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

Alessandro Laganà1,#,\*, Mario Acunzo1,#, Giulia Romano1, Alfredo Pulvirenti2, Dario Veneziano1,2, Luciano Cascione3, Rosalba Giugno2, Pierluigi Gasparini1, Dennis Shasha4, Alfredo Ferro2 & Carlo Maria Croce1,\*

1Department of Molecular Virology, Immunology and Medical Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, OH, USA

2Department of Clinical and Molecular Biomedicine, University of Catania, Italy

3IOR - Institute of Oncology Research, Bellinzona, Switzerland

4Department of Computer Science, Courant Institute of Mathematical Sciences, New York University, New York, NY, USA

#These Authors equally contributed to the work.

\*Corresponding Authors

**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 [McCarty].

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.

RNAi is now emerging as a promising therapeutic approach. 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.

Several ongoing and already completed RNAi-based clinical trials suggest encouraging results. 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 [Davidson, Davis].

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 [Cheng] .

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 instead in HIV-1 infected cells that were previously transduced with a double shRNA viral vector [ter Brake].

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 [Tiemann].

An alternative approach for targeting multiple genes is suggested by the endogenous microRNA (miRNA) way of action. miRNAs, indeed, are naturally intended to target multiple genes, often in multiple sites, due to the partial complementarity they exhibit to their targets. 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 to 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 down-regulation of up to 70% was obtained by top ranking miRNAs.

**RESULTS**

**The miR-Synth algorithm and the design features**

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

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 miRNA sequences, (iii) Scoring and ranking of the designed miRNAs.

The first step mainly relies on the concept of miRNA seed, which is the 5′ region of the miRNA, centered on nucleotides 2–7 (Fig. 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 of regulated targets [Bartel, Grimson]. 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 [Ref]. However, these cases are much less abundant than canonical sites which represent the predominant interaction model associated with greater target repression [Bartel, Grimson]. Among canonical sites, 7mer-m8 and 8-mer 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.

We estimated that, on average, 64.3% of human 3' UTR sequence pairs share at least a common 7nt pattern and this percentage decreases to 16%, 2.6% and 0.08% in the cases of sets of 3, 4 and 5 3'UTR sequences, respectively (see supplementary information). In light of this, and considering the fact that groups with more than 5 well conserved sequences might share at most one common 7nt pattern or one or more 6nt sites, we decided to set a maximum threshold of 8 target sequences that users can provide as input to the system. We believe this is not a real limitation, since 8 is already a considerable number of targets, unlikely to be practical in most applications. These sequences are screened for repeated patterns of 6 or 7 nucleotides (depending on user choice), which will constitute the binding sites for miRNA seeds (Fig. 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.

The second step of the algorithm consists of the actual miRNA sequence design. For each repeated pattern identified in the previous phase an anti-complementary miRNA 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 Fig. 1c. A 3nt mismatch after the seed is forced. The miRNA sequences thus obtained will be 22 nt long.

The designed miRNAs 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 (REF). These particular thresholds were chosen according to what has been observed in typical endogenous miRNA nucleotide composition (see supplementary information).

In this phase users can also choose to discard miRNA sequences sharing a seed with any endogenous miRNA. More details about the algorithm are given as supplementary information.

**Scoring and ranking of the synthetic miRNAs**

The third step of the miR-Synth pipeline consists in the evaluation and ranking of the designed miRNAs. We developed a scoring function based on six different features of validated 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, presence of ARE and CPE motifs upstream of the binding sites (REF). For any given miRNA, 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 conditional inference trees (ctree).

We have trained the system on a set of publicly available gene expression profiles following the over-expression of nine individual endogenous miRNAs. In particular, binding sites were predicted for each transfected miRNA on down-regulated 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 down-regulation 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 (Fig. 1d), 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 (Fig. 1e).

These two methods are used to evaluate the designed miRNAs. In particular, miRNAs are first ranked according to the M5P score and subsequently by ctree score.

We validated this scoring function by using a database of experimentally validated human miRNA/target interactions called miRTarBase as a test set. This dataset contains 495 cases of proven direct interactions, 490 cases for which direct binding wasn’t 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.

**Validation of single-target multi-site miRNAs**

Our miRNA 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 miRNAs, 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 [Suda, 2010].

