is a target of a-miR-M-60. pGL3-MET luciferase wild-type and mutated constructs were co-transfected with a-miR-M-60 or negative control in HEK-293A cells and luciferase assay was performed (error bars: ± SEM, $P < 0.05$).

We have combined well-established knowledge on miRNA targeting together with siRNA design rules and empirical observations on validated miRNA/target interactions into a pipeline which consists of three steps: (i) identification and filtering of repeated patterns; (ii) design and filtering of a-miR sequences; and (iii) scoring and ranking of the designed a-miRs.

The first step mainly relies on the concept of miRNA seed, which is the 5′ region of the miRNA, centered on nucleotides 2–7 (Figure 1a). The miRNA seed is the most conserved portion of metazoan miRNAs and allows the characterization of miRNA families. The seed generally matches complementary, often conserved, canonical sites on the 3′UTRs (UnTranslated Regions) of regulated targets (11,15). There is evidence that the lack of perfect seed pairing in functional binding sites is, at times, balanced by the presence of centered or 3′ compensatory sites. However, these cases are much less abundant than canonical sites which represent the predominant interaction model associated with greater target repression. Among canonical sites, 7mer-m8 and 8mer sites yield the strongest repression, while 6mer sites are associated with mild to very mild efficacy. In order to achieve a significant repression of the targets, we have chosen to consider only canonical sites, especially favoring 7mer-m8 and 8mer matches.

In order to estimate the number of human 3′UTR sequences that share at least a common 7nt pattern, we collected gene expression data associated with distinct diseases from the Gene Expression Atlas (GEA) (16) and focused on the upregulated genes, thus mimicking a plausible scenario for the employment of a-miRs. For each disease, we calculated all the possible combinations of two and three upregulated genes and counted how many of them share at least a 7mer 3′UTR site. We filtered out polyA-signal motifs, homopolymer motifs and sites matching the seeds of endogenous miRNAs. We performed this analysis on all upregulated gene pairs and triplets as detected in 83 different diseases, revealing that 97.3% of pairs and 81.32% of triplets

