**SUPPLEMENTARY INFORMATION**

**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

**S1. The miR-Synth pipeline**

miR-Synth is a computational tool for the design of multi-site multi-target artificial miRNAs (a-miRs) available through a web interface. It accepts up to 8 target sequences and returns a list of synthetic miRNAs potentially able to target simultaneously the input sequences or subsets of them, together with details about the predicted interactions.

The miR-Synth pipeline consists of the following three steps: (i) Identification and filtering of repeated patterns, (ii) Design and filtering of a-miR sequences, (iii) Scoring and ranking of the designed a-miRs. A detailed description of these steps is given below.

*(i) Identification and filtering of repeated patterns*

Input sequences are screened for repeated patterns of 6 or 7 nucleotides, which will constitute the seed binding regions: 6 nt for 6mer and 7mer-A1 matches, 7 nt for 7mer-m8 and 8mer matches.

These binding sites are then filtered based on the following user-provided specifications:

* A site must appear at least *n* times on each target (Default: n=2).
* A site must be present in every target (Default: NO).
* Users can also provide a list of sequences that must not be targeted. In this case, all the seed sites that appear at least once in any of the provided sequences will be removed.

 *(ii) Design and filtering of a-miR sequences*

For each seed site identified in the previous phase, all its instances are aligned together and an anti-complementary a-miR seed is created. The rest of the sequence is constructed by computing a consensus sequence through a simple profile technique, as depicted in Fig. S1. The a-miR sequences thus obtained will be 22 nt long. The designed a-miRs are then filtered based on their nucleotide composition. In particular:

* Sequences with GC content out of the user's specified range will be discarded (Default: 23% - 78%) (See section S2).
* Sequences containing stretches of six or more nucleotides of the same kind are discarded (See section S2).
* Sequences sharing a seed with an endogenous miRNA are discarded.

*(iii) Scoring and ranking of the designed a-miRs*

We developed a scoring function based on six different features of endogenous miRNA/target interactions coming from the literature. For any given a-miR, each feature is assigned a score ranging from 0 to 1. The features and the computation of the related scores is here described.

*Seed type*. We assigned a score to each of the 4 types of canonical seeds, based on the degree of repression typically yielded: 8mer = 1.0, 7mer-m8 = 0.7, 7mer-A1 = 0.5, 6mer = 0.4. Then, for each designed a-miR, the average seed score on its binding sites is computed.

*Degree of pairing of the a-miR 3' region*. This is simply computed as the fraction of paired bases in the a-miR/target duplex, excluding the seed area. Then, for each designed a-miR, the average 3’ match score on its binding sites is computed.

*AU content of the binding site and its surrounding regions*. This is computed as the fraction of A and U bases in the region which spans from 5 nt upstream to 5 nt downstream of the whole binding site. Then, for each designed a-miR, the average AU score on its binding sites is computed.

*a-miR nucleotide composition*. We assigned a score to the base pairs on each position of the a-miR/target duplex, based on the frequency of base pairs observed on a set of validated endogenous miRNA/target duplexes retrieved from the database Tarbase. For each a-miR binding site, this score is computed as the average value over the seed region.

*Structural accessibility of the binding sites*. For each binding site, local structural accessibility is calculated on a 160 nt window centered on the site, by using the program RNAplfold from the Vienna RNA Package (39) with the following parameters: *W*=80, L=40 (40, 41) and *u*=4. *W* is the sliding window size, *L* is the maximum base pairing distance, and *u* is the stretch of consecutive nts within the target site for which the accessibility is computed. We consider the probability of 4 nt sub-sequences to be unpaired and compute the average value over the bases of the binding site (42). Then, for each designed a-miR, the average accessibility score on its binding sites is computed.

*Presence of ARE and CPE motifs upstream of the binding sites*. For each binding site, the fraction of ARE (sequence: AUUUA) and CPE (sequence: UUUUUAU) motifs in the upstream region is computed. For each designed a-miR, the ARE/CPE score is computed as the average value over its binding sites.

*Repression Score*. For each designed a-miR, the total score is calculated by using the tree-based learning systems M5P and CTree, based on the scores of the 6 features described above. Both systems were trained on a set of publicly available gene expression profiles following the over-expression of nine individual endogenous human miRNAs (15) (miR-9-5p, miR-7-5p, miR-181a-5p, miR-148b-3p, miR-142-5p, miR-133a, miR-132-3p, miR-128 and miR-122-5p). In particular, binding sites on down-regulated genes were predicted for each transfected miRNA, then feature scores were calculated. The gene expression fold change after 24h 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, 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. 1d and tables st1 contains details about the generated M5P tree and the corresponding sets of feature weights. Fig.1e describes the generated CTree and the score classes.

*Ranking of the designed a-miRs*. The a-miRs produced by miR-Synth are first ranked according to the CTree score and subsequently by the M5P score. In particular, CTree splits the a-miRs into major classes, while M5P is used to rank a-miRs within each class.

**In-silico validation of the miR-Synth scoring function**

We validated the 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 of down-regulated genes for which direct binding wasn’t verified and 71 negative cases.

We randomly created 1,000 groups with 71 proven direct and 71 proven negative cases. For each group, we found that the top 10 interactions, as ranked by our approach, always contained a higher number of true direct interactions compared to 10,000 sets of 10 cases randomly chosen (6.5 out of 10 vs 5, P<0.0001). We obtained the same result when we considered random groups with 71 proven direct, 71 indirect and 71 proven negative cases and counted the number of proven direct interactions in the top 10 cases (5.8 out of 10 vs 3.3, P< 0.0001).

**S2. Basic miRNA and 3’ UTR descriptive statistics**

We performed some basic statistics on endogenous human mature miRNA sequences in order to identify typical values to be used by filters in the artificial miRNA selection process.

*Stretches of nucleotides of the same kind*

General siRNA guidelines recommend to avoid sequences with long stretches of the same kind of nucleotide. Endogenous human miRNAs commonly present stretches of 3 or 4 nucleotides of the same kind (72.2% and 26.9%, respectively), while stretches of 5, 6 and 7 nucleotides are more rare (7.5%, 1% and 0.4%, respectively) (Ref. miRBase Rel. 19 – 2042 mature sequences)(43) (See chart below). We set 5 nt as the threshold for stretches of the same kind allowed for an artificial miRNA.



*GC content*

General siRNA guidelines recommend to avoid sequences with GC content <30% and >60%. GC content in endogenous human miRNAs vary from 8.69% to 100% (Ref. miRBase Rel. 19 – 2042 mature sequences) (43). However, only 4% of them have GC content less than 23% and greater than 78%. We set 23%-78% as default range for artificial miRNAs’ GC content.



*Repeated patterns in human 3’ UTR sequences*

In order to appropriately set the maximum number of sequences that can be accepted as input to miR-Synth, we performed a simple test on sets of plausible target genes.

We collected gene expression data associated to different disease conditions from the Gene Expression Atlas (16) (http://www.ebi.ac.uk/gxa) and focused on the up-regulated genes, thus mimicking a typical plausible scenario for the use of artificial miRNAs. For each condition, we calculated all the possible combinations of two and three up-regulated genes and counted how many of them share 7mer 3' UTR sites. We filtered out polyA-signal motifs, homopolymer motifs and sites matching the seeds of endogenous miRNAs. We were able to perform this analysis for all pairs and triplets of up-regulated genes related to 83 different diseases and results show that 97.3% of pairs and 81.32% of triplets share at least one 7mer site. On average, pairs and triplets shared about 136 and 24 7mer sites, respectively. For 96 additional diseases we were able to perform the pairs analysis and cumulative results confirmed the results above, with 97.13% pairs sharing on average 132 7mer sites. We couldn't perform the triplets analysis on these additional diseases due to the high number of up-regulated genes which generated a number of triplets too computationally expensive to analyze. Table S10 shows detailed information on the tests performed.

Based on these tests, considering highly homologous sequences and the fact that users might also be interested in 6mer seeds, 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, more than would be practical in most applications.

**S3. *In silico* Analysis of Off-Target Effects**

In order to investigate the potential differences between a-miRs and siRNAs in terms of off-target effects, we performed an *in-silico* analysis on double-target a-miRs. Evidence shows that siRNAs may act like endogenous miRNAs and repress unintended targets by perfect complementarity of their seed region. Since siRNAs are designed to target a single gene, the simultaneous repression of c-MET and EGFR would take at least two different siRNAs. Thus, we designed siRNAs for c-MET and EGFR by using three different tools based on different algorithm and features: optiRNAi, RNAxs and siDirect2 (40, 44, 45). For each tool and each target, we chose the top 6 siRNAs that didn't share the seed sequence with any endogenous miRNA. We considered all the possible pairs of c-MET/EGFR siRNAs from each tool separately and, for each pair, calculated the number of potential off-target genes by matching their seed sequences with the whole 3' UTRome (34619 sequences).

Although seed match is not always a reliable indicator of an actual interaction, most functional interactions occur through perfect seed pairing. We considered both 6mer and 7mer matches and compared the number of off-targets of the 108 siRNA combinations with the off-targets of the top 6 a-miRs designed by miR-Synth. On average, a-miRs had about 8911 off-target 6mer and 3108 7mer matches as opposite to 13190 off-target 6mer and 6067 off-target 7mer matches for siRNAs. Figure S3 shows the distribution of the off-targets, and clearly demonstrates that double-targeting a-miRs are likely to have less off-target effects than pairs of single-targeting siRNAs.

**S4. Additional *In silico* Analysis of Potential Seed Binding Sites**

In order to assess the general applicability of our method, we refined the 7mer analysis described in section S2 by additionally running miR-Synth with default parameters\* on all pairs of up-regulated genes related to 8 diseases retrieved from the GEA dataset. These gene sets were chosen based on the known relevance to the associated disease and also on their [whose size?] size, in order to focus attention on the most relevant pairs. We thus analyzed a total of 13803 gene pairs.

miR-Synth was able to design at least an amiR for 95% of pairs and at least 6 amiRs for 83.5% of pairs. We computed the average feature and repression scores of the top 6 amiRs from each set (total 78649 amiRs) and compared them with the average scores of the top 6 c-MET/EGFR amiRs that we previously validated. These scores are reported on Table S11 and Figure S4 depicts their comparison charts. The scores obtained for the GEA dataset were comparable with the scores of the c-MET/EGFR amiRs. 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 in section S2.

We also computed the potential number of off-target genes for the whole set of 78649 amiRs. A gene was considered to be a potential off-target for an amiR if it contains at least a 7mer binding site for the amiR seed. The results showed an average of 3777 off-target genes per amiR, in agreement with the off-target analysis described in section S3 (See Table S11).

\* Default parameters. Endogenous miRNA seed filter: enabled. GC content: 23%-78%. Minimum number of binding sites on each sequence: 1. amiRs must target all input sequences: enabled. Seed types allowed: 7mer-m8/8mer.

**SUPPLEMENTARY REFERENCES**

39. Lorenz,R., Bernhart,S.H., Höner Zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. Algorithms for molecular biology. *Algorithms Mol Biol.*, **6**, 26.

40. Tafer,H., Ameres,S.L., Obernosterer,G., Gebeshuber,C.A., Schroeder,R., Martinez,J. and Hofacker,I.L. (2008) The impact of target site accessibility on the design of effective siRNAs. *Nat Biotechnol*, **26**, 578–583.

41. Marín,R.M. and Vanícek,J. (2010) Efficient use of accessibility in microRNA target prediction. *Nucleic Acids Res*, 10.1093/nar/gkq768.

42. Long,D., Lee,R., Williams,P., Chan,C.Y., Ambros,V. and Ding,Y. (2007) Potent effect of target structure on microRNA function. *Nat Struct Mol Biol*, **14**, 287–294.

43. Kozomara,A. and Griffiths-Jones,S. (2010) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res*, **39**, D152–D157.

44. Cui,W., Ning,J., Naik,U.P. and Duncan,M.K. (2004) OptiRNAi, an RNAi design tool. *Computer Methods and Programs in Biomedicine*, **75**, 67–73.

45. Naito,Y., Yoshimura,J., Morishita,S. and Ui-Tei,K. (2009) siDirect 2.0: updated software for designing functional siRNA with reduced seed-dependent off-target effect. *BMC Bioinformatics*, **10**, 392.

**SUPPLEMENTARY FIGURES**

**Fig. S1. Design of synthetic miRNA.** (a) An example of miRNA design process through consensus computation for three binding sites. The seed region and its binding sites are in red, the matches in the 3’ region are in blue. (b) Predicted interactions of the designed miRNA with each binding site.

 *seed site*

…5’-**AA**TT**GA**CC**GAAG**CAG**TAGCAATA**TT-3’… mRNA site1

…5’-**A**T**AGGAGTGA**T**GA**TA**TAGCAAT**CTG-3’… mRNA site2

…5’-**A**G**A**A**GAGTGAAG**G**CCTAGCAATA**TT-3’… mRNA site3

*…5’-****A****\*****A****\*****GAGTGAAG****\*\*\*****TAGCAATA****..-3’… Consensus.*

 ---------------------------------- \_

…3’-TTTCCTCACTTCTGG**ATCGTTA**T-5’ miRNA\_\_\_\_\_

(a)

…5’-AATTGACCGAAGCAG**TAGCAAT**ATT-3’… mRNA site1

|| || |||| **|||||||**| \_

…3’-TTTCCTCACTTCTGG**ATCGTTA**T-5’ miRNA\_\_\_\_\_

…5’-ATAGGAGTGATGATA**TAGCAAT**CTG-3’… mRNA site2

| |||||||| || **|||||||** \_

…3’-TTTCCTCACTTCTGG**ATCGTTA**T-5’ miRNA\_\_\_\_\_

…5’-AGAAGAGTGAAGGCC**TAGCAAT**ATT-3’… mRNA site3

| | |||||||| ||**|||||||**| \_

…3’-TTTCCTCACTTCTGG**ATCGTTA**T-5’ miRNA\_\_\_\_\_

(b)

**Fig. S2. Experimental test of low scoring a-miRs.** The bottom 6 a-miRs designed for c-MET and EGFR were tested by luciferase assay (See Table S8, a-miRS 120-125). 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: ± s.e.m., P<0.05). A mild repression of EGFR is yielded by a-miR-ME-88, a-miR-ME-169 and a-miR-188, while no significant repression of c-MET is observed with any of the tested a-miRs.

