Supplemental Material to: Anastasis Oulas, Nestoras Karathanasis, Annita Louloupi, Ioannis Iliopoulos, Kriton Kalantidis and Panayiota Poirazi. A new microRNA target prediction tool identifies a novel interaction of a putative miRNA with CCND2. RNA Biology 2012; 9(9); DOI: 10.4161/rna.21725;
http://www.landesbioscience.com/journals/rnabiology/article/21725/

Supplementary Material

A new microRNA target prediction tool identifies a novel interaction of a putative miRNA with CCND2

Anastasis Oulas[1,2], Nestoras Karathanasis[1,3], Annita Louloupi[3] , Ioannis Iliopoulos[4] , Kriton Kalantidis[1,3], and Panayiota Poirazi[1]

[1] Institute of Molecular Biology and Biotechnology-FORTH, Heraklion, Crete, Greece
[2] Institute of Marine Biology and Genetics-HCMR, Heraklion, Crete, Greece
[3] Department of Biology, University of Crete, Heraklion, Crete, Greece
[4] Division of Medical Sciences, University of Crete, Heraklion, Crete, Greece

SUPPLEMENTARY METHODS

Filtering of target prediction data using 5 benchmark and mock miRNAs against all human 3'UTRs

Filtering algorithm:

IF

score ≥ score threshold (user defined) &

free energy ≤ free energy threshold &

conservation ≥ conservation threshold

THEN

predicted candidate surpasses filters

score >=5	Energy < -8.0
3 < score <=4	Energy <= -12
2 < score <=3	Energy <= -14
1 < score <=2	Energy <= -16
0 < score <=1	Energy <= -18

Table S1. Filtering by score and free energy. Monotonic relationship between score and energy threshold derived form experimentally verified miRNA targets from Tarbase version 5

The rules used to filter our data are a combination of the HMM score, free energy as predicted by RNAcofold and conservation score in the seed region (see Materials and Methods). These features allow for the distinction of the weaker binding targets such as the 3′-compensatory category of miRNA target sites. The HMM score gives an indication of the seed match, whereby the higher the score the greater the seed match. Hence, candidates that display hybrid structures with strong seed binding will receive high HMM score (>5). If a candidate target site has a low HMM score (i.e. 3), we expect that base pairing at the 3' region will compensate for the lack of matches at the seed region and hence will have a low free energy value. These rules are based on a monotonic combination of the HMM score versus the minimum free energy and the conservation score. As the score decreases (more mismatches at seed region) the energy decreases as well, hence ensuring that candidate target sites with imperfect seed matches will only be classified as correct target sites if the ΔG is low enough to indicate 3'compensatory binding. Moreover, the filtering criteria are used in combination and hence no one filtering parameter is used to provide a cut-off decision; rather a unanimous decision is obtained after applying all three filtering parameters synergistically. The distributions of targets lying within and without the filtering boundaries are shown in Figure 2. As shown a high number of predicted targets exceed the thresholds and hence are filtered out. Importantly all of the experimentally verified miRNA targets from Tarbase are within the thresholds defined in Figure S1.