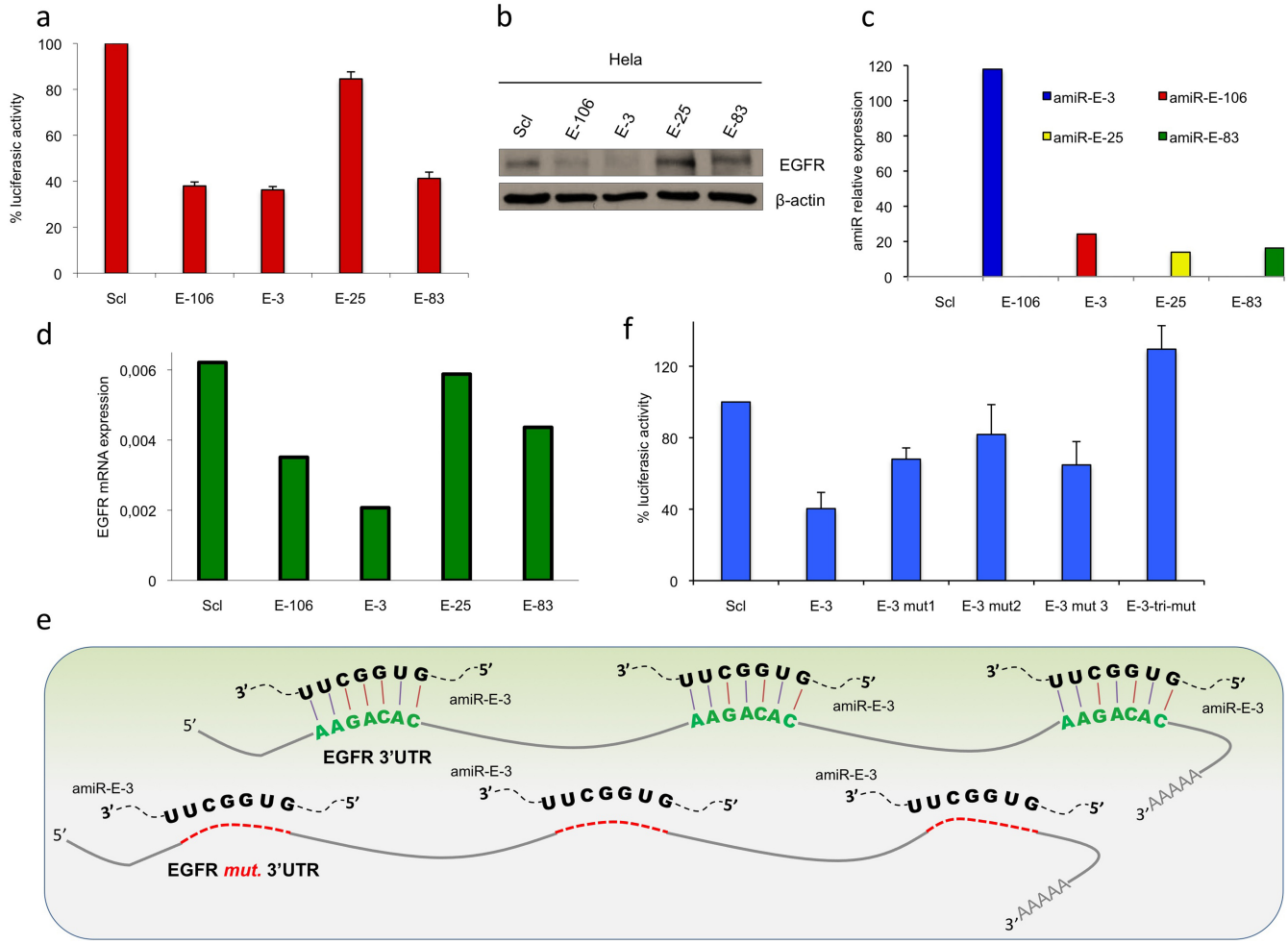

Figure 3. (**a**) pGL3-EGFR 3′UTR construct was co-transfected with a-miRs or negative control in HEK-293A cells and luciferase assay was performed (error bars: ± SEM, $P < 0.05$). (**b**) EGFR expression was assessed by western blot in HeLa cells transfected with a-miRs or negative control and harvested after 72 h. a-miR-E-3 and a-miR-E-106 enforced expression decreases endogenous levels of the EGFR protein. Loading control was obtained by using anti-β-actin antibody. (**c**) qRT-PCR of the transfected a-miRs in HeLa cells. (**d**) qRT-PCR of the EGFR mRNA after a-miRs enforced expression in HeLa cells. (**e**) Representation of the EGFR 3′UTRs binding sites for a-miR-E-3. In the figure pairing of the seed region of a-miR-E-3 with the three EGFR binding sites is shown. The deleted binding sites are indicated in red. (**f**) EGFR 3′UTR is target of a-miR-E-3. pGL3-EGFR luciferase wild-type and mutated constructs were co-transfected with a-miR-E-3 or negative control in HEK-293A cells and luciferase assay was performed (error bars: ± SEM, $P < 0.05$).

share at least one 7mer site. On average, pairs and triplets shared about 136 and 24 7mer sites, respectively (see Supplementary Section S2 for details and additional analysis). In light of this, and considering cases in which a set of highly similar sequences is chosen for targeting, we decided to set a maximum threshold of eight target sequences that users can provide as input to the system. This limitation makes sense, because eight is already a considerable number of targets, unlikely to be practical in most applications. These sequences are screened for repeated patterns of six or seven nucleotides (depending on user choice), which will constitute the binding sites for a-miR seeds (Figure 1b). These sites are then filtered based on user-provided specifications, e.g. a site must appear in multiple copies on the same target and/or it must be present at least once in every target. Moreover, users can also provide a list of sequences that must not be targeted. In this case, the system will remove all the sites that appear at least once in any of the provided sequences (Table 1).

The second step of the algorithm consists of the actual a-miR sequence design. For each repeated pattern identified in the previous phase an anticomplementary a-miR seed is created. The rest of the sequence is constructed by aligning the seed's binding sites and maximizing the match outside the seed region through a sequence profile technique, as depicted in Figure 1c. The a-miR sequences thus obtained will be 22 nt long.

The designed a-miRs are then filtered based on their nucleotide composition, combining well-established siRNA design rules with endogenous miRNA features. In particular, sequences with GC content out of the user's specified range (23–78% by default) or containing stretches of six or more nucleotides of the same kind are discarded (17,18). These particular thresholds were chosen according