**Fig. S3. In silico analysis of off-target effects.** Distribution of off-target numbers for double-target a-miRs and pairs of single-target siRNAs based on (a) 6mers and (b) 7mers.



(a)

****

(b)

**Fig. S4. Additional *In silico* Analysis of Potential Seed Binding Sites.** Average feature and repression scores of the top 6 amiRs from each GEA set and the top 6 c-MET/EGFR amiRs.

****

**SUPPLEMENTARY TABLES**

**Table S1.** Basic information about the tested miRNAs.

|  |
| --- |
| (a) 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 |
| (b) 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 |
| (c) 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 S2**. Sets of feature weights assigned by M5P tree (see fig. 1d).

|  |  |  |
| --- | --- | --- |
| Group (leaf) | Feature | Weight |
| G1 | Seed Match | + 0.004 |
| 3’ Match | + 0.0071 |
| AU Content | + 0.0164 |
| Nucleotide Composition | + 1.3323 |
| Structural Accessibility | + 0.0732 |
| ARE/CPE motifs | + 0.0005 |
| Constant | - 0.8396 |
| G2 | Seed Match | + 0.004 |
| 3’ Match | + 0.0071 |
| AU Content | + 0.0164 |
| Nucleotide Composition | + 1.0936 |
| Structural Accessibility | + 0.0525 |
| ARE/CPE motifs | + 0.0005 |
| Constant | - 0.6601 |
| G3 | Seed Match | + 0.004 |
| 3’ Match | + 0.0032 |
| AU Content | + 0.0078 |
| Nucleotide Composition | + 2.5724 |
| Structural Accessibility | + 0.0066 |
| ARE/CPE motifs | + 0.0005 |
| Constant | - 1.3783 |
| G4 | Seed Match | + 0.0075 |
| 3’ Match | + 0.0034 |
| AU Content | - 0.0465 |
| Nucleotide Composition | + 0.3517 |
| Structural Accessibility | + 0.0074 |
| ARE/CPE motifs | - 0.0162 |
| Constant | - 0.2045 |
| G5 | Seed Match | - 0.0044 |
| 3’ Match | 0 |
| AU Content | + 0.0007 |
| Nucleotide Composition | + 0.3131 |
| Structural Accessibility | + 0.0009 |
| ARE/CPE motifs | + 0.0002 |
| Constant | - 0.3749 |
| G6 | Seed Match | - 0.2156 |
| 3’ Match | 0 |
| AU Content | + 0.0007 |
| Nucleotide Composition | + 0.2042 |
| Structural Accessibility | + 0.0009 |
| ARE/CPE motifs | + 0.0002 |
| Constant | - 0.1842 |

**Table S3**. Tested synthetic miRNAs for c-MET.

**Input parameters**

Target sequence: MET – Accession number: NM\_001127500 (Entrez Gene ID: 4233

Source organism: Human

Seed type: 7mer-m8 / 8mer

Minimum number of binding sites on each sequence: 2

Minimum GC content allowed: 23%

Maximum GC content allowed: 78%

Homologous miRNA filter enabled: YES

Repeats filter enabled: YES

**miRNA list**

**1) miRNA ID: 60**

Sequence: UUUGAAACGGAGGCUGUCUAGA

Total binding sites: 3

*Scores*

Seed score: 1.0

3' Match score: 0.4761904761904762

AU Content score: 0.5333333333333333

Nucleotide Composition score: 0.6739926739926739

Structural Accessibility score: 0.2727995679

ARE/CPE score: 0.2916666666666667

-------------------------------------------------------------------------

C-tree score: -0.22488171570287555

M5P score: -0.26149350969291935

*Binding sites details*

Position: 413

Type: 8mer

Alignment:

 GUC G U

3’-AGAUCU GGA GCAAAGUU -5’ miRNA

 |||||| ||| ||||||||

5’-UUUAGA UCU UGUUUCAA -3’ mRNA

 AU AG A

Free energy: -13.1 Kcal/mol

Position: 700

Type: 8mer

Alignment:

 AGAUCUG AGG U

3’- UCGG CAAAGUU -5’ miRNA

 |||| |||||||

5’- GGCU GUUUCAA -3’ mRNA

 UGUUGCCAA G A

Free energy: -11.6 Kcal/mol

Position: 848

Type: 8mer

Alignment:

 AGAUC AGG U

3’- UGUCGG CAAAGUU -5’ miRNA

 |||||| |||||||

5’- ACAGUC GUUUCAA -3’ mRNA

 GCAAU AAA A

Free energy: -15.1 Kcal/mol

===================================

**2) miRNA ID: 118**

Sequence: UUUAUAAAGUCGAUACGUGUUU

Total binding sites: 3

*Scores*

Seed score: 1.0

3' Match score: 0.5

AU Content score: 0.7333333333333334

Nucleotide Composition score: 0.6776556776556776

Structural Accessibility score: 0.3292161828333333

ARE/CPE score: 0.6666666666666666

-------------------------------------------------------------------------

C-tree score: -0.22488171570287555

M5P score: -0.260479749391494

*Binding sites details*

Position: 876

Type: 8mer

Alignment:

 UAGCUGA U

3’- UUUGUGCA AAUAUU -5’ miRNA

 |||||||| ||||||

5’- AAACAUGU UUAUAA -3’ mRNA

 UCACACA A

Free energy: -7.21 Kcal/mol

Position: 1492

Type: 8mer

Alignment:

 UUUGUG CU U

3’- CAUAG GAAAUAUU -5’ miRNA

 ||||| ||||||||

5’- GUGUU UUUUAUAA -3’ mRNA

 AAA AUAUU A

Free energy: -6.8 Kcal/mol

Position: 2077

Type: 8mer

Alignment:

 UUU U UG U

3’- GUGCA AGC AAAUAUU -5’ miRNA

 ||||| ||| |||||||

5’- CACGU UUG UUUAUAA -3’ mRNA

 AAU CUA A

Free energy: -10.3 Kcal/mol

===================================

**3) miRNA ID:181**

Sequence: UUCUUUCUAAGGACGGGGCCGU

Total binding sites: 2

*Scores*

Seed score: 1.0

3' Match score: 0.5357142857142857

AU Content score: 0.6

Nucleotide Composition score: 0.7124542124542124

Structural Accessibility score: 0.32562529544999996

ARE/CPE score: 0.875

-------------------------------------------------------------------------

C-tree score: -0.22488171570287555

M5P score: -0.25342878705094485

*Binding sites details*

Position: 1941

Type: 8mer

Alignment:

 UGC GGAA U

3’- CGGGGCA UCUUUCU -5’ miRNA

 ||||||| |||||||

5’- GCUCUGU AGAAAGA -3’ mRNA

 UCU GGAA A

Free energy: -17.0 Kcal/mol

Position: 1986

Type: 8mer

Alignment:

 CG G C GA U

3’-UGC G G AG AUCUUUCU -5’ miRNA

 ||| | | || ||||||||

5’-AUG C C UC UAGAAAGA -3’ mRNA

 AA G A AA A

Free energy: -13.6 Kcal/mol

===================================

**4) miRNA ID:176**

Sequence: UCAGUACAAAACCUUGUGGCUU

Total binding sites: 2

*Scores*

Seed score: 1.0

3' Match score: 0.6785714285714286

AU Content score: 0.7

Nucleotide Composition score: 0.7472527472527473

Structural Accessibility score: 0.2774491527499999

ARE/CPE score: 0.875

-------------------------------------------------------------------------

C-tree score: -0.22488171570287555

M5P score: -0.24629628477351403

*Binding sites details*

Position: 1637

Type: 8mer

Alignment:

 UU G AAA

3’- CG UGUUCC ACAUGACU-5’ miRNA

 || |||||| ||||||||

5’- GC AUAAGG UGUACUGA-3’ mRNA

 UU G AAA

Free energy: -21.4 Kcal/mol

Position: 1827

Type: 8mer

Alignment:

 G AA

3’-UUC GUGUUCC AACAUGACU-5’ miRNA

 ||| ||||||| |||||||||

5’-AAG CACAAGG UUGUACUGA-3’ mRNA

 A AA

Free energy: -23.4 Kcal/mol

**Table S4**. Tested synthetic miRNAs for EGFR.

**Input Parameters**

Target sequence: EGFR – Accession number: NM\_005228 (Entrez Gene ID: 1956)

Source organism: Human

Seed type: 7mer-m8 / 8mer

Minimum number of binding sites on each sequence: 2

Minimum GC content allowed: 23%

Maximum GC content allowed: 78%

Homologous miRNA filter enabled: YES

Repeats filter enabled: YES

**miRNA list**

**1) miRNA ID:3**

Sequence: UGUGGCUUCACCUCCUGUAUCG

Total binding sites: 3

*Scores*

Seed score: 0.9 3'

Match score: 0.6428571428571429

AU Content score: 0.5

 Nucleotide Composition score: 0.6684981684981685

Structural Accessibility score: 0.1919591185491667

ARE/CPE score: 0.5

-------------------------------------------------------------------------

C-tree score: -0.22488171570287555

M5P score: -0.2411099107859798

*Binding sites details*

Position: 41

Type: 8mer

Alignment:

 CCAC U

3'-GCUAUG UCCU UUCGGUG -5' miRNA

 |||||| |||| |||||||

5'-CGAUAC AGGA AAGCCAC -3' mRNA

 CC CC A

Free energy: -21.1 Kcal/mol

Position: 434

Type: 8mer

Alignment:

 G AUG CAC U

3'- CU UCCUC UUCGGUG -5' miRNA

 || ||||| |||||||

5'- GA AGGAG AAGCCAC -3' mRNA

 U GCA CAC A

Free energy: -21.7 Kcal/mol

Position: 579

Type: 7mer-m8

Alignment:

 G C AC U
3'- CUAUGUC UCC UUCGGUG -5' miRNA

 ||||||| ||| |||||||

5'- GGUACAG AGG AAGCCAC -3' mRNA

 A U AU U

Free energy: -25.5 Kcal/mol

===================================

**2) miRNA ID:106**

Sequence: UGUGUGACACUGCGUAAGGGGG

Total binding sites: 2

*Scores*

Seed score: 1.0 3'

Match score: 0.6785714285714286

AU Content score: 0.7

Nucleotide Composition score: 0.7875457875457875

Structural Accessibility score: 0.4307963975

ARE/CPE score: 0.875

-------------------------------------------------------------------------

C-tree score: -0.22488171570287555

M5P score: -0.23793043342540024

*Binding sites details*

Position: 1158

Type: 8mer

Alignment:

 UC U

3'-GGGGGAAUGCG ACAGUGUG -5' miRNA

 ||||||||||| ||||||||

5'-CCUCCUUACGC UGUCACAC -3' mRNA

 UU A

Free energy: -34.5 Kcal/mol

Position: 1568

Type: 8mer

Alignment:

 GG GG UCA U

3'- G AAUGCG CAGUGUG -5' miRNA

 | |||||| |||||||

5'- C UUAUGU GUCACAC -3' mRNA

 UA AG UCA A

Free energy: -17.8 Kcal/mol

===================================

**3) miRNA ID:25**

Sequence: CAAAUGCUCGAGAGUCCGAUGU

Total binding sites: 2

*Scores*

Seed score: 0.85

3' Match score: 0.6785714285714286

AU Content score: 0.5

Nucleotide Composition score: 0.6739926739926739

Structural Accessibility score: 0.15839958197499998

ARE/CPE score: 0.375

-------------------------------------------------------------------------

C-tree score: -0.22488171570287555

M5P score: -0.2292631363469185

*Binding sites details*

Position: 193

Type: 8mer

Alignment:

 UGU CC G GC C

3'- AG UGA A UCGUAAA -5' miRNA

 || ||| | |||||||

5'- UC ACU U AGCAUUU -3' mRNA

 UCU AA G GA A

Free energy: -10.7 Kcal/mol

Position: 845

Type: 7mer-m8

Alignment:

 C

3'-UGUAGCCUGAG AG UCGUAAAC-5' miRNA

 ||||||||||| || ||||||||

5'-ACAUUGGAUUC UC AGCAUUUG-3' mRNA

 A

Free energy: -25.2 Kcal/mol

===================================

**4) miRNA ID:83**

Sequence: UAACAAUGCACUGGGGGCCCUG

Total binding sites: 2

*Scores*

Seed score: 0.85

3' Match score: 0.6071428571428571

AU Content score: 0.6499999999999999

Nucleotide Composition score: 0.6785714285714286

Structural Accessibility score: 0.35153470350000005

ARE/CPE score: 0.875

-------------------------------------------------------------------------

C-tree score: -0.22488171570287555

M5P score: -0.22794933305256426

*Binding sites details*

Position: 737

Type: 7mer-m8

Alignment:

 GUCCC CAC U

3'- GGGGGU GUAACAA -5' miRNA

 |||||| |||||||

5'- CUUCCA CAUUGUU -3' mRNA

 UUUGU UUC U

Free energy: -12.3 Kcal/mol

Position: 1644

Type: 8mer

Alignment:

 C CA

3'-GUC CGGGGGU CGUAACAAU-5' miRNA

 ||| ||||||| |||||||||

5'-CAG GCCCCUA GCAUUGUUA-3' mRNA

 A CA

Free energy: -27.4 Kcal/mol

**Table S5**. Tested Synthetic miRNAs for c-MET and EGFR.

**Input Parameters**

Target sequences:

MET – (Accession number: NM\_001127500 (Entrez Gene ID: 4233)

EGFR – Accession number: NM\_005228 (Entrez Gene ID: 1956)

Source organism: Human

Seed type: 7mer-m8 / 8mer

miRNAs must target all input sequences: YES

Minimum number of binding sites on each sequence: 1

Minimum GC content allowed: 23%

Maximum GC content allowed: 78%

Homologous miRNA filter enabled: YES

Repeats filter enabled: YES

**miRNA list**

**1) miRNA ID:141**

Sequence: UUCCAAUUCGAGGGGAGGUGGG

Total binding sites: 2

Sites on MET: 1

Sites on EGFR: 1

*Scores*

Seed score: 1.0

3' Match score: 0.6785714285714286

AU Content score: 0.7

Nucleotide Composition score: 0.6703296703296704

Structural Accessibility score: 0.18903029249999997

ARE/CPE score: 0.6875

-------------------------------------------------------------------------

C-tree score: -0.22488171570287555

M5P score: -0.26212105405543135

*Binding sites details*

Target: MET

Position: 1240

Type: 8mer

Alignment:

 G GAG U

3'-GGGUGGAG G C UUAACCU -5' miRNA

 |||||||| | | |||||||

5'-CCCACCUC C G AAUUGGA -3' mRNA

 G AA C A

Free energy: -26.0 Kcal/mol

Target: EGFR

Position: 1375

Type: 8mer

Alignment:

 GGG A C U

3'- UGGAGGGG G UUAACCU -5' miRNA

 |||||||| | |||||||

5'- AUUUCUCU C AAUUGGA -3' mRNA

 AAUA A A

Free energy: -15.4 Kcal/mol

===================================

**2) miRNA ID:23**

Sequence: UCAAUUUCGGUCCCGAGUUCCA

Total binding sites: 2

Sites on MET: 1

Sites on EGFR: 1

*Scores*

Seed score: 1.0

3' Match score: 0.4642857142857143

AU Content score: 0.65

Nucleotide Composition score: 0.6904761904761905

Structural Accessibility score: 0.19590275007499997

ARE/CPE score: 0.4375

-------------------------------------------------------------------------

C-tree score: -0.22488171570287555

M5P score: -0.25808594942969443

*Binding sites details*

Target: MET

Position: 298

Type: 8mer

Alignment:

 ACCUUGA CUGG

3'- GCC CUUUAACU-5' miRNA

 ||| ||||||||

5'- UGG GAAAUUGA-3' mRNA

 UUGAUUUCAUA

Free energy: -9.2 Kcal/mol

Target: EGFR

Position: 1299

Type: 8mer

Alignment:

 A C G U

3'- C UUGA CCC GGCUUUAACU-5' miRNA

 | |||| ||| ||||||||||

5'- G AACU GGG UUGAAAUUGA-3' mRNA

 U A A U

Free energy: -18.2 Kcal/mol

===================================

**3) miRNA ID:140**

Sequence: UCCAAUUGGACGGGAGGUGGGU

Total binding sites: 2

Sites on MET: 1

Sites on EGFR: 1

*Scores*

Seed score: 1.0

3' Match score: 0.5714285714285714

AU Content score: 0.6499999999999999

Nucleotide Composition score: 0.7307692307692307

Structural Accessibility score: 0.19208470612500006

ARE/CPE score: 0.6875

-------------------------------------------------------------------------

C-tree score: -0.22488171570287555

M5P score: -0.2498115468414106

*Binding sites details*

Target: MET

Position: 1239

Type: 8mer

Alignment:

 U G CAG

3'- GGGUGGAG G GUUAACCU-5' miRNA

 |||||||| | ||||||||

5'- CCCACCUC C CAAUUGGA-3' mRNA

 U G AAG

Free energy: -27.3 Kcal/mol

Target: EGFR

Position: 1374

Type: 8mer

Alignment:

 UGGG CAG

3'- UGGAGGG GUUAACCU-5' miRNA

 ||||||| ||||||||

5'- AUUUCUC CAAUUGGA-3' mRNA

 AAAUA UA

Free energy: -16.0 Kcal/mol

===================================

**4) miRNA ID:106**

Sequence: UUUCAUGAGCCCUAGACUGGGG

Total binding sites: 2

Sites on MET: 1

Sites on EGFR: 1

*Scores*

Seed score: 1.0

3' Match score: 0.6071428571428572

AU Content score: 0.65

Nucleotide Composition score: 0.7472527472527473

Structural Accessibility score: 0.280897115675

ARE/CPE score: 0.5625

-------------------------------------------------------------------------

C-tree score: -0.22488171570287555

M5P score: -0.24639068160688152

*Binding sites details*

Target: MET

Position: 944

Type: 8mer

Alignment:

 G AUCCCG U

3'- GGGUCAG AGUACUU -5' miRNA

 ||||||| |||||||

5'-  UUUAGUC UCAUGAA -3' mRNA

 AAAUAA A A

Free energy: -12.4 Kcal/mol

Target: EGFR

Position: 776

Type: 8mer

Alignment:

 U UCC U

3'-GGGG CAGA CG AGUACUU -5' miRNA

 |||| |||| || |||||||

5'-CCCC GUCU GC UCAUGAA -3' mRNA

 U U UG A

Free energy: -21.0 Kcal/mol

===================================

**5) miRNA ID:196**

Sequence: UGAGUUUCUCAGCGACGGACCG

Total binding sites: 2

Sites on MET: 1

Sites on EGFR: 1

*Scores*

Seed score: 1.0

3' Match score: 0.5714285714285714

AU Content score: 0.7

Nucleotide Composition score: 0.772893772893773

Structural Accessibility score: 0.2795060730000001

ARE/CPE score: 0.8125

-------------------------------------------------------------------------

C-tree score: -0.22488171570287555

M5P score: -0.24107103610939157

*Binding sites details*

Target: MET

Position: 1731

Type: 8mer

Alignment:

 G G C ACU U

3'- CCA GCAG G CUUUGAG -5' miRNA

 ||| |||| | |||||||

5'- GGU UGUC C GAAACUC -3' mRNA

 G G A ACU A

Free energy: -21.2 Kcal/mol

Target: EGFR

Position: 746

Type: 8mer

Alignment:

 GCC C CU U

3'- AGG AGCGA CUUUGAG -5' miRNA

 ||| ||||| |||||||

5'- UCC UUGUU GAAACUC -3' mRNA

 CAU A UU A

Free energy: -17.9 Kcal/mol

===================================

**6) miRNA ID:98**

Sequence: UUUCUUAAGCACGCCGUUGGGG

Total binding sites: 2

Sites on MET: 1

Sites on EGFR: 1

*Scores*

Seed score: 1.0

3' Match score: 0.5714285714285714

AU Content score: 0.8

Nucleotide Composition score: 0.783882783882784

Structural Accessibility score: 0.32562989699999995

ARE/CPE score: 0.6875

-------------------------------------------------------------------------

C-tree score: -0.22488171570287555

M5P score: -0.23874056862383553

*Binding sites details*

Target: MET

Position: 913

Type: 8mer

Alignment:

 G GG C CACG U

3'- G UUGC G AAUUCUU -5' miRNA

 | |||| | |||||||

5'- C GAUG C UUAAGAA -3' mRNA

 GUA AUA A A A

Free energy: -6.7 Kcal/mol

Target: EGFR

Position: 1649

Type: 8mer

Alignment:

 U C C G U

3'-GGGG UG CG AC AAUUCUU -5' miRNA

 |||| || || || |||||||

5'-CCCC AC GC UG UUAAGAA -3' mRNA

 U A AU A

Free energy: -16.4 Kcal/mol

**Table S6.** Complete list of synthetic miRNAs for c-MET

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Rank** | **ID** | **Sequence** | **Sites** | **M5P score** | **Ctree score** |
| 1 | 60 | UUUGAAACGGAGGCUGUCUAGA | 3 | -0.26149351 | -0.224881716 |
| 2 | 118 | UUUAUAAAGUCGAUACGUGUUU | 3 | -0.260479749 | -0.224881716 |
| 3 | 181 | UUCUUUCUAAGGACGGGGCCGU | 2 | -0.253428787 | -0.224881716 |
| 4 | 176 | UCAGUACAAAACCUUGUGGCUU | 2 | -0.246296285 | -0.224881716 |
| 5 | 108 | UAAAUAUUAGACGACGAAUAAG | 3 | -0.242355599 | -0.224881716 |
| 6 | 165 | UUCCACAGUCGGGAGCUUCUCU | 2 | -0.240768881 | -0.224881716 |
| 7 | 167 | UGCACUAUGGCGGAGCUUGCAG | 2 | -0.234875889 | -0.224881716 |
| 8 | 35 | UAUAUGACUCCUGUGUUGUCCC | 2 | -0.231553603 | -0.224881716 |
| 9 | 53 | UUGAGUUUAGCUUUAACAAAAC | 3 | -0.231103438 | -0.224881716 |
| 10 | 81 | UACAAACAACGUGAGGGGUCUG | 2 | -0.230296126 | -0.224881716 |
| 11 | 140 | UUUGAGUUAGGGUGUGACACAC | 2 | -0.229173718 | -0.224881716 |
| 12 | 13 | UCAAGUCCACAUGUUGUGCUGU | 2 | -0.226153654 | -0.224881716 |
| 13 | 89 | UUGCAGUGAAGUCGGGGCCGAA | 3 | -0.225760484 | -0.224881716 |
| 14 | 173 | UUCAUGCUCAGGUCCCUGGCAG | 2 | -0.225516386 | -0.224881716 |
| 15 | 128 | UUAUUUCAACGUGACUUUCUUA | 2 | -0.225290884 | -0.224881716 |
| 16 | 36 | UCAAGUUUACCACUCGCAUACG | 2 | -0.223165296 | -0.224881716 |
| 17 | 43 | UUCCUGCUAGUCUUUCCCGCUG | 2 | -0.222018082 | -0.224881716 |
| 18 | 69 | UGUAAUGACCCUGGGGUGUGUC | 2 | -0.216304624 | -0.224881716 |
| 19 | 152 | CCAAUUGCGACCGUGGUGGGUU | 2 | -0.213761756 | -0.224881716 |
| 20 | 178 | UAGUCCUGAACAUGGGAUGUGG | 2 | -0.211142676 | -0.224881716 |
| 21 | 124 | UGUACAUUAGGUCCUGGCCAUU | 2 | -0.210068602 | -0.224881716 |
| 22 | 161 | UAAUAGCUCCCCUGUUCAAUUG | 2 | -0.20750942 | -0.224881716 |
| 23 | 164 | UCCACAGACGCGAGCUUCUCUU | 2 | -0.205812794 | -0.224881716 |
| 24 | 22 | UUUCCUUACGCCCCUGUACUGU | 2 | -0.204985723 | -0.224881716 |
| 25 | 41 | GGCUUCAACAGCCGAUGUGGUA | 2 | -0.202127102 | -0.224881716 |
| 26 | 61 | UUUUGAAAGGGUGGCUGUCUAG | 3 | -0.201722794 | -0.224881716 |
| 27 | 155 | UUUUGUUUGGUAUGGCUGGCGU | 2 | -0.200881894 | -0.224881716 |
| 28 | 143 | UCAUGCUGACAUUUCUCCCAGU | 2 | -0.20029129 | -0.224881716 |
| 29 | 21 | CCUUAGAACGGUGUCUUUCUCC | 2 | -0.197161863 | -0.224881716 |
| 30 | 141 | UAUUACUUCAGUCAUUUUGGGU | 2 | -0.196153669 | -0.224881716 |
| 31 | 38 | CAGAAUUCAUCUUGUGGGGCCC | 2 | -0.195643221 | -0.224881716 |
| 32 | 131 | UAGCAUCAGAGUGUUCCGGUGG | 2 | -0.194422456 | -0.224881716 |
| 33 | 67 | CCAUACAAGAGUCCUUAGUGUC | 2 | -0.190860619 | -0.224881716 |
| 34 | 136 | UGUUCACACACCUGUGUUCUGG | 2 | -0.190305003 | -0.224881716 |
| 35 | 11 | GCCUAUAAAACUGCUUCCUUGG | 2 | -0.189374926 | -0.224881716 |
| 36 | 112 | UAUUGCCUAAAUCUUUUGUUCC | 2 | -0.189195173 | -0.224881716 |
| 37 | 135 | GUUCACAUACGUGUGUUCUGAG | 2 | -0.18825176 | -0.224881716 |
| 38 | 126 | UACAAAUUAGGUGUCUGUCUUC | 2 | -0.184152565 | -0.224881716 |
| 39 | 68 | AAAAUGUUGUGGGCGACACAAC | 3 | -0.183729956 | -0.224881716 |
| 40 | 154 | UUUGUUUCGUUUGGCUGGCGUG | 2 | -0.18348581 | -0.224881716 |
| 41 | 179 | UCUAUUGACCGGCUGUGGGUCU | 2 | -0.181952092 | -0.224881716 |
| 42 | 39 | UUAUGAAAGCCCGUGGUUCGGG | 2 | -0.179847408 | -0.224881716 |
| 43 | 153 | UUGUUUCCGGCUGCUCGCGCGG | 2 | -0.179749929 | -0.224881716 |
| 44 | 10 | CGCAAUUUACCGUCGUGGAAAG | 2 | -0.179455174 | -0.224881716 |
| 45 | 56 | UGUCCUGACCUUGUCUUGCGGG | 2 | -0.176768878 | -0.224881716 |
| 46 | 184 | GGACAAGUAAGUAAUUCGCAGA | 2 | -0.175046667 | -0.224881716 |
| 47 | 175 | GUACAUUUGGCUAUCGGCAUUU | 2 | -0.175019209 | -0.224881716 |
| 48 | 77 | UACAUUUAACUGGGGGCGUUUU | 2 | -0.17496555 | -0.224881716 |
| 49 | 117 | UAUAAACAACAUUUGUGUGAUG | 2 | -0.172624901 | -0.224881716 |
| 50 | 110 | UAUAAAUACGGUUCGUGCCUUU | 2 | -0.172598868 | -0.224881716 |
| 51 | 172 | UCAUUAAUGGGAGGCCGAGUUU | 2 | -0.172256161 | -0.224881716 |
| 52 | 114 | UACAUGUUGACUGUGAUGCGAU | 2 | -0.171576103 | -0.224881716 |
| 53 | 5 | GCCAUUGGAGGGUGGGCGGUGG | 2 | -0.213069622 | -0.192354835 |
| 54 | 4 | GACAAAGUCGCGCCCGGGGCGG | 2 | -0.21303433 | -0.192354835 |
| 55 | 122 | GUCCUGUUCUAGGGUUGUCUCG | 2 | -0.210619715 | -0.192354835 |
| 56 | 130 | GUUCUCAACGUGGGGGUUUCUA | 2 | -0.207090962 | -0.192354835 |
| 57 | 129 | GACUAAAUAAAUCCUGACGACU | 2 | -0.204147926 | -0.192354835 |
| 58 | 32 | GGCUGCAGCGGUUGGGUGCGUG | 2 | -0.203230939 | -0.192354835 |
| 59 | 26 | GAUGCUCUCCGCGCUGGGAGAG | 2 | -0.202700524 | -0.192354835 |
| 60 | 113 | GUGUAUUGCGAAUCAGUACUCU | 2 | -0.201519939 | -0.192354835 |
| 61 | 24 | CAAUUCCUACCGUCCCCGGGUU | 2 | -0.200488314 | -0.192354835 |
| 62 | 72 | AUGUUUACCAGGUGGGUGCCUG | 2 | -0.200111258 | -0.192354835 |
| 63 | 145 | GAAUGUAUGGAGCUGUGCGUAG | 2 | -0.199036351 | -0.192354835 |
| 64 | 116 | AUAAACAUCAAUUGUGUGAUGC | 2 | -0.196067856 | -0.192354835 |
| 65 | 64 | ACACACAAUUCUGCUUCGUGUC | 2 | -0.194817406 | -0.192354835 |
| 66 | 80 | ACAAACAUCGAGAGGGGUCUGU | 2 | -0.194775336 | -0.192354835 |
| 67 | 14 | AUACAAGUCGCUUGUCCGGGCG | 2 | -0.194211864 | -0.192354835 |
| 68 | 100 | AAAAUACAGUGAUUAGCCGGGG | 2 | -0.192455249 | -0.192354835 |
| 69 | 168 | GUCAUUAAAGGCAGUCUGUCUU | 2 | -0.188782353 | -0.192354835 |
| 70 | 138 | CAUGUUCAGUCUGCUGUGUUCU | 2 | -0.188472869 | -0.192354835 |
| 71 | 109 | CUAAAUAUAGGUCGAUGCCUAG | 2 | -0.186946414 | -0.192354835 |
| 72 | 16 | CAAUACAACAGCUAGCCGGGGG | 2 | -0.182203855 | -0.192354835 |
| 73 | 66 | CAACACACGGAAGUGCUUUGUA | 2 | -0.180580894 | -0.192354835 |
| 74 | 12 | CAAGUCCUCAAGUUGUGCUGUG | 2 | -0.178653125 | -0.192354835 |
| 75 | 103 | GAGGUGGGACCCUGCCUCGGGG | 2 | -0.178390727 | -0.192354835 |
| 76 | 50 | CUCAAGAACCGCUCUACAUGUG | 2 | -0.178213915 | -0.192354835 |
| 77 | 163 | GAGGGUGGCCAGUCCUGGCAGG | 2 | -0.177917914 | -0.192354835 |
| 78 | 27 | GUGAUGCUGACUCGGAUCUGUG | 2 | -0.173280732 | -0.192354835 |
| 79 | 132 | GAGUAGCAACGUCCUCAUUCCG | 2 | -0.17101295 | -0.192354835 |
| 80 | 144 | CUUUACAAAGGUGCGGGGUUUU | 2 | -0.069237272 | -0.192354835 |
| 81 | 23 | AAUUCCUUCCCUCCCCGGGUUU | 2 | -0.41024469 | -0.175175976 |
| 82 | 105 | CAUUUAUAUCCGCGGGGCGCGG | 2 | -0.395943892 | -0.175175976 |
| 83 | 55 | GUCUUCUGCCAUGUCUGUGCUC | 2 | -0.3751398 | -0.175175976 |
| 84 | 171 | CAUUAAUCGGUGGCCGAGUUUC | 2 | -0.331313964 | -0.175175976 |
| 85 | 125 | UUGUCAUCGUGGUUCUUCACUU | 2 | -0.310409533 | -0.175175976 |
| 86 | 45 | UUUUCCUGGAACAACUGCCCAU | 2 | -0.294615584 | -0.175175976 |
| 87 | 90 | GUUGCAGUCACCUCUGGUCUGU | 2 | -0.293919674 | -0.175175976 |
| 88 | 146 | CUUUCUUUCGACAGCCCAGAUU | 3 | -0.288951912 | -0.175175976 |
| 89 | 58 | GCCUGUUCACCGGCUGCCGCGC | 2 | -0.278983644 | -0.175175976 |
| 90 | 149 | CCAUCCACGGAUUCCUUGCCCC | 2 | -0.278349633 | -0.175175976 |
| 91 | 147 | CCUUUCUUACGUCUGCGCAGGU | 2 | -0.275113467 | -0.175175976 |
| 92 | 71 | CCCUAAUGCCAGAUUGACCCCC | 2 | -0.270184291 | -0.175175976 |
| 93 | 115 | CAAACAUGGACUGUGUGCUGCU | 2 | -0.238706749 | -0.175175976 |
| 94 | 88 | CACACCACACCCCCUCUGCCGU | 2 | -0.234691371 | -0.175175976 |
| 95 | 57 | CUCCUGUCGGCAGCAUCACUCU | 2 | -0.211722458 | -0.175175976 |
| 96 | 180 | GUUCUUUCGUGUGAGGCGUUCU | 2 | -0.207660043 | -0.175175976 |
| 97 | 85 | CUGGCAACUCCGCGAGACCCCG | 2 | -0.196272954 | -0.175175976 |
| 98 | 79 | CAAACAUUAAAAUGGGAAUGUU | 2 | -0.185803729 | -0.175175976 |
| 99 | 52 | CUCUGUGAAGAAGUAGCCCUCA | 2 | -0.172565783 | -0.175175976 |
| 100 | 47 | AAUAUUUCGACCUGCCUUUUCC | 2 | -0.171902699 | -0.175175976 |
| 101 | 40 | CUUCAAUUCGGCUUUUGAUUUG | 2 | -0.167104685 | -0.175175976 |
| 102 | 160 | GUUUCUAAUCAGUGCUCUGCGC | 2 | -0.152025812 | -0.175175976 |
| 103 | 1 | GCUUUGACACAGUUGCCGCUCU | 2 | -0.148885354 | -0.175175976 |
| 104 | 19 | CAUUUAAAUUUGAAUACACACC | 3 | -0.145943715 | -0.175175976 |
| 105 | 95 | GGCUAUUCCCCCGGGGUGAGCC | 2 | -0.139559865 | -0.175175976 |
| 106 | 157 | CACACCCUGAAGGUAGCUUGCU | 2 | -0.112879723 | -0.175175976 |
| 107 | 46 | CAUUUCCUCGACGCAUUUCCCA | 2 | -0.112134257 | -0.175175976 |
| 108 | 18 | CUUUAAAUCAGUACGCCGGCCC | 2 | -0.074502814 | -0.175175976 |
| 109 | 137 | AUGUUCACUCAGCUGUGUUCUG | 2 | -0.071339198 | -0.175175976 |
| 110 | 34 | AUAUGACACCAGUGUUGUCCCG | 2 | -0.070561764 | -0.175175976 |
| 111 | 83 | GAUCCUGUGGGUUUAUACACAC | 2 | -0.05988612 | -0.056348738 |

**Table S7.** Complete list of synthetic miRNAs for EGFR

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Rank** | **ID** | **Sequence** | **Sites** | **M5P score** | **Ctree score** |
| 1 | 3 | UGUGGCUUCACCUCCUGUAUCG | 3 | -0.241109911 | -0.224881716 |
| 2 | 106 | UGUGUGACACUGCGUAAGGGGG | 2 | -0.237930433 | -0.224881716 |
| 3 | 25 | CAAAUGCUCGAGAGUCCGAUGU | 2 | -0.229263136 | -0.224881716 |
| 4 | 83 | UAACAAUGCACUGGGGGCCCUG | 2 | -0.227948932 | -0.224881716 |
| 5 | 80 | GUUUACAAUCCUCUGCCGUUGG | 2 | -0.215213766 | -0.224881716 |
| 6 | 30 | CUCUUGCUCCAUGCGUGUCUGC | 2 | -0.213776638 | -0.224881716 |
| 7 | 11 | GCUGUGGGGGAUUGGGCCAAUG | 2 | -0.213676632 | -0.224881716 |
| 8 | 99 | UCUGAUCUCCCCGCCAAAGGUC | 2 | -0.213178355 | -0.224881716 |
| 9 | 87 | UGAUUUCACCCCAGCUUGACUG | 2 | -0.210567041 | -0.224881716 |
| 10 | 34 | UGAUAAAUACCUGGGGGUGGGG | 2 | -0.209980431 | -0.224881716 |
| 11 | 29 | UCUUGCUGCAAGCGUGUCUGCA | 2 | -0.203586399 | -0.224881716 |
| 12 | 101 | UAAAGGAGUAAUCUGUGCUGCU | 2 | -0.203327846 | -0.224881716 |
| 13 | 18 | UGACAAAGCCAUGGGAGGGGGC | 2 | -0.200018437 | -0.224881716 |
| 14 | 39 | UAUACUUUGAAUUCUUUGCUGU | 2 | -0.194319957 | -0.224881716 |
| 15 | 13 | UAACCAGUGACGGGGCCGGGGA | 2 | -0.188424699 | -0.224881716 |
| 16 | 49 | UUUGAAGUCUUGUUCAAGGGGG | 2 | -0.188173837 | -0.224881716 |
| 17 | 79 | UUUACAAACCACUGCCGUUGGC | 2 | -0.186924671 | -0.224881716 |
| 18 | 95 | UAGGAGAUAGCUGUGCGGGUGC | 2 | -0.185762278 | -0.224881716 |
| 19 | 21 | GGAAGACAUUCGAGUGCGUCUU | 2 | -0.183769423 | -0.224881716 |
| 20 | 84 | GAGUUUCAAUACUACGGCCUGG | 2 | -0.182995268 | -0.224881716 |
| 21 | 5 | UGCUGUGGCGAGGUGGUGGGUA | 2 | -0.175358928 | -0.224881716 |
| 22 | 53 | UUCUUCCUACAUGGAUGGGCCG | 2 | -0.174828175 | -0.224881716 |
| 23 | 68 | GGCCAUGAGCCGCUUCUGGGCC | 2 | -0.172446058 | -0.224881716 |
| 24 | 52 | UCUUCCUUCAAGGAUGGGCCGG | 2 | -0.057443635 | -0.224881716 |
| 25 | 97 | GAUUAGGACUAACGGUGGGGCG | 2 | -0.21222741 | -0.192354835 |
| 26 | 12 | AACCAGUCACCGGGCCGGGGAG | 2 | -0.20811832 | -0.192354835 |
| 27 | 66 | GAACCACUCCACCCUCGGUGGG | 2 | -0.206795205 | -0.192354835 |
| 28 | 59 | CAUCCUCUCCAGGUGUUGUGGC | 2 | -0.204106947 | -0.192354835 |
| 29 | 19 | AAGACAAACUUGUGAGACUUCC | 3 | -0.203066258 | -0.192354835 |
| 30 | 75 | GAAUUCCAGGGCCCCCGUUUCG | 2 | -0.203044354 | -0.192354835 |
| 31 | 109 | GAAGGAACAAAUUGCCUGCGUG | 2 | -0.20129902 | -0.192354835 |
| 32 | 77 | CAGUCUAACCGUGAUGUCCGGC | 2 | -0.195392425 | -0.192354835 |
| 33 | 9 | GAGGACCUCGAGCGGCUCGGCC | 2 | -0.194200831 | -0.192354835 |
| 34 | 10 | GGUCUAAGCCGGAAUGCGGGCC | 2 | -0.194120035 | -0.192354835 |
| 35 | 15 | CAAAGGAAGCGUUCGUCCGUGU | 2 | -0.193665793 | -0.192354835 |
| 36 | 27 | GCGUUUCUCGAUUGGCCUGGCA | 2 | -0.186045989 | -0.192354835 |
| 37 | 31 | AAGGGACACACUGGCUGCUGGU | 2 | -0.185644896 | -0.192354835 |
| 38 | 41 | ACAUAUACAGAUUGUGUUUUUC | 2 | -0.185605154 | -0.192354835 |
| 39 | 105 | GGGAACAUAAAGUCUGUGUGUG | 2 | -0.18253095 | -0.192354835 |
| 40 | 98 | CAAAGGACGCCUGGCAUCCCGU | 2 | -0.181851227 | -0.192354835 |
| 41 | 46 | GAAGUAAAUCCCGAUAGCGACC | 2 | -0.179760435 | -0.192354835 |
| 42 | 16 | ACAAAGGAGACCAGGUACCCAA | 3 | -0.176861267 | -0.192354835 |
| 43 | 40 | CAUAUACUGAAUGUGUUUUUCG | 2 | -0.047420846 | -0.192354835 |
| 44 | 74 | GUUUCUUCGAAGCCCAGGUGGG | 2 | -0.386216637 | -0.175175976 |
| 45 | 51 | CUUCCUUGAACGAUGGGCCGGG | 2 | -0.25020063 | -0.175175976 |
| 46 | 57 | GUUGUGGCGACUGCUGCUUGCG | 2 | -0.201286859 | -0.175175976 |
| 47 | 22 | GUUUGAAGACUAGGGAAUGGAG | 2 | -0.199922587 | -0.175175976 |
| 48 | 35 | CUCUUUGACACAUAAGCUUCUG | 2 | -0.173773465 | -0.175175976 |
| 49 | 92 | CAAUGCUGAACGGGCCCUGGCG | 2 | -0.168048984 | -0.175175976 |
| 50 | 72 | GCUACUGUGGGUGGCACGUGUU | 2 | -0.166946483 | -0.175175976 |
| 51 | 102 | AUCUUCCAAAAGUAGCGCAAUC | 2 | -0.163898438 | -0.175175976 |
| 52 | 94 | UAUUCCACUGACCCCGCUGUGG | 2 | -0.162939094 | -0.175175976 |
| 53 | 104 | GGAACAUAAGGGCUGGGUGUAA | 2 | -0.157510275 | -0.175175976 |
| 54 | 33 | GAUAAAUUCCAGGGGGUGGGGA | 2 | -0.127141007 | -0.175175976 |
| 55 | 44 | CAAUCAAUUCGGACCCUGACAU | 2 | -0.076113199 | -0.175175976 |
| 56 | 71 | CCAUGACUCCUUACCGCUGGGC | 2 | -0.074602147 | -0.056348738 |
| 57 | 108 | CUGUUGUGCGGGGAGCUGCAAG | 2 | -0.074028151 | -0.056348738 |
| 58 | 17 | GACAAAGGUGAGCUGCUUCCCA | 2 | -0.064169449 | -0.056348738 |
| 59 | 43 | AAUCAAUACGCACCCUGACAUA | 2 | -0.061562959 | -0.056348738 |