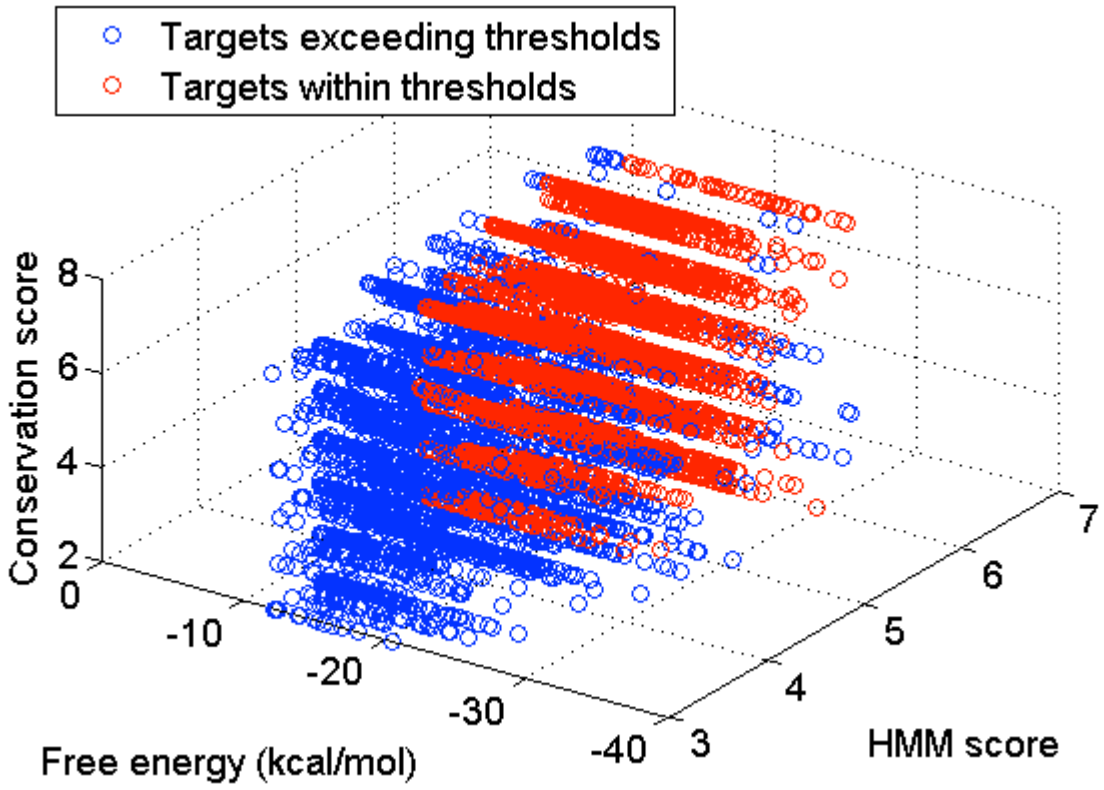


Figure S1. **Filtering targets using different parameters**. 3D plot of targets after using filtering parameters using predicted targets from 5 benchmark miRNA and all 3UTRs in the human genome. Conservation score >= 2, HMM score >= 3 and free energy < -8.0 (as predicted by RNAcofold). All experientially supported targets which exceed the conservation threshold fall within the red area of the plot which represents the targets retained after filtering. The blue circles represent miRNA targets which do not surpass the filtering thresholds and hence are not considered as true miRNA-targets All experimentally verified miRNA targets obtained from Tarbase and tested here are represented as red circles.

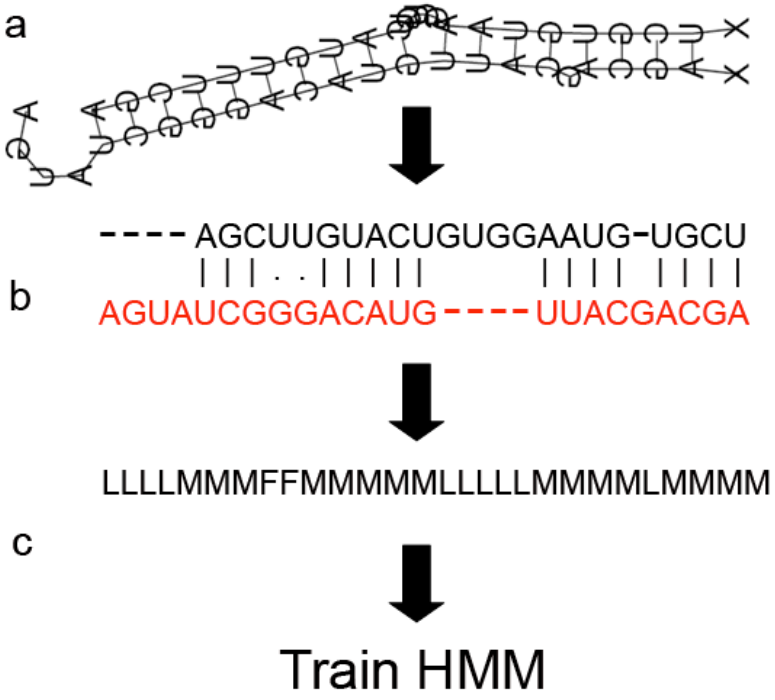


Figure S2. **RNAcofold to HMM**. (a) Output from RNAcofold. (b) Text-like representation of the output from (a). (c) String-like conversion of the RNAcofold output using Ls (loops), Ms (matches) and Fs (GU wobbles). String-like structures are aligned in order to construct a multiple sequence alignment, utilized to train an HMM model.