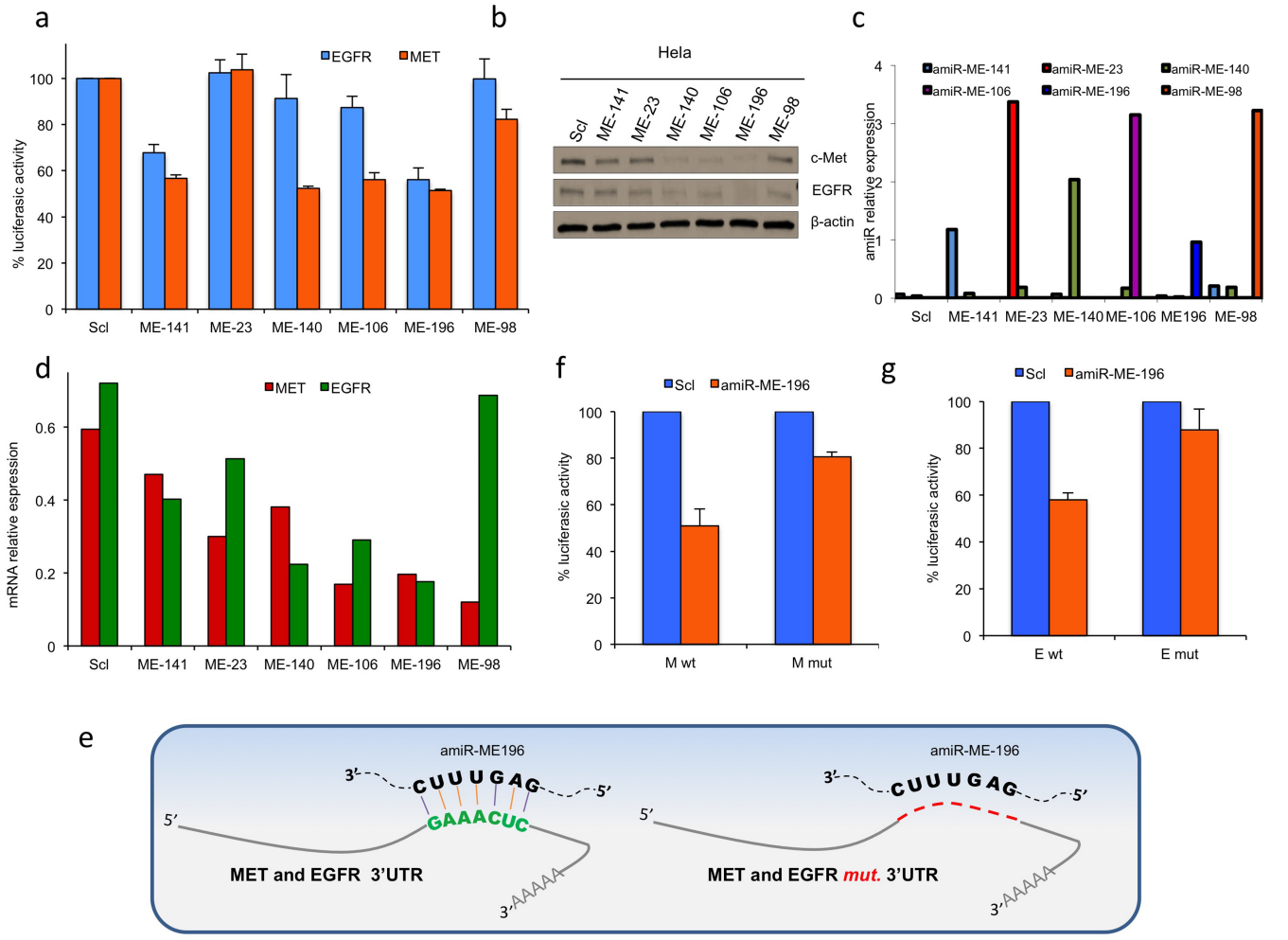

Figure 4. (**a**) pGL3-MET 3′UTR and pGL3-EGFR 3′UTR were co-transfected with a-miRs or negative control in HEK-293A cells and luciferase assay was performed (error bars: ± SEM, $P < 0.05$). (**b**) EGFR and c-MET expression was assessed by western blot in HeLa cells transfected with a-miRs or negative control and harvested after 72 h. Loading control was obtained using anti-β-actin antibody. (**c**) qRT-PCR of the transfected a-miRs in HeLa. (**d**) qRT-PCR of the c-MET and EGFR mRNA after a-miRs enforced expression in HeLa cells. (**e**) Representation of the c-MET and EGFR 3′UTRs binding sites for a-miR-ME-196. In the figure, pairing of the seed region of a-miR-ME-196 with the c-MET/EGFR binding site is shown. The deleted binding site is indicated in red. (**f**) MET 3′UTR is target of a-miR-ME-196. pGL3-MET luciferase wild-type and mutated constructs were co-transfected with a-miR-ME-196 or negative control in HEK-293A cells and luciferase assay was performed. (**g**) EGFR 3′UTR is a target of a-miR-ME-196. pGL3-EGFR luciferase wild-type and mutated constructs were co-transfected with a-miR-ME-196 or negative control in HEK-293A cells and luciferase assay was performed (error bars: ± SEM, $P < 0.05$).

to what has been observed in typical endogenous miRNA nucleotide composition (see Supplementary Section S2).

In this phase users can also choose to discard a-miR sequences sharing a seed with any endogenous miRNA. Moreover, users can enable the prediction of potential off-target genes. A filter allows the removal of those a-miRs whose seed is predicted to bind more than a user-provided maximum number of off-target genes. Alternatively, the user can request the top 10 a-miRs with the smallest number of off-target hits. This is an important feature, since a single a-miR may target even thousands of different genes. This issue will be further discussed in the Validation and Discussion sections. More details about the algorithm and the filters are given as supplementary information (see Supplementary Section S1).