**Table S8.** Complete list of synthetic miRNAs for c-MET and EGFR.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Rank** | **ID** | **Sequence** | **Sites on c-MET** | **Sites on EGFR** | **M5P score** | **Ctree score** |
| 1 | 141 | UUCCAAUUCGAGGGGAGGUGGG | 1 | 1 | -0.262121054 | -0.224881716 |
| 2 | 23 | UCAAUUUCGGUCCCGAGUUCCA | 1 | 1 | -0.258085949 | -0.224881716 |
| 3 | 140 | UCCAAUUGGACGGGAGGUGGGU | 1 | 1 | -0.249811547 | -0.224881716 |
| 4 | 106 | UUUCAUGAGCCCUAGACUGGGG | 1 | 1 | -0.246390682 | -0.224881716 |
| 5 | 196 | UGAGUUUCUCAGCGACGGACCG | 1 | 1 | -0.241071036 | -0.224881716 |
| 6 | 98 | UUUCUUAAGCACGCCGUUGGGG | 1 | 1 | -0.23874806 | -0.224881716 |
| 7 | 204 | UGUCAUCCAGGCUUGCCGAUUC | 1 | 1 | -0.234493846 | -0.224881716 |
| 8 | 137 | UACCUGCCUGCGCGUGAGAGGG | 1 | 1 | -0.232062736 | -0.224881716 |
| 9 | 18 | GACCUGAUCGAUUUCAUCUGCG | 1 | 1 | -0.231204838 | -0.224881716 |
| 10 | 24 | UUCAAUUUGGGACCCGAGUUCC | 1 | 1 | -0.2289542 | -0.224881716 |
| 11 | 134 | UUCCAUCCGGGCCGUCUCUGCU | 1 | 1 | -0.228564665 | -0.224881716 |
| 12 | 55 | UUGCAGUGAAGUCGUGGCAGAA | 3 | 1 | -0.228046307 | -0.224881716 |
| 13 | 194 | UGUGUGACUCUGCGUAAGUGUA | 1 | 2 | -0.225800008 | -0.224881716 |
| 14 | 66 | GGAAACCACGCGUGGCUCCGUG | 1 | 1 | -0.225675946 | -0.224881716 |
| 15 | 105 | UGAUGACUGUUGGCUGUGACAC | 1 | 1 | -0.225395337 | -0.224881716 |
| 16 | 104 | GAAAUUAUAAGAGGCAGGUUCU | 1 | 1 | -0.220162553 | -0.224881716 |
| 17 | 33 | UACUAUCUGGGCUGAGUUUCCC | 1 | 1 | -0.219504707 | -0.224881716 |
| 18 | 6 | UUCCUUAGCAACCCGUCAUUCU | 1 | 1 | -0.218806489 | -0.224881716 |
| 19 | 119 | UCAAAUUACCAGUCUCGUUUUG | 1 | 1 | -0.217631437 | -0.224881716 |
| 20 | 127 | GUUUACAAUUCUCUGUGGUUAA | 2 | 2 | -0.215926957 | -0.224881716 |
| 21 | 182 | GUCAUUGACUGUCAUGCCGUCU | 1 | 1 | -0.214027874 | -0.224881716 |
| 22 | 139 | CUUGCGAGCACGGAAGCGGCGC | 1 | 1 | -0.213417504 | -0.224881716 |
| 23 | 190 | CUUUGUAGGGAGCUGGGCUAUG | 1 | 1 | -0.21275742 | -0.224881716 |
| 24 | 5 | UGAAUCCACACUGGGCUGGAAC | 1 | 1 | -0.2123791 | -0.224881716 |
| 25 | 1 | GGACAAAGCCAGGGGAGUUGCU | 1 | 2 | -0.21178937 | -0.224881716 |
| 26 | 199 | UGUCUUCUCCACUGUGCCCGCA | 1 | 1 | -0.20861562 | -0.224881716 |
| 27 | 123 | GAUCUUUACACUUUUGCAGCGG | 1 | 1 | -0.20765557 | -0.224881716 |
| 28 | 44 | CAAUGAAAUUGGCCAUGGUCCC | 1 | 1 | -0.205235717 | -0.224881716 |
| 29 | 50 | GACAAACAACACAAGGUAACUG | 2 | 1 | -0.203109841 | -0.224881716 |
| 30 | 210 | GGAUGCGUAGGUGCUCUUGCCC | 1 | 1 | -0.202085781 | -0.224881716 |
| 31 | 149 | UAUUCCAUGGGGGCGGUUUCGC | 1 | 1 | -0.200549294 | -0.224881716 |
| 32 | 111 | CCACAUUCAGAGGUGCGGGGUA | 1 | 1 | -0.200046043 | -0.224881716 |
| 33 | 169 | CACUUUAAUACCUUACGGCGCC | 1 | 1 | -0.197034181 | -0.224881716 |
| 34 | 207 | CAGAGACACAACUUGCGUGCCA | 1 | 1 | -0.196664641 | -0.224881716 |
| 35 | 211 | CCCGAGUUGAAUCUUGUGGUGC | 1 | 1 | -0.196288461 | -0.224881716 |
| 36 | 131 | UUCUUUGCGGAGGCUGGGAGCG | 1 | 1 | -0.196200539 | -0.224881716 |
| 37 | 16 | CAGAAUUCGUCCUGCGGGGUCU | 2 | 1 | -0.195419698 | -0.224881716 |
| 38 | 29 | CUGAGUUUGGUUUUAACAAAAU | 3 | 1 | -0.194295453 | -0.224881716 |
| 39 | 60 | CUGAGCCUCCCAGGUGGAGGUU | 1 | 1 | -0.193240516 | -0.224881716 |
| 40 | 85 | UCUUCCAAGGCUAGGGAUGUCU | 1 | 1 | -0.192665497 | -0.224881716 |
| 41 | 151 | GGGCUAUUCCACCUAGGGGUGA | 1 | 1 | -0.192460042 | -0.224881716 |
| 42 | 187 | CUGCUGGGCGGACGCUGGGCUU | 1 | 1 | -0.192256358 | -0.224881716 |
| 43 | 126 | UUUACAAACCAGCGGCGUUAGC | 1 | 2 | -0.186986322 | -0.224881716 |
| 44 | 181 | CCAUUGAACAACAUGCCGUCAG | 1 | 1 | -0.186901168 | -0.224881716 |
| 45 | 43 | AAAAUGUUCAGGACGCAAUAUC | 3 | 1 | -0.183772168 | -0.224881716 |
| 46 | 3 | UACAAGUCGCACUAUCGCGCAU | 1 | 1 | -0.183220195 | -0.224881716 |
| 47 | 195 | GAGUUUCACUCUGACGGACUGG | 1 | 2 | -0.182550608 | -0.224881716 |
| 48 | 4 | UUAACAAUCGAGGUGGGGCUCU | 1 | 1 | -0.178444851 | -0.224881716 |
| 49 | 202 | UAUACUCUCACAGUUAGUAGCU | 1 | 1 | -0.17390764 | -0.224881716 |
| 50 | 46 | UCCAUCAGACCGGAGUAAGGUC | 1 | 1 | -0.173460973 | -0.224881716 |
| 51 | 42 | UUAUUGACGUCGCGAGUCUCAA | 1 | 1 | -0.173403994 | -0.224881716 |
| 52 | 150 | GUAUUCCAGGAUGGCGGCUGUG | 1 | 1 | -0.214676798 | -0.192354835 |
| 53 | 121 | CAUAUUUAGGUCCUGAUUUCCU | 1 | 1 | -0.214215256 | -0.192354835 |
| 54 | 90 | CAAUAUUCAACCUGGCUGCGGU | 1 | 1 | -0.213315585 | -0.192354835 |
| 55 | 37 | GAAACACUCCAUUCCUAGGUGU | 1 | 1 | -0.211610023 | -0.192354835 |
| 56 | 19 | GUACCUGAAAACCACACUACUC | 1 | 1 | -0.209476917 | -0.192354835 |
| 57 | 198 | GGUAAGUCAGAGGACGAUCUGC | 1 | 1 | -0.208297186 | -0.192354835 |
| 58 | 143 | AAAACUCCGGCAUCGCCUUGCU | 1 | 1 | -0.205182651 | -0.192354835 |
| 59 | 161 | GAAUUCCAGUGCAUGCAUCAUG | 1 | 2 | -0.203956341 | -0.192354835 |
| 60 | 117 | AAAUUACUACACUUGCUUUGGG | 1 | 1 | -0.203833332 | -0.192354835 |
| 61 | 179 | GCUGUCAUAGCCCCCUGUCUGA | 1 | 1 | -0.199636827 | -0.192354835 |
| 62 | 128 | CCAAUGAAGGACUCCGUGAUCC | 1 | 1 | -0.199015364 | -0.192354835 |
| 63 | 68 | GAGUUUGAUCGCAGCGGUGGGC | 1 | 1 | -0.198809637 | -0.192354835 |
| 64 | 51 | ACAAACAACGGUGUCUAAGGGU | 1 | 1 | -0.198105301 | -0.192354835 |
| 65 | 175 | GUUCACAGACGGGUGUCUUUGG | 1 | 1 | -0.194653159 | -0.192354835 |
| 66 | 80 | AAGGAAUGGGACUUCCCAAAUU | 1 | 1 | -0.193807589 | -0.192354835 |
| 67 | 82 | CAAAGGAAACACUCUUCCGAAA | 1 | 2 | -0.193715957 | -0.192354835 |
| 68 | 63 | CAAAUUAGCGCGGGCGGGUGGG | 1 | 1 | -0.193548681 | -0.192354835 |
| 69 | 177 | CCCAGUCUCACGGUCCUUGUGC | 1 | 1 | -0.192488442 | -0.192354835 |
| 70 | 164 | GGUAGCAACGCCUACCUGCCAG | 1 | 1 | -0.189526401 | -0.192354835 |
| 71 | 48 | CAAAGGGAGACUUUUCCUGCUG | 1 | 1 | -0.186758484 | -0.192354835 |
| 72 | 129 | GAGCUGUGCCGGGUUGGUCCGA | 1 | 1 | -0.184923686 | -0.192354835 |
| 73 | 67 | AGUUUGAACGGAGCGGUGGGCA | 1 | 1 | -0.184721793 | -0.192354835 |
| 74 | 83 | CCAAAGGAGACCAUGUACCCAA | 1 | 3 | -0.181226913 | -0.192354835 |
| 75 | 112 | AAACAUCUCGUUGUGCUGUGUC | 1 | 1 | -0.179846756 | -0.192354835 |
| 76 | 118 | CAAAUUACAACUCUUGCUUUGG | 1 | 1 | -0.179833043 | -0.192354835 |
| 77 | 57 | GAGGUGGACCAUGCUGCGGGCU | 1 | 1 | -0.177332532 | -0.192354835 |
| 78 | 70 | GUGGAUUUGAAGGGCCCAGGCG | 1 | 1 | -0.173363013 | -0.192354835 |
| 79 | 135 | GGACAGAGCCCCUUCUCCUCCU | 1 | 1 | -0.17316924 | -0.192354835 |
| 80 | 81 | AAAGGAAUCGGUCUUCCCAAAU | 1 | 1 | -0.172564314 | -0.192354835 |
| 81 | 197 | GUAAGUCUGACGACGAUCUGCG | 1 | 1 | -0.171923146 | -0.192354835 |
| 82 | 100 | GUUUCAUAGAAACUUCCUGAAU | 1 | 1 | -0.636826091 | -0.175175976 |
| 83 | 107 | AUUUCAUGUGCGCUAGACUGGG | 1 | 1 | -0.61727485 | -0.175175976 |
| 84 | 97 | GUGUCAUCGUGGCUUGCUAAUU | 2 | 1 | -0.593206114 | -0.175175976 |
| 85 | 99 | CUUUCUUAUGCUCGCCGUUGGG | 1 | 1 | -0.417249477 | -0.175175976 |
| 86 | 115 | CUUUGAGUAAACAGCUUUUGAA | 1 | 1 | -0.394743264 | -0.175175976 |
| 87 | 160 | AAUUCCACUCAAUGAAGCAUGG | 1 | 2 | -0.382526629 | -0.175175976 |
| 88 | 32 | GUCUUCUGUAGUGUCUCUGAAC | 2 | 1 | -0.371348388 | -0.175175976 |
| 89 | 113 | AAAACAUCACGAUGUGCUGUGU | 1 | 1 | -0.354712867 | -0.175175976 |
| 90 | 133 | CCUUUCUUACGACUGCGGAGGG | 2 | 1 | -0.348274732 | -0.175175976 |
| 91 | 76 | GGCUCAUGGGAGGAAUCCUCGG | 1 | 1 | -0.323637775 | -0.175175976 |
| 92 | 28 | GUUUCUGUCUACGCGUCGCCGU | 1 | 1 | -0.318472847 | -0.175175976 |
| 93 | 178 | CUGUCAUUGCGCCCUGUCUGAC | 1 | 1 | -0.300834678 | -0.175175976 |
| 94 | 56 | GUUGCAGUCCCCUCCGGGAAGC | 2 | 1 | -0.295746177 | -0.175175976 |
| 95 | 215 | GUCUAAGGUGCAGUUCCUUCCC | 1 | 1 | -0.289242383 | -0.175175976 |
| 96 | 157 | GAAGCAUCGGGUGGUUGUGCUG | 1 | 1 | -0.283672502 | -0.175175976 |
| 97 | 13 | CAUUGGUCGCAGGGCUGUGGGA | 1 | 1 | -0.276685141 | -0.175175976 |
| 98 | 45 | CCAUCAGUCCCGAGUAAGGUCC | 1 | 1 | -0.275227159 | -0.175175976 |
| 99 | 166 | GAUACAGUUCGGAGCCGGCCUG | 1 | 1 | -0.274337558 | -0.175175976 |
| 100 | 10 | GCCUCUGGAACUGUUGCGGCGU | 1 | 1 | -0.265060228 | -0.175175976 |
| 101 | 201 | CUUCCUUGAACGAUGGGGUGAG | 1 | 2 | -0.249959324 | -0.175175976 |
| 102 | 20 | GGUACCUGGGAUGUCUUGAUUC | 1 | 1 | -0.248825674 | -0.175175976 |
| 103 | 183 | UUUUCCUUGAGGCACUUCUGCG | 1 | 1 | -0.242047438 | -0.175175976 |
| 104 | 69 | GAUUUCUUCGCUGCUGGAGUGG | 1 | 1 | -0.227626176 | -0.175175976 |
| 105 | 96 | GAUCUAUGGGGCUUAGGUCCUG | 1 | 1 | -0.215919214 | -0.175175976 |
| 106 | 188 | CUUGCUGGCACCGCCCCUGGCU | 1 | 1 | -0.207794047 | -0.175175976 |
| 107 | 213 | CCCUGUUGACCGGAGGGCUGCC | 1 | 1 | -0.206015154 | -0.175175976 |
| 108 | 173 | GUUGUGGCGACAGCUGCUUGUU | 1 | 2 | -0.198492211 | -0.175175976 |
| 109 | 186 | GGUUAACCGCAGGGGUCACGGG | 1 | 1 | -0.188473609 | -0.175175976 |
| 110 | 142 | CUUCCAAUACGUGGGGAGGUGG | 1 | 1 | -0.178464233 | -0.175175976 |
| 111 | 86 | AUCUUCCAUGGGUAGGAAAAUC | 1 | 2 | -0.1783074 | -0.175175976 |
| 112 | 84 | CCCAAAGGUGGGGCUGCACUUC | 1 | 1 | -0.17819961 | -0.175175976 |
| 113 | 147 | GCUUAACAAACCUGUCGGGGCU | 1 | 1 | -0.160963016 | -0.175175976 |
| 114 | 209 | GAUGCGUUGGAGCUCUUGCCCC | 1 | 1 | -0.154353507 | -0.175175976 |
| 115 | 145 | GCUAAUGCCCCCUUGGCCGCCG | 1 | 1 | -0.14778502 | -0.175175976 |
| 116 | 61 | GUGUAGUCGACUGGAUUGGGGG | 2 | 1 | -0.147318953 | -0.175175976 |
| 117 | 185 | UCCUGAUGGACUUCUGCGGGGG | 1 | 1 | -0.143788441 | -0.175175976 |
| 118 | 116 | AAUUACUUCAGACAUUUAGGGA | 2 | 1 | -0.139345459 | -0.175175976 |
| 119 | 132 | CUUCUUUGGGGUGGCUGGGAGC | 1 | 1 | -0.136639271 | -0.175175976 |
| 120 | 2 | GAACCAUUCCACCCUGGGUGGA | 1 | 1 | -0.133880309 | -0.175175976 |
| 121 | 88 | AUAUUCUUCGAGGCUGCGGUUC | 1 | 1 | -0.127301118 | -0.175175976 |
| 122 | 169 | GACACUUUAACUGUGUCCGUGG | 1 | 1 | -0.110378135 | -0.175175976 |
| 123 | 188 | UCUUGCUGCCAUCAUGCCUGGA | 1 | 2 | -0.086690594 | -0.175175976 |
| 124 | 167 | AAACCAGUGUCGGUACCGGACA | 1 | 2 | -0.075914566 | -0.175175976 |
| 125 | 170 | CAAAUAUAGGACUUUGAAUGUU | 1 | 1 | -0.061478711 | -0.056348738 |