Profile HMMs

The HMMER (Eddy SR, 1998) software package was used to build a HMM capable of predicting RNA or DNA Profiles. HMMs are generative probabilistic models which are frequently used to address serious theoretical problems. For correct statistical inference, it is necessary to be able to calculate a probability distribution P(S|M) for the probability of sequences S given a modelM, and have this quantity sum to one over the 'space' of all sequences. Generative models work by recursive enumeration of possible sequences from a finite set of rules—rules that in an HMM are represented by states, state transitions and symbol emission probabilities. HMMER uses a Profile HMM architecture called Plan 7 which is illustrated in Figure S3. Profile HMMs are statistical models of multiple sequence alignments. They capture position-specific information about how conserved each column of the alignment is, and which residues are most likely.

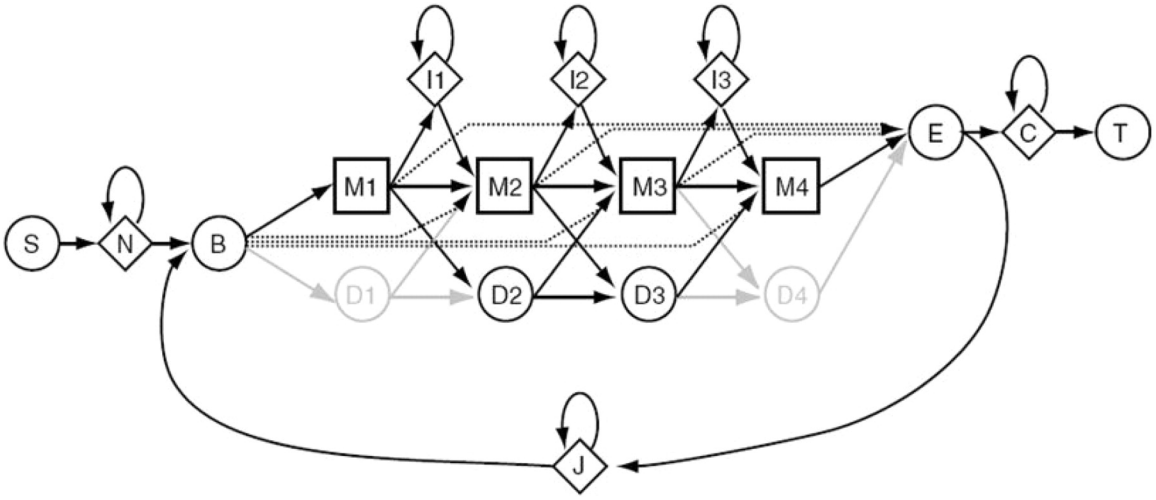


Figure S3. HMMER Plan 7 architecture. Squares indicate match states (modeling consensus positions in the alignment). Diamonds indicate insert states (modeling insertions relative to consensus) and special random sequence emitting states. Circles indicate delete states (modeling deletions relative to consensus) and special begin/end states. Arrows indicate state transitions. Figure was adopted from Eddy SR, 1998.