Scoring and ranking of a-miRs

The third step of the miR-Synth pipeline consists in the evaluation and ranking of the designed a-miRs. We developed a scoring function based on six different features of validated endogenous miRNA/target interactions: seed type, pairing of the miRNA 3′ region, AU content of the binding site and its surrounding regions, miRNA nucleotide composition, structural accessibility of the binding site and presence of ARE (AU Rich Element) and CPE (Cytoplasmic Polyadenylation Element) motifs upstream of the binding sites (15,19–22). For any given a-miR, each feature is assigned a score ranging from 0 to 1 and a total repression score is calculated by combining the tree-based multiple linear regression learning system M5P with CTree (23,24).

We have trained the system on a set of publicly available gene expression profiles following the over-expression

Table 1. Details about the tested miRNAs

Synthetic miRNAs for c-MET						
Rank	**ID**	**Sequence**	**Sites**	**Seed types**	**M5P score**	**CTree score**
1	60	UUUGAAACGGAGGCUGUCUAGA	3	8mer/8mer/8mer	−0.261	−0.225
2	118	UUUAUAAAGUCGAUACGUGUUU	3	8mer/8mer/8mer	−0.260	−0.225
3	181	UUCUUUCUAAGGACGGGGCCGU	2	8mer/8mer	−0.253	−0.225
4	176	UCAGUACAAAACCUUGUGGCUU	2	8mer/8mer	−0.246	−0.225
Synthetic miRNAs for EGFR						
Rank	**ID**	**Sequence**	**Sites**	**Seed types**	**M5P score**	**CTree score**
1	3	UGUGGCUUCACCUCCUGUAUCG	3	8mer/8mer/7mer-m8	−0.241	−0.225
2	106	UGUGUGACACUGCGUAAGGGGG	2	8mer/8mer	−0.238	−0.225
3	25	CAAAUGCUCGAGAGUCCGAUGU	2	8mer/7mer-m8	−0.229	−0.225
4	83	UAACAAUGCACUGGGGGCCCUG	2	8mer/7mer-m8	−0.228	−0.225
Synthetic miRNAs for c-MET and EGFR						
Rank	**ID**	**Sequence**	**Sites**	**Seed types**	**M5P score**	**CTree score**
1	141	UUCCAAUUCGAGGGGAGGUGGG	1+1	8mer/8mer	−0.262	−0.225
2	23	UCAAUUUCGGUCCCGAGUUCCA	1+1	8mer/8mer	−0.258	−0.225
3	140	UCCAAUUGGACGGGAGGUGGGU	1+1	8mer/8mer	−0.249	−0.225
4	106	UUUCAUGAGCCCUAGACUGGGG	1+1	8mer/8mer	−0.246	−0.225
5	196	UGAGUUUCUCAGCGACGGACCG	1+1	8mer/8mer	−0.241	−0.225
6	98	UUUCUUAAGCACGCCGUUGGGG	1+1	8mer/8mer	−0.239	−0.225

More details are given as supplementary information (Supplementary Tables S2–S4).

of nine individual endogenous miRNAs (15). In particular, binding sites were predicted for each transfected miRNA on downregulated genes, then feature values were calculated. The gene expression fold change was used as a measure of the degree of repression induced by the miRNA. Thus, lower values mean stronger downregulation of the target. Only transcripts with single binding sites for the transfected miRNAs were considered, in order to reduce the chances of indirect effects.

According to the M5P tree (Figure 1d, Supplementary Table S2), the most discriminant features were the nucleotide composition of the miRNA, the type of seed and the AU content of the binding site.

Depending on the values of these three, six different sets of weights were assigned to all of the features. Only the seed type and the nucleotide composition of the miRNA were considered as discriminant features by CTree (Figure 1e).

These two methods are used to evaluate the designed a-miRs. In particular, a-miRs are first ranked according to the CTree score and subsequently by the M5P score. CTree splits the a-miRs into major classes, while M5P is used to rank a-miRs within each class.

We validated this scoring function by using a database of experimentally validated human miRNA/target interactions called miRTarBase as a test set (25). This dataset contains 495 cases of proven direct interactions, 490 cases for which direct binding was not verified and 71 negative cases. We considered 1000 randomly created groups with the same number of proven direct and proven negative cases. For each group, the top 10 interactions, as ranked by our approach, always contained a higher number of true direct interactions compared to sets of 10 cases randomly chosen ($P < 0.0001$). A more detailed description of the scoring features, classification and validation processes is given as supplementary information (see Supplementary Section S1).