**Table S9.** Mutagenesis’ primers for c-MET and EGFR.

|  |  |  |
| --- | --- | --- |
| **Primer name** | **Target gene** | **Sequence** |
| M-60-mut1 FW | c-MET | 5'-ccactcatttagaattctagtgaacacttttgtgtgttgtatgg-3' |
| M-60-mut1 RV | c-MET | 5'-ccatacaacacacaaaagtgttcactagaattctaaatgagtgg-3' |
| M-60-mut2 FW | c-MET | 5'-tgttgccaaggctggactcctggactcaag-3' |
| M-60-mut2 RV | c-MET | 5'-cttgagtccaggagtccagccttggcaaca-3' |
| M-60-mut3 FW | c-MET | 5'-gaatatttataggcaatacagtcaaagaatagcatcacacaaaacatgtttata-3' |
| M-60-mut3 RV | c-MET | 5'-tataaacatgttttgtgtgatgctattctttgactgtattgcctataaatattc-3' |
| ME-196-mutM FW | c-MET | 5'-gagggtgtgtcacactgaatagttgagtttggct-3' |
| ME-196-mutM RV | c-MET | 5'-agccaaactcaactattcagtgtgacacaccctc-3' |
| E-3-mut1 FW | EGFR | 5'-tcgatacccaggaccaagcaggtcctcc-3' |
| E-3-mut1 RV | EGFR | 5'-ggaggacctgcttggtcctgggtatcga-3' |
| E-3-mut2 FW | EGFR | 5'-tgtgagcaaggagcacaaagtcttccagaggatg-3' |
| E-3-mut2 RV | EGFR | 5'-catcctctggaagactttgtgctccttgctcaca-3' |
| E-3-mut3 FW | EGFR | 5'-gtcatggcaggtacagtaggatatctgtcccttcc-3' |
| E-3-mut3 RV | EGFR | 5'-ggaagggacagatatcctactgtacctgccatgac |
| ME-196-mutE FW | EGFR | 5'-tccattccattgttttgagtatgctgcccctgtc-3' |
| ME-196-mutE RV | EGFR | 5'-gacaggggcagcatactcaaaacaatggaatgga-3' |