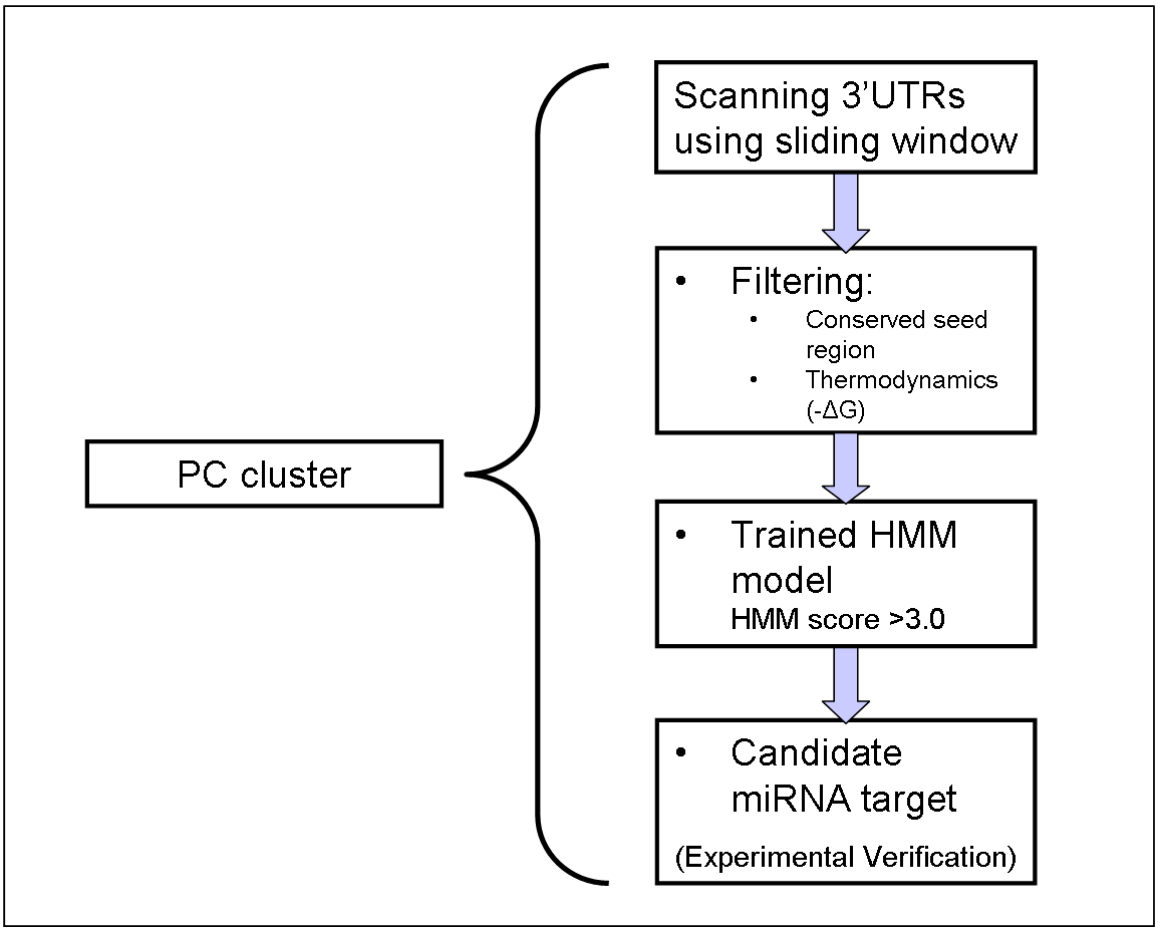


Figure S4. Targetprofiler flowchart. A sliding window of 32nt scans full 3'UTR sequences from the human genome shifting 1nt at time. Specific filtering parameters (see main text) are then enforced to discard sequences that do not meet the algorithms filtering criteria. The trained HMM model then assigns a likelihood score to the target site underhand providing candidate miRNA targets for experimental verification.

a

Wild type interaction:

```
5′ AGTGGGGGCCGAGTTGTTCCCCCAGCCTGCCAA  CCND2
                              |||||||
             3′   ATGGAGAGGGGGACGGTC  c-miR-Ch9
```

Wild type: GGGGGCCGAGTTGTTCCCCCAGCCTGCCAA

Mutated interaction:

```
5′ AGTGGGGGCCGAGTTGTTCCCCCAGCGAGGGAA  CCND2
                              |   |   |
             3′   ATGGAGAGGGGGACGGTC  c-miR-Ch9
```

Mutant: GGGCCGAGTTGTTCCCCCAGCGAGGGAA

b

P1-FORWARD: TCTAGAATGGTGACTGACCCTTGAGC
P2-REVERSE: TCTAGAATCCTGCTAGCAATGGGATG
Mutagenesis Left : CAGCGAGGGAAATTTTGATCCTTCCCCTCTTT
Mutagenesis Right: ATTTCCCTCGCTGGGGGAACAACTCG

c

5-TACCTCTCCCCCTGCCAG-3

Figure S5. **Cloned sequences and primers. (a)** Mutant seed regions designed for pGL4-10 + mut – Triplet and pGL4-10 + mut-3'UTR. **(b)** Primers for CCND2 3'UTR amplification and PCR mutagenesis for pGL4-10 + mut-3'UTR. **(c)** LNA sequence