Validation of single-target multi-site a-miRs

Our a-miR design system was validated on c-MET and EGFR, two well-known genes involved in lung cancer. This choice constitutes a good example of beneficial employment of multi-target a-miRs, given the reciprocal and complementary relationship between EGFR and c-MET in acquired resistance to kinase inhibitors in lung cancer, and the necessity of concurrent inhibition of both to further improve patient outcomes (26).

We designed two different sets of multi-site a-miRs exclusively targeting c-MET and EGFR, respectively. The system returned 111 a-miRs for c-MET and 59 a-miRs for EGFR (Supplementary Tables S6 and S7). For each of the two genes, we focused on the top four a-miRs as ranked by our scoring system (Supplementary Tables S3 and S4). Supplementary Table S1(a) and (b) summarizes the main features of these a-miRs. The eight a-miRs thus taken into consideration had at least two binding sites on their targets, with a predominant presence of 8mer matches. To verify direct targeting, the wild-type 3′UTRs of c-MET and EGFR were cloned into pGL3 control vectors downstream of the luciferase open reading frame. a-miRs for c-MET and EGFR were individually co-transfected with the c-MET and EGFR 3′UTR constructs, respectively, in HEK-293A cells. This resulted in a significant inhibition of the luciferase activity induced by two c-MET a-miRs and three EGFR a-miRs, as compared to the negative control (Figures 2a and 3a). Moreover, western-blot and qRT-PCR assays showed that over-expression of a-miRs in HeLa cells strongly reduced the endogenous protein and mRNA levels of c-MET and EGFR as compared to control (Figures 2b and d and 3b and d), in agreement with the luciferase assay results. Expression of transfected a-miRs in HeLa transfected cells was confirmed by qRT-PCR (Figures 2c and 3c). Among the five functional a-miRs, *a-miR-M-60* and *a-miR-E-3* ranked first and yielded strong downregulation of c-MET and EGFR 3′UTRs luciferase activity, respectively (Figures 2a and 3a). Hence, as further analysis, we performed mutagenesis of *a-miR-M-60* and *a-miR-E-3* binding sites within the MET and EGFR 3′UTRs, which abolished the ability of these a-miRs to regulate luciferase

expression, thus confirming that the binding sites are functional (Figures 2e and f and 3e and f).

Validation of multi-target synthetic a-miRs

We subsequently designed a-miRs intended to target both c-MET and EGFR concurrently. The algorithm returned a total of 125 a-miRs with 7mer-m8/8mer matches on the UTRs of both genes (Supplementary Table S8). We selected the top six a-miRs as ranked by our scoring function (Supplementary Tables S1(c) and S5). All of them had one 8mer binding site on each gene.

To verify multiple direct targeting of c-MET and EGFR, the designed a-miRs were individually co-transfected with both wild-type c-MET and EGFR 3′UTR constructs into HEK-293A cells. *a-miR-ME-196* and *a-miR-ME-141* induced a significant inhibition of the luciferase activity for both constructs, while *a-miR-ME-140* and *a-miR-ME-106* yielded a significant repression of c-MET only, as compared to the negative control (Figure 4a). Moreover, overexpression of the a-miRs in HeLa cells induced a strong repression of the endogenous c-MET and EGFR proteins and mRNAs in three cases and a mild downregulation in the three remaining cases, as compared to the control (Figure 4b and d). Interestingly, although not all tested a-miRs were functional at the luciferase level, the effects on the endogenous proteins, whose repression represents our primary goal, was much stronger. This could be due to the intrinsic limitations of the luciferase assay, being based on an artificial construct. Nevertheless, out of the six tested a-miRs, *a-miR-ME-196* was chosen for further investigation because of its greater downregulation at both the protein and the luciferase level (Figure 4a, b and d). The expression of *a-miR-ME-196* in HeLa transfected cells was confirmed by qRT-PCR (Figure 4c). Mutagenesis of the a-miR binding site within the c-MET and EGFR 3′UTRs eliminated its ability to regulate luciferase expression, thus confirming that the binding site is functional (Figure 4e and f). In order to further demonstrate the robustness of the miR-Synth scoring function and the additional benefits of incorporating features other than the seed match, we tested the bottom six a-miRs designed for c-MET and EGFR and found that three of these a-miRs yielded a mild repression of EGFR, lower than observed for the best top six a-miRs, and that none of them was able to significantly repress c-MET, despite their good seed matches (7mer/8mer) (Supplementary Figure S2).