**Table S10. Shared 7mer sequences for up-regulated gene pairs and triplets in diseases.** For each disease, the table reports the number of up-regulated genes, the total number of pairs (triplets), the number of pairs (triplets) sharing at least a 7mer sequence and the average number of shared 7mers for pairs (triplets). The pairs analysis was performed for a total of 179 diseases, the triplets analysis was performed for a subset of 83 diseases, due to the high number of combinations for the remaining 96 diseases.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CONDITION** | **GENES** | **PAIRS** | **SHARING 7mer** | **SHARING 7mer %** | **AVG 7mer #** | **TRIPLETS** | **SHARING 7mer** | **SHARING 7mer %** | **AVG 7mer #** |
| benign hyperplastic enlarged lobular unit | 3 | 3 | 3 | 100.00% | 48.33333333 | 1 | 1 | 100.00% | 5 |
| pancreatic ductal carcinoma | 3 | 3 | 3 | 100.00% | 89.66666667 | 1 | 1 | 100.00% | 1 |
| autoimmune pancreatitis | 4 | 6 | 6 | 100.00% | 160.6666667 | 4 | 4 | 100.00% | 21.5 |
| spina bifida | 4 | 6 | 6 | 100.00% | 58.16666667 | 4 | 3 | 75.00% | 2.75 |
| hepatitis C virus, hepatocellular carcinoma | 4 | 6 | 6 | 100.00% | 61.5 | 4 | 4 | 100.00% | 7.5 |
| ductal carcinoma in situ | 5 | 10 | 10 | 100.00% | 68.5 | 10 | 10 | 100.00% | 7.4 |
| Muenke syndrome | 5 | 10 | 10 | 100.00% | 35.6 | 10 | 7 | 70.00% | 2.2 |
| sporadic amyotrophic lateral sclerosis | 5 | 10 | 10 | 100.00% | 63.7 | 10 | 9 | 90.00% | 7 |
| squamous cell carcinoma | 5 | 10 | 10 | 100.00% | 79.1 | 10 | 8 | 80.00% | 4.8 |
| actinic keratosis | 6 | 15 | 13 | 86.67% | 27.53333333 | 20 | 9 | 45.00% | 1.05 |
| HIV-1 infected | 7 | 21 | 21 | 100.00% | 107.8571429 | 35 | 27 | 77.14% | 10.54285714 |
| non-progressive HIV infection | 7 | 21 | 21 | 100.00% | 134.4285714 | 35 | 30 | 85.71% | 22.85714286 |
| Nonsyndromic Sagittal Synostosis | 10 | 45 | 45 | 100.00% | 405.0666667 | 120 | 120 | 100.00% | 116.55 |
| pauciarticular rheumatoid arthritis | 11 | 55 | 49 | 89.09% | 23.30909091 | 165 | 63 | 38.18% | 1.460606061 |
| peripheral arterial disease (PAD) | 12 | 66 | 66 | 100.00% | 363.4242424 | 220 | 220 | 100.00% | 94.52727273 |
| diabetic nephropathy | 13 | 78 | 76 | 97.44% | 29.23076923 | 286 | 150 | 52.45% | 1.98951049 |
| spondyloarthropathy | 15 | 105 | 103 | 98.10% | 62.1047619 | 455 | 336 | 73.85% | 6.323076923 |
| invasive ductal carcinoma | 16 | 120 | 105 | 87.50% | 101.2833333 | 560 | 416 | 74.29% | 15.25 |
| familial amyotrophic lateral sclerosis | 19 | 171 | 161 | 94.15% | 61.01169591 | 969 | 590 | 60.89% | 6.282765738 |
| adrenal gland pheochromocytoma | 24 | 276 | 266 | 96.38% | 98.30072464 | 2024 | 1522 | 75.20% | 12.08745059 |
| invasive breast cancer | 30 | 435 | 435 | 100.00% | 152.2114943 | 4060 | 3797 | 93.52% | 30.09753695 |
| Cornelia de Lange Syndrome | 35 | 595 | 585 | 98.32% | 99.53277311 | 6545 | 5415 | 82.73% | 14.07715814 |
| squamous cell lung carcinoma | 35 | 595 | 573 | 96.30% | 180.3428571 | 6545 | 5377 | 82.15% | 30.02490451 |
| asthma | 40 | 780 | 681 | 87.31% | 45.57948718 | 9880 | 4882 | 49.41% | 4.400809717 |
| malignant peripheral nerve sheath tumor | 41 | 820 | 817 | 99.63% | 98.07195122 | 10660 | 9187 | 86.18% | 13.69474672 |
| **CONDITION** | **GENES** | **PAIRS** | **SHARING 7mer** | **SHARING 7mer %** | **AVG 7mer #** | **TRIPLETS** | **SHARING 7mer** | **SHARING 7mer %** | **AVG 7mer #** |
| polyarticular rheumatoid arthritis | 42 | 861 | 860 | 99.88% | 119.2810685 | 11480 | 10817 | 94.22% | 18.87360627 |
| Barrett's esophagus | 43 | 903 | 897 | 99.34% | 130.8593577 | 12341 | 10863 | 88.02% | 19.32063852 |
| irritable bowel syndrome | 43 | 903 | 902 | 99.89% | 241.944629 | 12341 | 12013 | 97.34% | 52.63900818 |
| severe acute respiratory syndrome | 43 | 903 | 781 | 86.49% | 20.75526024 | 12341 | 3589 | 29.08% | 1.273235556 |
| type 1 diabetes | 47 | 1081 | 1079 | 99.81% | 177.2155412 | 16215 | 14771 | 91.09% | 32.44724021 |
| high-stage neuroblastoma | 51 | 1275 | 1240 | 97.25% | 90.25098039 | 20825 | 15741 | 75.59% | 11.6087395 |
| chronic HIV infection | 55 | 1485 | 1329 | 89.49% | 76.07138047 | 26235 | 15409 | 58.73% | 8.961425577 |
| peripheral T-cell lymphoma, not otherwise specified | 58 | 1653 | 1650 | 99.82% | 294.2486388 | 30856 | 29837 | 96.70% | 71.09612393 |
| dilated cardiomyopathy | 65 | 2080 | 2000 | 96.15% | 94.88125 | 43680 | 30438 | 69.68% | 13.52545788 |
| myelodysplastic syndrome | 66 | 2145 | 2077 | 96.83% | 99.55804196 | 45760 | 32500 | 71.02% | 13.45124563 |
| meningococcal sepsis | 71 | 2485 | 2440 | 98.19% | 166.1006036 | 57155 | 47188 | 82.56% | 29.22606946 |
| essential thrombocythemia | 73 | 2628 | 2567 | 97.68% | 128.1449772 | 62196 | 53673 | 86.30% | 20.28738183 |
| type 2 diabetes | 74 | 2701 | 2595 | 96.08% | 105.1384672 | 64824 | 51352 | 79.22% | 15.1589689 |
| well-differentiated liposarcoma | 79 | 3081 | 2929 | 95.07% | 97.47711782 | 79079 | 57547 | 72.77% | 13.16340621 |
| oral squamous cell carcinoma | 81 | 3240 | 3237 | 99.91% | 159.8743827 | 85320 | 79429 | 93.10% | 31.41555321 |
| pulmonary arterial hypertension | 82 | 3321 | 3142 | 94.61% | 52.78831677 | 88560 | 56217 | 63.48% | 5.328669828 |
| type 1 diabetes recent onset | 83 | 3403 | 3347 | 98.35% | 127.8521892 | 91881 | 73325 | 79.80% | 19.82117086 |
| type 1 diabetes longstanding | 87 | 3741 | 3529 | 94.33% | 97.74525528 | 105995 | 71706 | 67.65% | 13.08957026 |
| infiltrating ductal carcinoma | 89 | 3916 | 3808 | 97.24% | 141.1407048 | 113564 | 96013 | 84.55% | 22.40036455 |
| fibrosarcoma | 100 | 4950 | 4925 | 99.49% | 122.1981818 | 161700 | 144000 | 89.05% | 19.93038343 |
| pre-diabetes | 103 | 5253 | 5233 | 99.62% | 280.6613364 | 176851 | 167293 | 94.60% | 65.14948742 |
| monoclonal gammopathy of unknown significance (MGUS) | 106 | 5565 | 5315 | 95.51% | 61.61581312 | 192920 | 126641 | 65.64% | 6.877047481 |
| polycystic ovary syndrome | 108 | 5778 | 5643 | 97.66% | 94.45292489 | 204156 | 157192 | 77.00% | 13.73157291 |
| trisomy 13 | 109 | 5886 | 5845 | 99.30% | 221.2966361 | 209934 | 192391 | 91.64% | 44.54315166 |
| multiple sclerosis | 115 | 6555 | 6461 | 98.57% | 212.0147979 | 246905 | 232958 | 94.35% | 47.40512343 |
| anaplastic mixed oligoastrocytoma | 121 | 7260 | 7198 | 99.15% | 359.7424242 | 287980 | 271249 | 94.19% | 95.05410792 |
| anaplastic astrocytoma | 125 | 7750 | 7662 | 98.86% | 232.9984516 | 317750 | 286410 | 90.14% | 49.49768057 |
| lipoma | 129 | 8256 | 8072 | 97.77% | 111.4487645 | 349504 | 282588 | 80.85% | 16.35377564 |
| HIV-1 seropositive | 130 | 8385 | 7985 | 95.23% | 93.07477639 | 357760 | 255103 | 71.31% | 12.57650101 |
| uterine fibroid | 130 | 8385 | 8171 | 97.45% | 126.563864 | 357760 | 287781 | 80.44% | 20.30838551 |
| **CONDITION** | **GENES** | **PAIRS** | **SHARING 7mer** | **SHARING 7mer %** | **AVG 7mer #** | **TRIPLETS** | **SHARING 7mer** | **SHARING 7mer %** | **AVG 7mer #** |
| small cell lung carcinoma | 137 | 9316 | 9203 | 98.79% | 156.048948 | 419220 | 364277 | 86.89% | 26.07427604 |
| low-stage neuroblastoma | 142 | 10011 | 9968 | 99.57% | 191.410948 | 467180 | 439979 | 94.18% | 36.47753971 |
| anaplastic oligodendroglioma | 146 | 10585 | 10350 | 97.78% | 201.9948984 | 508080 | 447446 | 88.07% | 37.27146709 |
| Apert syndrome | 146 | 10585 | 9734 | 91.96% | 33.89560699 | 508080 | 257465 | 50.67% | 3.081548181 |
| juvenile idiopathic arthritis subtype rheumatoid factor polyarthritis | 149 | 11026 | 10807 | 98.01% | 328.9018683 | 540274 | 507787 | 93.99% | 87.26470087 |
| type 1 diabetes clinical onset | 156 | 12090 | 11908 | 98.49% | 149.7488834 | 620620 | 525536 | 84.68% | 25.96444201 |
| borderline tuberculoid leprosy | 167 | 13861 | 13832 | 99.79% | 179.5463531 | 762355 | 724591 | 95.05% | 33.56945124 |
| FAB: M2 acute myeloid leukemia | 169 | 14196 | 14137 | 99.58% | 149.3664413 | 790244 | 705980 | 89.34% | 26.94021214 |
| well-differentiated sarcoma | 169 | 14196 | 14052 | 98.99% | 134.5011975 | 790244 | 668073 | 84.54% | 20.68257779 |
| Roberts syndrome | 193 | 18528 | 18311 | 98.83% | 229.2575561 | 1179616 | 1100586 | 93.30% | 49.70269054 |
| multiple sclerosis specimens at 9th month pregnancy | 194 | 18721 | 17372 | 92.79% | 64.35356017 | 1198144 | 766507 | 63.97% | 7.362650065 |
| juvenile idiopathic arthritis subtype oligoarthritis | 212 | 22366 | 22311 | 99.75% | 233.6836269 | 1565620 | 1516958 | 96.89% | 54.33057766 |
| juvenile idiopathic arthritis subtype enthesitis related arthritis | 217 | 23436 | 22780 | 97.20% | 233.4350998 | 1679580 | 1516047 | 90.26% | 51.00654271 |
| trisomy 18 | 237 | 27966 | 26252 | 93.87% | 152.4038475 | 2190670 | 1663927 | 75.96% | 27.75153035 |
| dedifferentiated liposarcoma | 241 | 28920 | 28342 | 98.00% | 140.9757953 | 2303960 | 1992967 | 86.50% | 22.84920875 |
| extra-adrenal sympathetic paraganglioma | 243 | 29403 | 29010 | 98.66% | 147.9761929 | 2362041 | 2107507 | 89.22% | 26.49058293 |
| ischemic cardiomyopathy | 243 | 29403 | 28501 | 96.93% | 116.5983743 | 2362041 | 1823410 | 77.20% | 16.64772457 |
| FAB: M1 acute myeloid leukemia | 289 | 41616 | 40900 | 98.28% | 251.1676038 | 3981264 | 3651887 | 91.73% | 58.46850447 |
| non-basal-like breast cancer | 307 | 46971 | 46139 | 98.23% | 139.5513828 | 4775385 | 3999680 | 83.76% | 22.72177552 |
| early HIV infection | 312 | 48516 | 47688 | 98.29% | 94.85433671 | 5013320 | 4167990 | 83.14% | 14.76788755 |
| alveolar rhabdomyosarcoma | 335 | 55945 | 54843 | 98.03% | 117.7410314 | 6209895 | 5058248 | 81.45% | 16.97406333 |
| embryonal rhabdomyosarcoma | 370 | 68265 | 68043 | 99.67% | 205.6611587 | 8373840 | 7916247 | 94.54% | 43.4653029 |
| age-related macular degenerat. | 393 | 77028 | 75269 | 97.72% | 114.5133718 | 10039316 | 8440042 | 84.07% | 19.01462211 |
| Parkinson's disease | 393 | 77028 | 75287 | 97.74% | 140.7437685 | 10039316 | 8588292 | 85.55% | 22.43083503 |
| esophageal adenocarcinoma | 409 | 83436 | 80215 | 96.14% | 80.80157246 | 11319484 | 8420925 | 74.39% | 10.33293779 |
| lepromatous leprosy | 411 | 84255 | 80006 | 94.96% | 117.4213875 | 11486765 | 8866187 | 77.19% | 17.47043071 |
| Ewing sarcoma | 442 | 97461 | 92361 | 94.77% | 101.4525913 | 14294280 | 10249797 | 71.71% | 15.20315756 |
| permanent atrial fibrillation | 480 | 114960 | 111589 | 97.07% | 111.7794885 | 18316960 | 14633078 | 79.89% | 15.9088833 |
| psoriasis | 6223 | 19359753 | 18117504 | 93.58% | 73.97965021 | - | - | - | - |
| **CONDITION** | **GENES** | **PAIRS** | **SHARING 7mer** | **SHARING 7mer %** | **AVG 7mer #** | **TRIPLETS** | **SHARING 7mer** | **SHARING 7mer %** | **AVG 7mer #** |
| glioma | 5767 | 16626261 | 16089924 | 96.77% | 129.2130834 | - | - | - | - |
| periodontitis | 4974 | 12367851 | 11929214 | 96.45% | 105.7976144 | - | - | - | - |
| undifferentiated sarcoma | 4726 | 11165175 | 10830504 | 97.00% | 123.4823469 | - | - | - | - |
| B-cell lymphoma | 4533 | 10271778 | 9745557 | 94.88% | 83.89081988 | - | - | - | - |
| pancreatic cancer | 4443 | 9867903 | 9600833 | 97.29% | 123.0190387 | - | - | - | - |
| adenoma | 4173 | 8704878 | 8204709 | 94.25% | 95.12696881 | - | - | - | - |
| metastatic prostate cancer | 4135 | 8547045 | 8187151 | 95.79% | 108.8187944 | - | - | - | - |
| ovarian serous carcinoma | 4086 | 8345655 | 7903749 | 94.70% | 99.93915816 | - | - | - | - |
| T-cell acute lymphoblastic leukemia | 4070 | 8280415 | 7858536 | 94.91% | 104.5052505 | - | - | - | - |
| colon carcinoma | 4038 | 8150703 | 7856047 | 96.38% | 104.6948436 | - | - | - | - |
| leiomyosarcoma | 3945 | 7779540 | 7606116 | 97.77% | 149.6241285 | - | - | - | - |
| acute quadriplegic myopathy | 3886 | 7548555 | 7401357 | 98.05% | 185.2487793 | - | - | - | - |
| lung adenocarcinoma | 3737 | 6980716 | 6750072 | 96.70% | 94.47827873 | - | - | - | - |
| superficial transitional cell carcinoma with surrounding carcinoma in situ lesion | 3706 | 6865365 | 6717158 | 97.84% | 133.6403731 | - | - | - | - |
| amyotrophic lateral sclerosis | 3670 | 6732615 | 6637960 | 98.59% | 230.8792269 | - | - | - | - |
| chronic lymphocytic leukemia | 3663 | 6706953 | 6586303 | 98.20% | 128.2747887 | - | - | - | - |
| monoclonal gammopathy of uncertain significance | 3603 | 6489003 | 6123818 | 94.37% | 98.41798039 | - | - | - | - |
| anaplastic large cell lymphoma | 3463 | 5994453 | 5719267 | 95.41% | 101.6029226 | - | - | - | - |
| breast cancer | 3329 | 5539456 | 5367588 | 96.90% | 128.3381346 | - | - | - | - |
| adrenocortical carcinoma | 3237 | 5237466 | 5112620 | 97.62% | 131.1314338 | - | - | - | - |
| malignant melanoma | 3212 | 5156866 | 4972497 | 96.42% | 96.15725365 | - | - | - | - |
| B cell chronic lymphocytic leukemia | 3147 | 4950231 | 4789146 | 96.75% | 112.6814821 | - | - | - | - |
| osteoarthritis | 3135 | 4912545 | 4824516 | 98.21% | 185.6338859 | - | - | - | - |
| multiple myeloma | 3103 | 4812753 | 4624343 | 96.09% | 107.869172 | - | - | - | - |
| oligodendroglioma | 3073 | 4720128 | 4576952 | 96.97% | 160.4520384 | - | - | - | - |
| myeloma | 3069 | 4707846 | 4659438 | 98.97% | 215.255814 | - | - | - | - |
| carcinoma in situ lesion | 2987 | 4459591 | 4328834 | 97.07% | 134.2936189 | - | - | - | - |
| superficial transitional cell carcinoma without surrounding carcinoma in situ lesion | 2883 | 4154403 | 3883799 | 93.49% | 59.88577757 | - | - | - | - |
| **CONDITION** | **GENES** | **PAIRS** | **SHARING 7mer** | **SHARING 7mer %** | **AVG 7mer #** | **TRIPLETS** | **SHARING 7mer** | **SHARING 7mer %** | **AVG 7mer #** |
| experimental malaria-infected | 2692 | 3622086 | 3501772 | 96.68% | 110.8744867 | - | - | - | - |
| primary Sjogren syndrome | 2521 | 3176460 | 3130230 | 98.54% | 225.2827824 | - | - | - | - |
| sarcoma | 2425 | 2939100 | 2898962 | 98.63% | 163.110114 | - | - | - | - |
| dermatomyositis | 2423 | 2934253 | 2845859 | 96.99% | 106.7246601 | - | - | - | - |
| gastric cancer | 2374 | 2816751 | 2717710 | 96.48% | 125.2864698 | - | - | - | - |
| cervical cancer | 2368 | 2802528 | 2747167 | 98.02% | 145.2816985 | - | - | - | - |
| MDS-associated acute myeloid leukemia | 2323 | 2697003 | 2631139 | 97.56% | 120.8700639 | - | - | - | - |
| papillary thyroid carcinoma | 2238 | 2503203 | 2429582 | 97.06% | 121.7085174 | - | - | - | - |
| vulvar intraepithelial neoplasia | 2238 | 2503203 | 2439096 | 97.44% | 132.4364057 | - | - | - | - |
| nasopharyngeal carcinoma | 2236 | 2498730 | 2446952 | 97.93% | 147.4521517 | - | - | - | - |
| benign nevus | 2232 | 2489796 | 2428995 | 97.56% | 136.5208238 | - | - | - | - |
| renal cell carcinoma | 2158 | 2327403 | 2254714 | 96.88% | 124.6513788 | - | - | - | - |
| nodal peripheral T-cell lymphoma | 2111 | 2227105 | 2173757 | 97.60% | 134.8245318 | - | - | - | - |
| muscle invasive carcinoma | 2107 | 2218671 | 2093335 | 94.35% | 63.05102649 | - | - | - | - |
| juvenile idiopathic arthritis subtype systemic | 2082 | 2166321 | 2077267 | 95.89% | 110.1071263 | - | - | - | - |
| glioblastoma multiforme | 2063 | 2126953 | 2063838 | 97.03% | 136.2678418 | - | - | - | - |
| leiomyoma | 2019 | 2037171 | 1988937 | 97.63% | 122.8906758 | - | - | - | - |
| T-cell lymphoblastic lymphoma | 2016 | 2031120 | 2002605 | 98.60% | 150.3069395 | - | - | - | - |
| monoclonal gammopathy of undetermined significance (MGUS) | 1992 | 1983036 | 1903713 | 96.00% | 98.96232746 | - | - | - | - |
| chronic myelogenous leukemia | 1978 | 1955253 | 1923937 | 98.40% | 107.4380824 | - | - | - | - |
| chromophobe renal cell carcinoma | 1943 | 1886653 | 1830021 | 97.00% | 113.9408959 | - | - | - | - |
| Kaposi's sarcoma | 1932 | 1865346 | 1822497 | 97.70% | 155.3534358 | - | - | - | - |
| colorectal adenocarcinoma | 1831 | 1675365 | 1630893 | 97.35% | 93.0189493 | - | - | - | - |
| extranodal NK/T-cell lymphoma | 1746 | 1523385 | 1481708 | 97.26% | 123.9344151 | - | - | - | - |
| asymptomatic HIV-1 infection | 1737 | 1507716 | 1401661 | 92.97% | 75.39446885 | - | - | - | - |
| cancer | 1725 | 1486950 | 1445217 | 97.19% | 132.2933125 | - | - | - | - |
| peripheral T-cell lymphoma unspecified | 1714 | 1468041 | 1456676 | 99.23% | 220.9613703 | - | - | - | - |
| ulcerative colitis | 1694 | 1433971 | 1405688 | 98.03% | 137.3092845 | - | - | - | - |
| colon cancer | 1650 | 1360425 | 1311909 | 96.43% | 109.1793873 | - | - | - | - |
| **CONDITION** | **GENES** | **PAIRS** | **SHARING 7mer** | **SHARING 7mer %** | **AVG 7mer #** | **TRIPLETS** | **SHARING 7mer** | **SHARING 7mer %** | **AVG 7mer #** |
| synovial sarcoma | 1643 | 1348903 | 1319452 | 97.82% | 132.175877 | - | - | - | - |
| lung cancer | 1637 | 1339066 | 1316446 | 98.31% | 131.9556534 | - | - | - | - |
| obesity | 1577 | 1242676 | 1216678 | 97.91% | 138.8246438 | - | - | - | - |
| testicular seminoma | 1507 | 1134771 | 1083059 | 95.44% | 89.89054972 | - | - | - | - |
| astrocytoma | 1454 | 1056331 | 1026003 | 97.13% | 151.65717 | - | - | - | - |
| polyarticular juvenile idiopathic arthritis | 1443 | 1040403 | 1005212 | 96.62% | 93.05630126 | - | - | - | - |
| non-ischemic cardiomyopathy | 1406 | 987715 | 986358 | 99.86% | 228.2709911 | - | - | - | - |
| CLL | 1402 | 982101 | 967445 | 98.51% | 190.243534 | - | - | - | - |
| B-precursor acute lymphoblastic leukemia | 1377 | 947376 | 899173 | 94.91% | 107.1051536 | - | - | - | - |
| Hodgkins lymphoma | 1373 | 941878 | 916863 | 97.34% | 123.5695992 | - | - | - | - |
| acute lymphocytic leukemia | 1352 | 913276 | 896115 | 98.12% | 142.8513516 | - | - | - | - |
| angioimmunoblastic T-cell lymphoma | 1325 | 877150 | 863742 | 98.47% | 167.4694864 | - | - | - | - |
| acute malaria, natural infection Cameroon | 1313 | 861328 | 834256 | 96.86% | 98.77786859 | - | - | - | - |
| myxoid liposarcoma | 1304 | 849556 | 792108 | 93.24% | 75.59459059 | - | - | - | - |
| Mycobacterium tuberculosis infected | 1292 | 833986 | 812712 | 97.45% | 137.5496183 | - | - | - | - |
| malignant pleural mesothelioma | 1284 | 823686 | 805266 | 97.76% | 135.2700131 | - | - | - | - |
| Crohn's disease | 1208 | 729028 | 707316 | 97.02% | 115.7521782 | - | - | - | - |
| clear cell sarcoma of the kidney | 1196 | 714610 | 704751 | 98.62% | 162.3180756 | - | - | - | - |
| sporadic basal-like breast cancer | 970 | 469965 | 451502 | 96.07% | 106.6553403 | - | - | - | - |
| Burkitts lymphoma | 954 | 454581 | 442041 | 97.24% | 102.0810153 | - | - | - | - |
| treated malaria, natural infection Cameroon | 942 | 443211 | 423717 | 95.60% | 74.93753991 | - | - | - | - |
| familial hypercholesterolemia | 919 | 421821 | 403578 | 95.68% | 68.96949654 | - | - | - | - |
| FAB: M4 acute myeloid leukemia | 903 | 407253 | 391592 | 96.15% | 136.4891087 | - | - | - | - |
| childhood acute myeloid leukemia | 897 | 401856 | 391768 | 97.49% | 137.2343327 | - | - | - | - |
| head and neck squamous cell carcinoma | 856 | 365940 | 358257 | 97.90% | 103.009149 | - | - | - | - |
| plasma cell leukemia | 855 | 365085 | 350869 | 96.11% | 75.63576427 | - | - | - | - |
| Pick's disease | 783 | 306153 | 305011 | 99.63% | 339.7707747 | - | - | - | - |
| T cell acute lymphoblastic leukemia | 773 | 298378 | 289987 | 97.19% | 106.8400184 | - | - | - | - |
| adrenocortical adenoma | 735 | 269745 | 263791 | 97.79% | 182.0624479 | - | - | - | - |
| **CONDITION** | **GENES** | **PAIRS** | **SHARING 7mer** | **SHARING 7mer %** | **AVG 7mer #** | **TRIPLETS** | **SHARING 7mer** | **SHARING 7mer %** | **AVG 7mer #** |
| Waldenstrom's macroglobulinemia | 735 | 269745 | 259674 | 96.27% | 93.22384474 | - | - | - | - |
| systemic lupus erythematosus | 719 | 258121 | 253454 | 98.19% | 148.5012223 | - | - | - | - |
| AIDS | 694 | 240471 | 226705 | 94.28% | 94.60828957 | - | - | - | - |
| astrocytic tumor | 649 | 210276 | 207984 | 98.91% | 147.7867279 | - | - | - | - |
| bladder carcinoma | 636 | 201930 | 195674 | 96.90% | 80.79105135 | - | - | - | - |
| lymph node-negative breast carcinoma | 616 | 189420 | 184114 | 97.20% | 110.1350491 | - | - | - | - |
| gastric carcinoma | 615 | 188805 | 181475 | 96.12% | 73.34891025 | - | - | - | - |
| chronic obstructive pulmonary disease | 559 | 155961 | 154531 | 99.08% | 209.3933419 | - | - | - | - |
| neuroblastoma | 534 | 142311 | 140563 | 98.77% | 170.8778731 | - | - | - | - |