We designed two different sets of multi-site miRNAs exclusively targeting c-MET and EGFR respectively. The system returned 102 miRNAs for c-MET and 44 miRNAs for EGFR. For each of the two genes, we focused on the top 4 miRNAs as ranked by our scoring system. Table 1a/b summarizes the main features of these miRNAs. The 8 miRNAs thus taken into consideration had at least 2 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. miRNAs for c-MET and EGFR were individually co-transfected with the c-MET and EGFR 3’-UTR constructs respectively, in HEK293 cells. This resulted in a significant inhibition of the luciferase activity induced by two c-MET miRNAs and three EGFR miRNAs, as compared to the negative control (Fig. 2 and 3). Moreover, western-blot and qRT-PCR assays showed that over-expression of miRNAs in HeLa cells strongly reduced the endogenous protein and mRNA levels of c-MET and EGFR as compared to control (Fig. 2 and 3), in agreement with the luciferase assay results. Expression of transfected miRNAs in HeLa transfected cells was confirmed by qRT-PCR (Fig. 2 and 3). Among the five functional miRNAs, amiR-M-60 and amiR-E-3 yielded stronger down-regulation of c-MET and EGFR 3’ UTRs luciferase activity, respectively (Fig. 2 and 3). Hence, as further analysis, we performed mutagenesis of amiR-M-60 and amiR-E-3 binding sites within the MET and EGFR 3’-UTRs, which abolished the ability of these miRNAs to regulate luciferase expression, thus confirming that the binding sites are functional (Fig. 2 and 3).

**Validation of multi-target synthetic miRNAs**

We subsequently designed synthetic miRNAs intended to target both c-MET and EGFR concurrently. The algorithm returned a total of 103 miRNAs (Table 1c). We selected the top 6 miRNAs as ranked by our scoring function. All of them had one 8mer binding site on each gene.

To verify multiple direct targeting of c-MET and EGFR, the designed miRNAs were individually co-transfected with both wild-type c-MET and EGFR 3' UTR constructs into HEK293 cells. amiR-ME-196 and amiR-ME-141 induced a significant inhibition of the luciferase activity for both constructs, while amiR-ME-140 and amiR-ME-106 yielded a significant repression of c-MET only, as compared to the negative control (Fig. 4). Moreover, over-expression of the artificial miRNAs in HeLa cells induced a strong repression of the endogenous c-MET and EGFR proteins and mRNAs in three cases and a mild down-regulation in the three remaining cases, as compared to the control (Fig. 4). Interestingly, although not all tested miRNAs 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 miRNAs, amiR-ME-196 was chosen for further investigation because of its greater down-regulation at both the protein and the luciferase level (Fig. 4). The expression of amiR-ME-196 in HeLa transfected cells was confirmed by qRT-PCR (Fig. 4). Mutagenesis of the miRNA binding site within the MET and EGFR 3’-UTRs eliminated its ability to regulate luciferase expression, thus confirming that the binding site is functional (Fig. 4).

**The miR-Synth web interface**

miR-Synth is freely available for academic use through a web interface. Users can provide up to 8 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 miRNAs simultaneously targeting all of the provided sequences or to include miRNAs targeting subsets of them as well. A list of sequences (or their IDs) that must not be directly targeted by the designed miRNAs 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 or other mutations in the targets could negatively affect miRNA binding.

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 miRNAs for a pair of targets with default parameters takes 30 seconds 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 miRNA, 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 as supplementary information.

**DISCUSSION**

RNAi constitutes a powerful tool for the regulation of gene expression. Recent progress in the development of increasingly efficient carriers for the intracellular delivery of small RNAs, such as nano-particles and viral systems, has made the establishment of therapeutics based on this promising technology imaginable. Moreover, 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.

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. 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 miRNA 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. Although this may in principle represent a substantial advantage in the employment of artificial miRNAs in place of siRNAs/shRNAs, it certainly deserves additional consideration supported by further experimental validation.

The miR-Synth pipeline allows the rational design of artificial miRNAs 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 down-regulation, and improve the selection of effective molecules.

**References**

...