SUPPLEMENTARY RESULTS

Target Type	HMM Score	Location (chr:start-end)	Sequence (targetX&XmiRNA)	Bracket notation	Conservation score	Strand
7mer-m8	6.2	10:60193696019401	GAAGAGGACACCAGCCCAAGCUGGACC UGCCAUX&XCUGGCAGGGGGAGAGGUA	.(((....((...(((((((..&..)))))))...)).))).	7	1
8mer	6.2	11:26973747-26973779	GAGGUUCAAGGUGCUGCUUUGCAUGCC UGCCAAX&XCUGGCAGGGGGAGAGGUA	.....((((((...(.(((((((. .&..)))))))).)..))))))	7	1
7mer-m8	6.2	1:156619755-156619787	AACCUGUCAGCUUGCACCAUCCCCACCU GCCACX&XCUGGCAGGGGGAGAGGUA	.....(((.(((((.(((((((. .&..)))))))))).)))))).	7	1
8mer	6.2	11:72794632-72794664	CUCAAAAGGUGAUUUUGUCCUUAGACCU GCCAAX&XCUGGCAGGGGGAGAGGUA	.....(((.((((....(((((((.. &..))))))))))).))))..	7	-1
7mer-m8	6.2	12:3771033-3771065	CCCGCUGUUAAACUGCAUAGGGCAGCC UGCCACX&XCUGGCAGGGGGAGAGGUA	.....(((((((..&..)))))))..	7	-1
8mer	6.2	12:4282963-4282995	AGUGGGGGCCGAGUUGUUCCCCCAGCC UGCCAAX&XCUGGCAGGGGGAGAGGUA	(((((..(((((((..&..)))))))).)) .).))..	7	1
7mer-m8	6.2	1:35849529-35849561	GUAGGAGGUUUAGUGGCUGCUCUGGCC UGCCAUX&XCUGGCAGGGGGAGAGGUA	.....(((.((((..(((((((. .&..)))))))).)))).))).	7	-1
8mer	6.2	14:62243818-62243850	UAUUGCAUGUCCAGCUGGAUUCUGGCC UGCCAAX&XCUGGCAGGGGGAGAGGUA	.....(((...(((..(((((((.. &..)))))))))))..))).	7	-1
7mer-m8	6.2	16:70239699-70239731	CAGCUGUUCUGUAUCAGUCCUACCACC UGCCAUX&XCUGGCAGGGGGAGAGGUA	.....(((.((.(((((((..& ..)))))))))...)))..	7	-1
8mer	6.2	17:44836546-44836578	UUGCAUCCUGCUGGGGCUGAACAUGCC UGCCAAX&XCUGGCAGGGGGAGAGGUA	.....(((.....(.(((((((..&..)))))))..)....))).	7	-1
8mer	6.2	19:44061675-44061707	AGCUUCCCCCAAGAAGUCCCCGCCACCU GCCAAX&XCUGGCAGGGGGAGAGGUA	.....(((.((.(((((((.. &..)))))))).)).).)))..	7	-1
8mer	6.2	19:45595285-45595317	AGGCAGCUGGUGGCUUUGCCCUCCACC UGCCAAX&XCUGGCAGGGGGAGAGGUA	.....(((((((.(((((((..&..)))))))).))))).))))	7	-1
8mer	6.2	2:166362026-166362058	UGCCUACCUGUCAAACUGUGUGAAACCU GCCAAX&XCUGGCAGGGGGAGAGGUA	......(((........(((((((..&..)))))))......))).	7	1
7mer-m8	6.2	2:48517574-48517606	UGAAUUCGAGUAUUUUAAUGUUAUACCU GCCAUX&XCUGGCAGGGGGAGAGGUA	.....((((((((.. &..))))))))..... .	7	1
7mer-m8	6.2	3:31653720-31653752	UAAAAGUGAAAGAGAAAGGGUUUUUCCU GCCACX&XCUGGCAGGGGGAGAGGUA	.....(((..(((((((.. &..)))))))..))).	7	1
8mer	6.2	4:114731538-114731570	CAGUAAAUAUAUUGAGCCAUGUUAACCU GCCAAX&XCUGGCAGGGGGAGAGGUA	.....(((.(.((..(((((((. .&..)))))))..)).)))).	7	-1
8mer	6.2	5:14561539-14561571	AGAAGUUCUUUCUCAUUCUCUUUCACCU GCCAAX&XCUGGCAGGGGGAGAGGUA	.....(((((..(.(((((((. .&..)))))))).))))) ..	7	1
8mer	6.2	5:71536909-71536941	AUCUAGUUAAGUCGCUGAACAAUUACCU GCCAAX&XCUGGCAGGGGGAGAGGUA	.....(((.....(((((((..&. .)))))))......))).	7	1
8mer	6.2	8:92477834-92477866	GCAUCUAUAAAAGUAAAUUCUAGUGCCU GCCAAX&XCUGGCAGGGGGAGAGGUA	.....(((((..(.(((((((.. &..)))))))).).)))). ..	7	1
7mer-m8	6.2	X:40964086-40964118	UUCAUCUACUUAGACUUUUUAAAUGCCU GCCAUX&XCUGGCAGGGGGAGAGGUA	.....(((((((...(.(((((((. .&..)))))))).))))))).	7	1
7mer-m8	6.2	X:44727602-44727634	AAUGCUGUUAUUUUUUCCAGAUUUACCU GCCAUX&XCUGGCAGGGGGAGAGGUA	.....((((((((((.....(((((((. .&..)))))))))))))))))	7	1
8mer	6.2	X:48787540-48787572	UUAUUGGGAGACUUUUGUCUUCCAGCC UGCCAAX&XCUGGCAGGGGGAGAGGUA	.....(.((((((..(((((((..&..)))))))))))))).) ..	7	1
7mer-m8	6.7	1:157210240-157210272	GGCUGGGGAGUGUUUAUUUUAAGAUCC UGCCAUX&XCUGGCAGGGGGAGAGGUA	.....((((((....(((((((. .&..)))))))...))))))	7	1
7mer-m8	6.7	1:157466940-157466972	UGGACUGUGCCUAUGGAUUUGGAUUCC UGCCAUX&XCUGGCAGGGGGAGAGGUA	.....(((((.....(((((((((. .&..)))))))))...)))))	7	-1
8mer	6.7	1:208661990-208662022	AACAGUAACGAGUAGCCAGAGUACUCCU GCCAAX&XCUGGCAGGGGGAGAGGUA	.....(((.....(((((((((.. &..)))))))))....))).	7	1
8mer	6.7	16:28755701-28755733	CGUUCCCCAGGGGAGCUGGGGAAUUCC UGCCAAX&XCUGGCAGGGGGAGAGGUA	.....(((.....(((((((((..&..))))))))).....))).	7	1
8mer	6.7	17:24980601-24980633	AAAUGGAGACUUCCAAUCACCAGCUCCU GCCAAX&XCUGGCAGGGGGAGAGGUA	.....(((..(((((((((..&. .)))))))).....))).	7	-1