On a final note, in order to assess the general applicability of our method, we additionally ran miR-Synth on 14,325 pairs of upregulated genes in eight diseases retrieved from the GEA dataset mentioned above. miR-Synth was able to design at least an a-miR for 95% of pairs and at least six a-miRs for 86.9% of pairs. The feature and global scores of the top six a-miRs from GEA were very comparable to the scores of the validated c-MET/EGFR top six a-miRs. In particular, this held true for features such as AU content and structural accessibility, which solely depend on the target sequence, thus confirming the results obtained with the 7mer analysis described above. However, when we applied the off-target filter, we found that only 43% of gene pairs shared at least a 7mer with no more than 2000 off-target hits, and the percentage dropped to 5.6% when we considered gene pairs sharing a 7mer with no more than 1000 off-targets. This is an intrinsic factor of any a-miR, due to the short length of the seed region, which with no doubt requires proper consideration. Our experiments showed, however, that a perfect seed match is not the only indicator of effectiveness and that other features must be taken into account. In light of all this, our off-target prediction analysis and filters constitute a useful tool to help the user select the best a-miRs. More details are given as supplementary information (see Supplementary Sections S3 and S4).

The miR-Synth web interface

miR-Synth is freely available for academic use through a web interface (http://microrna.osumc.edu/mir-synth). Users can provide up to eight UTR sequences or select them from a menu by their name, Refseq accession number or Entrez gene ID. Although the system was trained on human miRNAs, it allows selection of targets from other species as well, such as mouse and rat. Users can either request to design a-miRs simultaneously targeting all of the provided sequences or to include a-miRs targeting subsets of them as well. A list of sequences (or their IDs) that must not be directly targeted by the designed a-miRs can also be provided.

In the available options users can specify the kind of seed matches allowed (6mer and/or 7mer-m8/8mer), the GC% content range (default is 23–78%) and whether the endogenous miRNA filter should be applied. Sequence masks can also be provided, in order to specify portions of the input sequences that should not be targeted. This can be a useful option when the presence of SNPs (Single Nucleotide Polymorphism) or other mutations in the targets could negatively affect a-miR binding (27,28).

Finally, users can choose to view the list of potential off-target genes, which is obtained through the computation of seed matches on the whole database of UTR sequences from the selected species.

The system is fast. For example, the design of a-miRs for a pair of targets with default parameters takes 30 s at most. However, given the variability in the number of input sequences and the different options that can be selected, which could substantially increase computation time, users are provided with the results page link by e-mail once the computation has completed. For each individual a-miR, details about interaction features and their binding sites are given, including partial and global scores along with the list of off-target genes and the number of their potential binding sites, if requested.

Technical details about the development of the web interface are provided in the online methods.

DISCUSSION

RNAi constitutes a powerful tool for the regulation of gene expression (29,30). Recent progress in the development of increasingly efficient carriers for the intracellular delivery of small RNAs, such as nanoparticles and viral systems, has made the establishment of therapeutics based on this promising technology imaginable (31–33). Moreover, new strategies for oral delivery of antisense nucleotides and recent findings suggesting that exogenous miRNA, such as

those of plant origin, simply introduced through food intake could be active and functional in recipient cells, opens a new scenario in which RNAi could constitute an appealing and concrete therapeutical tool for cancer, viral infections and other diseases caused or progressively maintained by the over-expression of multiple genes (34–36).

Although the rules for the design of efficient siRNA and shRNA are nowadays well established, sequence design methodologies can nevertheless be further improved, especially to reduce off-target effects.

siRNAs are designed to regulate specific targets through perfect complementarity, but evidence shows that the presence of one or more perfect matches in 3′UTR sequences with the siRNA seed region is associated with considerable off-target effects and represents a widespread and unintended consequence of siRNA-mediated silencing (37,38). This phenomenon, which reflects the natural behavior of miRNAs, suggests a possible approach for designing fewer molecules that may reduce the expression of many targets. In fact, our experiments show that a single a-miR may be able to repress at least two unrelated genes at the same time, while it may likely take a pool of different siRNAs/shRNAs to obtain the significant inhibition of a single gene. It is very important to point out that, in principle, there is no difference between a single multi-target a-miR and a single-target siRNA in terms of basic seed-based off-target effects. Any very short nucleotide sequence, such as a 7mer, is likely to appear in a substantial number of UTRs. Unsurprisingly therefore, an *in silico* test confirmed that double-targeting a-miRs are likely to have fewer off-targets than pairs of single-targeting siRNAs (see Supplementary Section S3). This indicates a substantial advantage in the employment of a-miRs in place of siRNAs/shRNAs.

The miR-Synth pipeline allows the rational design of a-miRs by taking multiple factors into consideration. It integrates current knowledge regarding miRNA/target interaction and features simple yet powerful options which allow, for example, to investigate off-target effects and design molecules virtually not affected by SNPs and other polymorphisms.

Future work includes refinement of the design process and further analysis of miRNA/target interactions, in order to better understand the causal connection between the targeting features and the degree of downregulation, and improve the selection of effective molecules.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online, including [39–45].