**Figures and Tables**

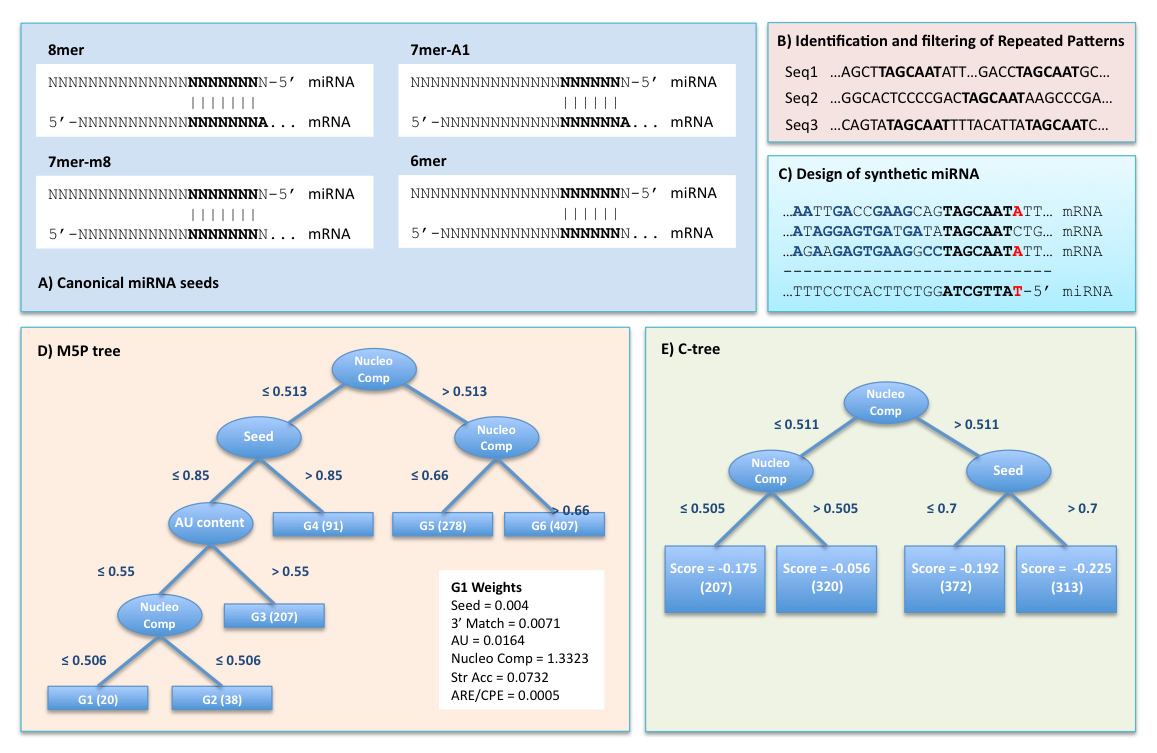


Fig. 1

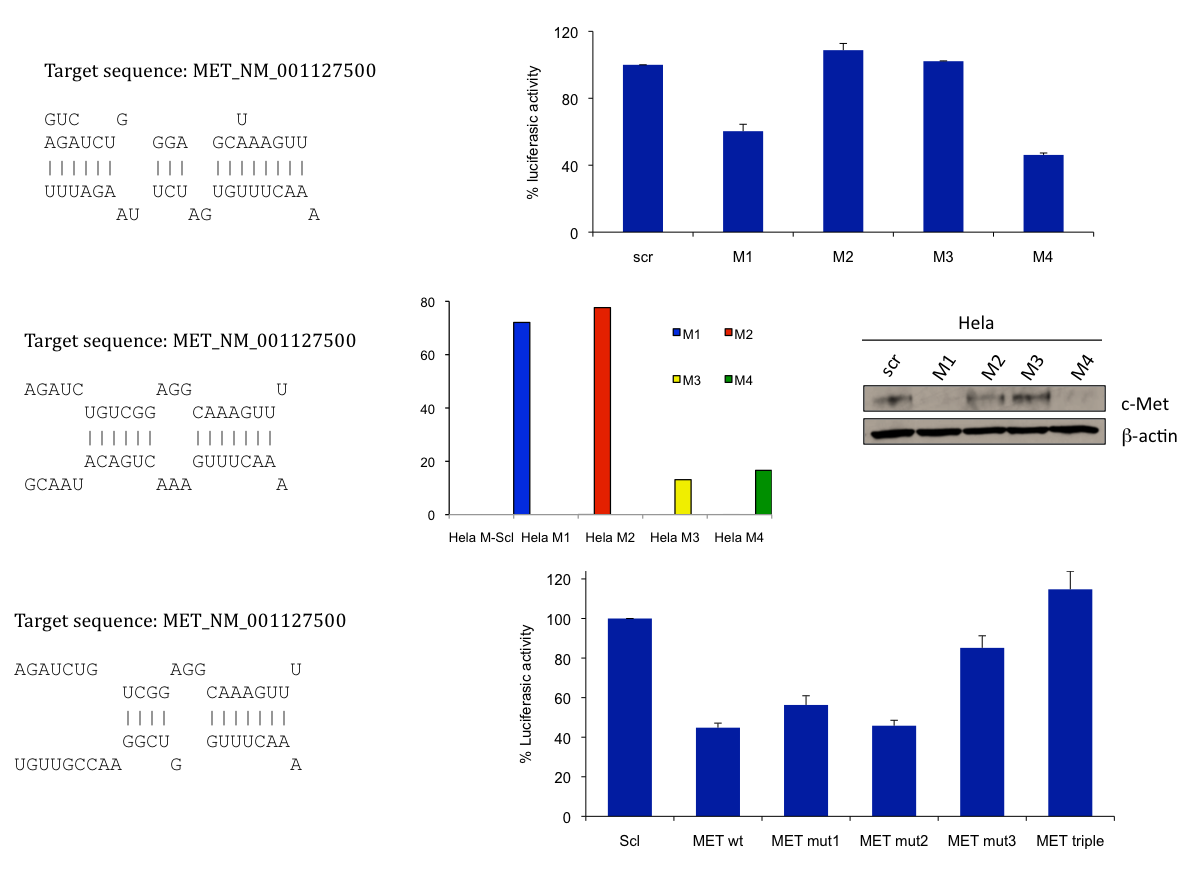