7mer-m8	6.7	19:55989973-55990005	GCCCGGCCUCCCGCCCAUGGGGUCUCCUGCCAUX&XCUGGCAGGGGGAGAGGUA	.....((((((((((..&..))))))))))....	7	1
7mer-m8	6.7	20:47955090-47955122	CCUGUUGGCUUGUGAAAUGAGCCCUCCUGCCACX&XCUGGCAGGGGGAGAGGUA	.....((((((((((..&..))))))))).....	7	-1
7mer-m8	6.7	21:15346087-15346119	GAAACCCAUUUAACUGUCACACACUCCUGCCACX&XCUGGCAGGGGGAGAGGUA	.....((.((.....((((((((((..&..)))))))))).))...	7	1
7mer-m8	6.7	5:172496015-172496047	AAAAAAAAUUGCAUUUUAUAUGAUUCCUGCCAUX&XCUGGCAGGGGGAGAGGUA	.....((((((((((..&..))))))))).....	7	1
8mer	6.7	5:65902001-65902033	UGAAUUUCUACGGAGCUUGAUGAUUCCUGCCAAX&XCUGGCAGGGGGAGAGGUA	.....((((.....((((((((((..&..)))))))))...)))).	7	1
7mer-m8	6.7	8:74868493-74868525	GUAACAGGAAAAGUUUUCAUUAACUCCUGCCAUX&XCUGGCAGGGGGAGAGGUA	.....((((((((((..&..))))))))).....	7	-1

Table S2 . c-miRCh9 predicted targets. The table shows details for the total number of predicted targets for c-miRCh9. The target site on CCND2 is highlighted in grey.