ACKNOWLEDGMENTS

We thank Nicola Bombieri, Alessandro Danese and Francesco Martinelli for their precious help with the testing of miR-Synth. We thank the anonymous reviewers for their helpful comments and suggestions.

FUNDING

National Institutes of Health/National Cancer Institute [R01-CA135030, P01-CA081534 to A.L., U01-CA152758, U01-CA166905 to M.A.]; U.S. National Science Foundation [DBI-0445666, DBI-0421604, NSF IOS-0922738, MCB-0929339, N2010 IOB-0519985]; National Institutes of Health [2R01GM032877–25A1 to D.S.].

Conflict of interest statement. None declared.

REFERENCES

1. Petrelli,A. and Valabrega,G. (2009) Multitarget drugs: the present and the future of cancer therapy. *Expert Opin. Pharmacother.*, **10**, 589–600.
2. McCarty,M.F. (2004) Targeting multiple signaling pathways as a strategy for managing prostate cancer: Multifocal signal modulation therapy. *Integr. Cancer Ther.*, **3**, 349–380.
3. Sinha,G. (2012) Interest resparks in RNAi. *Nat. Biotechnol.*, **30**, 1012.
4. Davis,M.E., Zuckerman,J.E., Choi,C.H.J., Seligson,D., Tolcher,A., Alabi,C.A., Yen,Y., Heidel,J.D. and Ribas,A. (2010) Evidence of RNAi in humans from systemically administered siRNA via targeted nanoparticles. *Nature*, **464**, 1067–1070.
5. Rao,D.D., Vorhies,J.S., Senzer,N. and Nemunaitis,J. (2009) siRNA vs. shRNA: similarities and differences. *Adv. Drug Deliv. Rev.*, **61**, 746–759.
6. Davidson,B.L. and McCray,P.B. (2011) Current prospects for RNA interference-based therapies. *Nat. Rev. Genet.*, **12**, 329–340.
7. Cheng,T.L., Teng,C.F., Tsai,W.H., Yeh,C.W., Wu,M.P., Hsu,H.C., Hung,C.F. and Chang,W.T. (2009) Multitarget therapy of malignant cancers by the head-to-tail tandem array multiple shRNAs expression system. *Cancer Gene Ther.* , **16**, 516–531.
8. ter Brake,O., Konstantinova,P., Ceylan,M. and Berkhout,B. (2006) Silencing of HIV-1 with RNA interference: a multiple shRNA approach. *Mol. Ther.*, **14**, 883–892.
9. Tiemann,K., Hohn,B., Ehsani,A., Forman,S.J., Rossi,J.J. and Saetrom,P. (2010) Dual-targeting siRNAs. *RNA*, **16**, 1275–1284.
10. Passioura,T., Gozar,M.M., Goodchild,A., King,A., Arndt,G.M., Poidinger,M., Birkett,D.J. and Rivory,L.P. (2009) Interfering ribonucleic acids that suppress expression of multiple unrelated genes. *BMC Biotechnol.*, **9**, 57.
11. Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
12. Acunzo,M., Visone,R., Romano,G., Veronese,A., Lovat,F., Palmieri,D., Bottoni,A., Garofalo,M., Gasparini,P., Condorelli,G. *et al.* (2012) miR-130a targets MET and induces TRAIL-sensitivity in NSCLC by downregulating miR-221 and 222. *Oncogene*, **31**, 634–642.
13. Acunzo,M., Romano,G., Palmieri,D., Laganà,A., Garofalo,M., Balatti,V., Drusco,A., Chiariello,M., Nana-Sinkam,P. and Croce,C.M. (2013) Cross-talk between MET and EGFR in non-small cell lung cancer involves miR-27a and Sprouty2. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 8573–8578.
14. Thomas,M., Lieberman,J. and Lal,A. (2010) Desperately seeking microRNA targets. *Nat. Struct. Mol. Biol.*, **17**, 1169–1174.
15. Grimson,A., Farh,K.K.-H., Johnston,W.K., Garrett-Engele,P., Lim,L.P. and Bartel,D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, **27**, 91–105.
16. Petryszak,R., Burdett,T., Fiorelli,B., Fonseca,N.A., Gonzalez-Porta,M., Hastings,E., Huber,W., Jupp,S., Keays,M., Kryvych,N. *et al.* (2013) Expression Atlas update–a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments. *Nucleic Acids Res.*, doi:10.1093/nar/gkt1270.
17. Reynolds,A., Leake,D., Boese,Q., Scaringe,S., Marshall,W.S. and Khvorova,A. (2004) Rational siRNA design for RNA interference. *Nat. Biotechnol.*, **22**, 326–330.
18. Dykxhoorn,D.M., Novina,C.D. and Sharp,P.A. (2003) Killing the messenger: short RNAs that silence gene expression. *Nat. Rev. Mol. Cell Biol.*, **4**, 457–467.
19. Betel,D., Koppal,A., Agius,P., Sander,C. and Leslie,C. (2010) Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.*, **11**(8), R90.
20. Shao,Y., Chan,C.Y., Maliyekkel,A., Lawrence,C.E., Roninson,I.B. and Ding,Y. (2007) Effect of target secondary structure on RNAi efficiency. *RNA*, **13**, 1631–1640.