Fig. 2

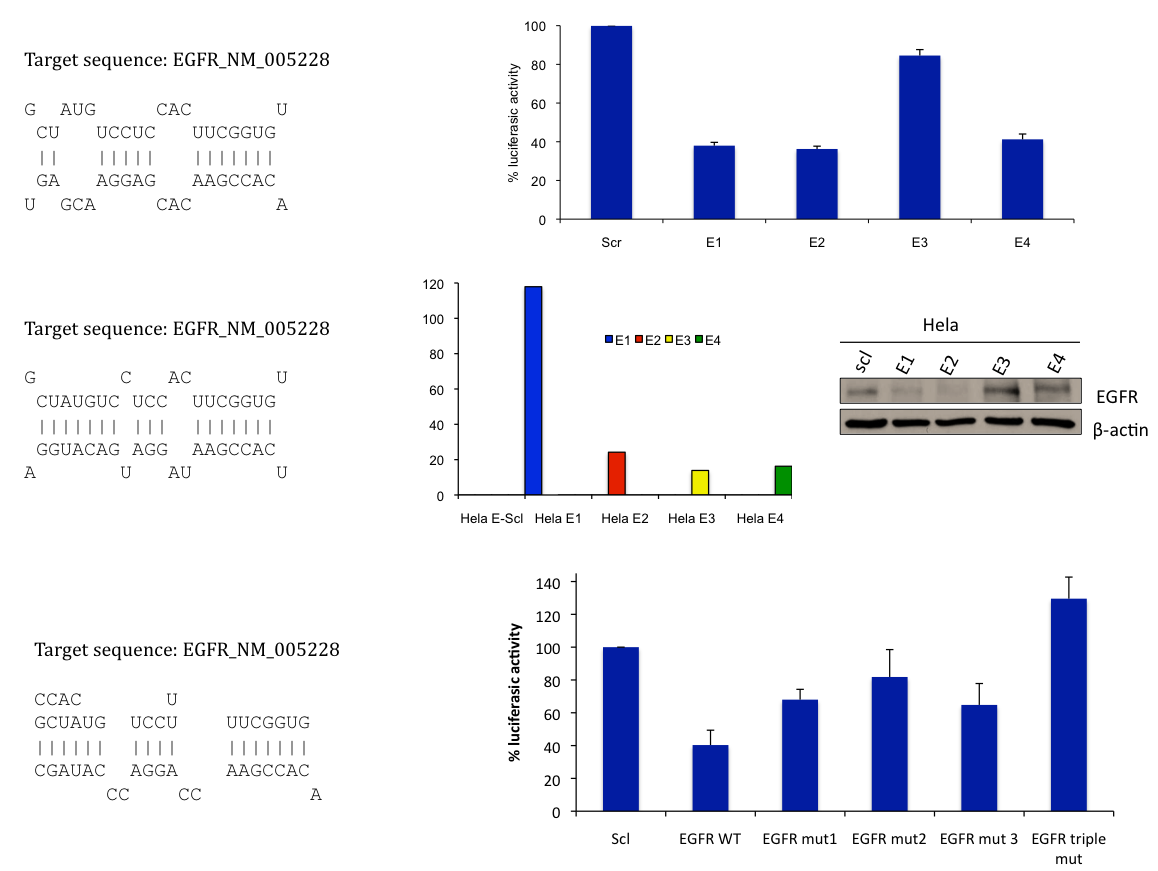


Fig. 3

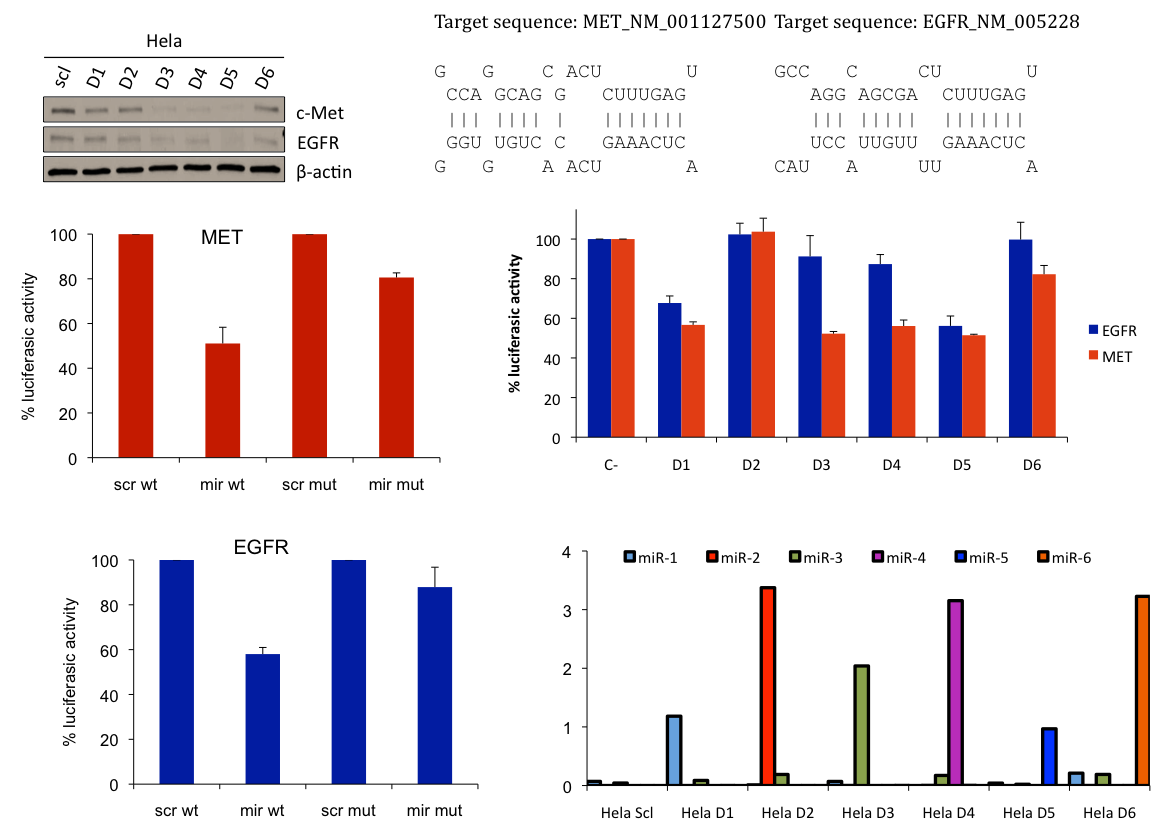


Fig. 4

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 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 |

Table 1