(a)	Targetprofiler	TargetScan 4.2	Diana-MicroT 3.0	PicTar
Targetprofiler	3667	1630	1803	938
TargetScan 4.2		3732	2475	1274
Diana-MicroT 3.0			5770	1421
PicTar				2923
(b)				
	Targetprofiler	TargetScan 4.2	Diana-MicroT 3.0	PicTar
Targetprofiler	100%	44.45%	49.17%	25.58%
TargetScan 4.2	43.68%	100%	66.32%	34.14%
Diana-MicroT 3.0	31.25%	42.89%	100%	24.63%
PicTar	32.09%	43.59%	48.61%	100%

Table S3 Common targets between prediction tools. The table shows a pairwise comparison of the number of common miRNA gene targets predicted by each pair of tools for the 5 benchmark miRNAs (a). The overall fraction of miRNA targets predicted in common by each pair of tools is also shown as a percentage in (b).

Primer extension methodology for experimental identification of the mature miRNA sequence for a novel miRNA candidate

We devised a novel methodology based on primer extension for identifying the most probable miRNA mature sequence from a putative precursor. Specifically, instead of using one primer complement to the mature sequence, which in our case was unknown, we designed three different overlapping primers that are complementary to the positive strand of the precursor sequence, namely the strand producing a small RNA (see Figure S6a). Using these primers we performed three primer extension reactions. The length of the extended primers further defines the location of the mature sequence on the precursor. The primer may bind to the precursor and/or the mature sequence. We assume that binding of the primer to both the mature as well as the precursor sequence will result in competition of binding and this will be evident in the banding patterns resulting from the primer extension reactions.

More specifically, for the first primer extension reaction the longest expected product is 19nt and the results show a band in the vicinity of 19nts (Figure S6b column 1). This product could be the result of either the extension of the precursor or the extension of the mature. Hence, the exact binding remains to be verified by investigating the products of the other 2 primer extension reactions. If this band is a direct result of the extension of the mature, then the other 2 primers extension reaction would be expected to produce sequences of length more than 31nts, resulting from the binding and extension of the primer to the precursor sequence. Clearly this is not the case.

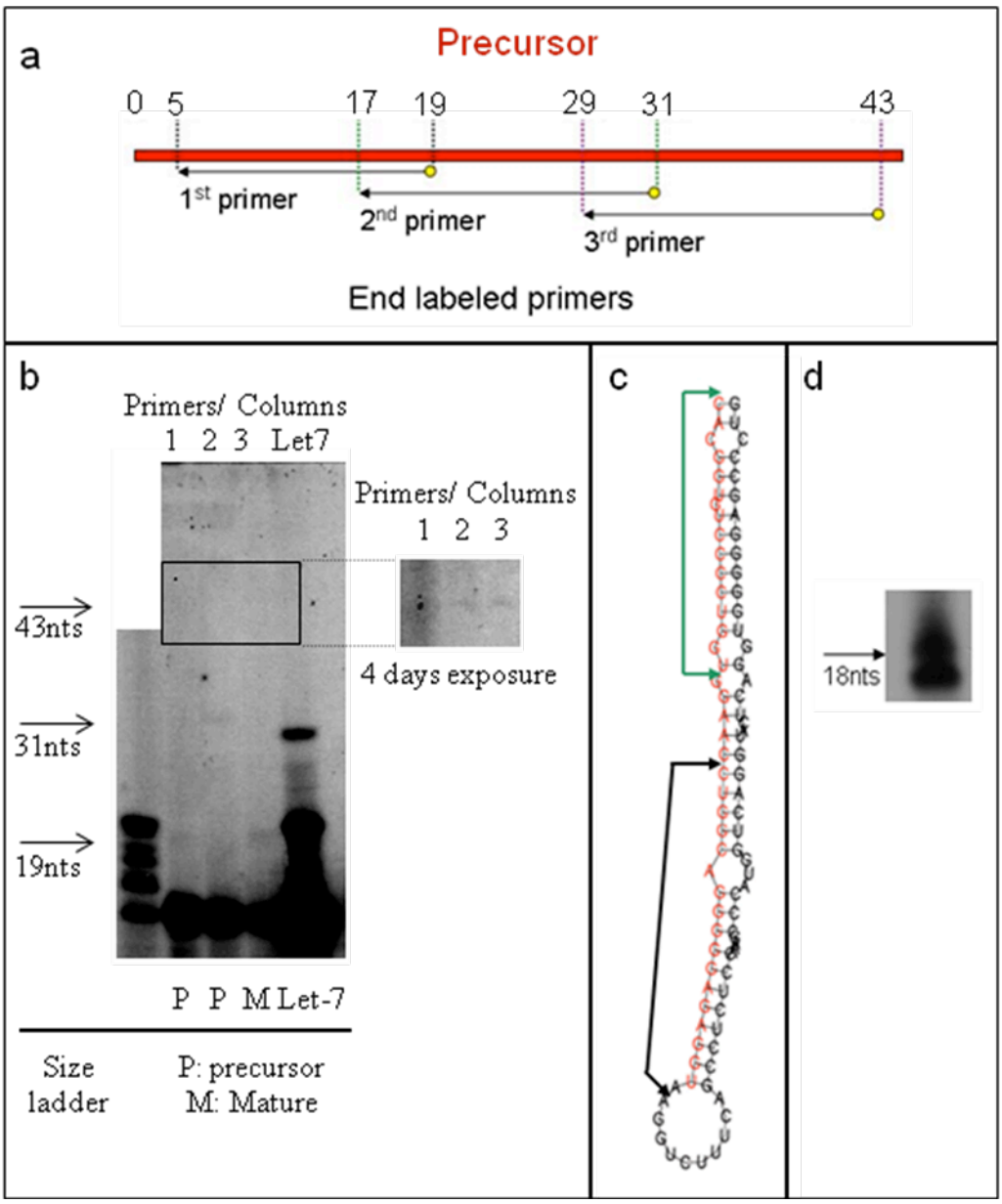
The second primer extension reaction produces one band (Figure S6b column 2) which is potentially derived due to the extension of the precursor. For this primer extension reaction the longest expected product through precursor binding and elongation is 31nts and the observed band corresponds to this length.