21. Kertesz,M., Iovino,N., Unnerstall,U., Gaul,U. and Segal,E. (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**, 1278–1284.
22. Sun,G., Li,H. and Rossi,J.J. (2010) Sequence context outside the target region influences the effectiveness of miR-223 target sites in the RhoB 3'UTR. *Nucleic Acids Res.*, **38**(1), 239-252.
23. Wang,Y. and Witten,I.H. (1997) Inducing model trees for continuous classes. In: van Someren,M and Widmer,G (eds). *Poster Papers of the 9th European Conference on Machine Learning (ECML 97)*.Prague, Czech Republic, 128–137.
24. Hothorn,T., Hornik,K. and Zeileis,A. (2006) Unbiased recursive partitioning: A conditional inference framework. *J. Comput. Graph. Stat.*, **15**, 651–674.
25. Hsu,S.D., Lin,F.M., Wu,W.Y., Liang,C., Huang,W.C., Chan,W.L., Tsai,W.T., Chen,G.Z., Lee,C.J., Chiu,C.M. *et al.* (2010) miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res.*, **39**, D163–D169.
26. Suda,K., Tomizawa,K., Osada,H., Maehara,Y., Yatabe,Y., Sekido,Y. and Mitsudomi,T. (2011) Conversion from the "oncogene addiction" to "drug addiction" by intensive inhibition of the EGFR and MET in lung cancer with activating EGFR mutation. *Lung Cancer*, doi:10.1016/j.lungcan.2011.11.007.
27. Haas,U., Sczakiel,G. and Laufer,S.D. (2012) MicroRNA-mediated regulation of gene expression is affected by disease-associated SNPs within the 3'-UTR via altered RNA structure. *RNA Biol.*, **9**, 924–937.
28. Saunders,M.A., Liang,H. and Li,W.-H. (2007) Human polymorphism at microRNAs and microRNA target sites. *Proc. Natl. Acad. Sci. U. S. A.*, **104**, 3300–3305.
29. Dykxhoorn,D.M. and Lieberman,J. (2005) The silent revolution: RNA interference as basic biology, research tool, and therapeutic. *Annu. Rev. Med.*, **56**, 401–423.
30. Rao,D.D., Wang,Z., Senzer,N. and Nemunaitis,J. (2013) RNA interference and personalized cancer therapy. *Discov. Med.*, **15**, 101–110.
31. Daka,A. and Peer,D. (2012) RNAi-based nanomedicines for targeted personalized therapy. *Adv. Drug Deliv. Rev.*, **64**, 1508–1521.
32. Kesharwani,P., Gajbhiye,V. and Jain,N.K. (2012) A review of nanocarriers for the delivery of small interfering RNA. *Biomaterials*, **33**, 7138–7150.
33. Breitbach,C.J., Burke,J., Jonker,D., Stephenson,J., Haas,A.R., Chow,L.Q.M., Nieva,J., Hwang,T.-H., Moon,A., Patt,R. *et al.* (2011) Intravenous delivery of a multi-mechanistic cancer-targeted oncolytic poxvirus in humans. *Nature*, **477**, 99–102.
34. Akhtar,S. (2009) Oral delivery of siRNA and antisense oligonucleotides. *J Drug Target.*, **17**, 491–495.
35. Aouadi,M., Tesz,G.J., Nicoloro,S.M., Wang,M., Chouinard,M., Soto,E., Ostroff,G.R. and Czech,M.P. (2009) Orally delivered siRNA targeting macrophage Map4k4 suppresses systemic inflammation. *Nature*, **458**, 1180–1184.
36. Zhang,L., Hou,D., Chen,X., Li,D., Zhu,L., Zhang,Y., Li,J., Bian,Z., Liang,X., Cai,X. *et al.* (2012) Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA. *Cell Res.*, **22**, 107–126.
37. Jackson,A.L., Burchard,J., Schelter,J., Chau,B.N., Cleary,M., Lim,L. and Linsley,P.S. (2006) Widespread siRNA 'off-target' transcript silencing mediated by seed region sequence complementarity. *RNA*, **12**, 1179–1187.
38. Birmingham,A., Anderson,E.M., Reynolds,A., Ilsley-Tyree,D., Leake,D., Fedorov,Y., Baskerville,S., Maksimova,E., Robinson,K., Karpilow,J. *et al.* (2006) 3′ UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat. Methods*, **3**, 199–204.