**Table S11. Additional In silico Analysis of Potential Seed Binding Sites.** (a) For each disease, the table reports the number of up-regulated genes, the total number of gene pairs, the number of pairs for which miR-Synth was able to design at least 1 amiR, the number of pairs for which miR-Synth was able to design at least 6 amiRs, the average number of amiRs designed per gene pairs and the average number of off-target genes. (b) For each disease, the table reports the average scores of the top 6 amiRs over all the analyzed gene pairs: seed, 3’ Match, AU content, Nucleotide Composition, Structural Accessibility, ARE/CPE, CTree Repression Score, M5P Repression Score. The last row reports the average scores of the top 6 amiRs designed for c-MET/EGFR.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **CONDITION** | **GENES** | **PAIRS** | **PAIRS W/ >= 1 miRNA** | **PAIRS W/ >= 6 miRNA** | **AVG # of miRNAs per pair** | **AVG # of Off-Target Genes** |
| Invasive Breast Cancer | 30 | 435 | 434 | 409 | 82.25806452 | 3980.755407 |
| Asthma | 40 | 780 | 637 | 396 | 27.99058085 | 4093.625545 |
| High-Stage Neuroblastoma | 51 | 1275 | 1231 | 987 | 47.55483347 | 3679.910158 |
| Squamos Cell Lung Carcinoma | 35 | 595 | 552 | 497 | 107.2699275 | 3603.680188 |
| Multiple Sclerosis | 115 | 6555 | 6403 | 6082 | 120.6250195 | 3762.465289 |
| Type 1 Diabetes | 47 | 1081 | 1069 | 1002 | 97.4686623 | 3633.405227 |
| Type 2 Diabetes | 74 | 2701 | 2569 | 2215 | 60.63020631 | 3771.705228 |
| Irritable Bowel Syndrome | 43 | 903 | 900 | 859 | 139.2566667 | 3798.238714 |

(a)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CONDITION** | **AVG SEED** | **AVG 3’ MATCH** | **AVG AU** | **AVG COMP** | **AVG ACCESS** | **AVG ARE/CPE** | **AVG CTREE** | **ACG M5P** |
| Invasive Breast Cancer | 0.922215101 | 0.533702048 | 0.661147301 | 0.685263726 | 0.303613011 | 0.385437578 | -0.221723785 | -0.233281505 |
| Asthma | 0.846072746 | 0.546901489 | 0.60038809 | 0.637929832 | 0.295351521 | 0.238617111 | -0.213010443 | -0.216689219 |
| High-Stage Neuroblastoma | 0.877086543 | 0.546050975 | 0.583099364 | 0.653802258 | 0.270270072 | 0.299313267 | -0.217683689 | -0.222985895 |
| Squamos Cell Lung Carcinoma | 0.902322392 | 0.547189182 | 0.60553261 | 0.674069524 | 0.276265963 | 0.374868168 | -0.220707322 | -0.228246128 |
| Multiple Sclerosis | 0.93565328 | 0.527883951 | 0.682220817 | 0.689364018 | 0.296370607 | 0.443902109 | -0.222929863 | -0.237074424 |
| Type 1 Diabetes | 0.91916729 | 0.537375975 | 0.637271842 | 0.685603715 | 0.287272585 | 0.405984194 | -0.222228681 | -0.232710954 |
| Type 2 Diabetes | 0.890827283 | 0.541157879 | 0.614856368 | 0.668287004 | 0.272537639 | 0.393227289 | -0.219119314 | -0.225360652 |
| Irritable Bowel Syndrome | 0.939662943 | 0.536906791 | 0.653253179 | 0.693272621 | 0.282969442 | 0.444709252 | -0.22337393 | -0.238224249 |
| c-MET/EGFR amiRs | 1 | 0.577380952 | 0.691666667 | 0.732600733 | 0.24245455 | 0.645833333 | -0.224881716 | -0.249371388 |

(b)