The observed banding patterns of the third reaction (Figure S6b column 3) reveal two products, one at 43nts and one at 19nts. We believe that the long band, which could be visualized only after 4 days of exposure, is due to precursor elongation since the longest expected product through precursor binding and extension would be 43nts. We consider the short band, which was detectable after overnight exposure, as a result of the extension of the mature sequence and we use this as our reference point for the prediction of the c-miR-ch9 mature sequence.

Based on that, we determine that the 5' start of the mature sequence is at the 26th nucleotide of the precursor's sequence. Having identified the 5' end of our mature sequence, and knowing that our mature sequence is 18nts long, we are able to deduce the 3' end of the mature sequence (Figure S6c). According to the results from our primer extension methodology we predict that the mature sequence for the potential miRNA (c-miR-ch9) is 5' CUGGCAGGGGGAGAGGUA.

Due to the novelty of the methodology (primer extension has never been used before for the prediction of the miRNA mature sequence) and the fact that some of the bands were faint we decided to verify our prediction with a northern blot analysis.

We carry out a northern blot analyses using a DNA probe complement to the mature sequence as predicted by our primer extension reaction (black arrows in Figure S6c) and we also use a negative control, a DNA probe complement to the adjacent sequence (green arrows in Figure S6c, data not shown). Additionally to increase sensitivity and signal of the experiment we also used an LNA probe complementary to the mature sequence (Figure S6d).



FigureS6. MiRNA mature prediction methodology

a) Three primers were designed to bind to the whole length of the verified positive strand of the hypothetical miRNA (Oulas et al, NAR, 2009) c-miR-ch9. b) MiRNA mature prediction methodology results. Each column displays the banding patterns from the corresponding primers shown in (a). The banding patterns of the first and the second columns result from the annealing of the first and the second primers to the precursor (denoted by P) sequence. While the banding pattern on the third column is a consequence of the annealing of the third primer to the precursor, 43nts band, and the mature (M) sequence, 19nts band.

c) Predicted potential mature miRNA sequence. Based on the results in (b) and the fact that the expected mature is 18nts long (Oulas et al, NAR, 2009) we conclude that the mature sequence is between the 24th and the 41stnt on the verified positive (5p) strand.

The black arrows indicate the predicted mature miRNA sequence. The green arrows indicate the sequence which was used as a negative control for the northern blot.
d) Northern blot analysis. In order to verify the predicted mature miRNA sequence we perform a northern blot analysis using a DNA probe complement to the predicted mature, (black arrows in c), and as a negative control a DNA probe complement to the adjacent sequence, (green arrows in c). In order to increase our experiments sensitivity and signal an LNA probe complement to the predicted mature was also used. The upper observed band is ~18 nts long as expected from previous published data (Oulas et al, NAR, 2009).

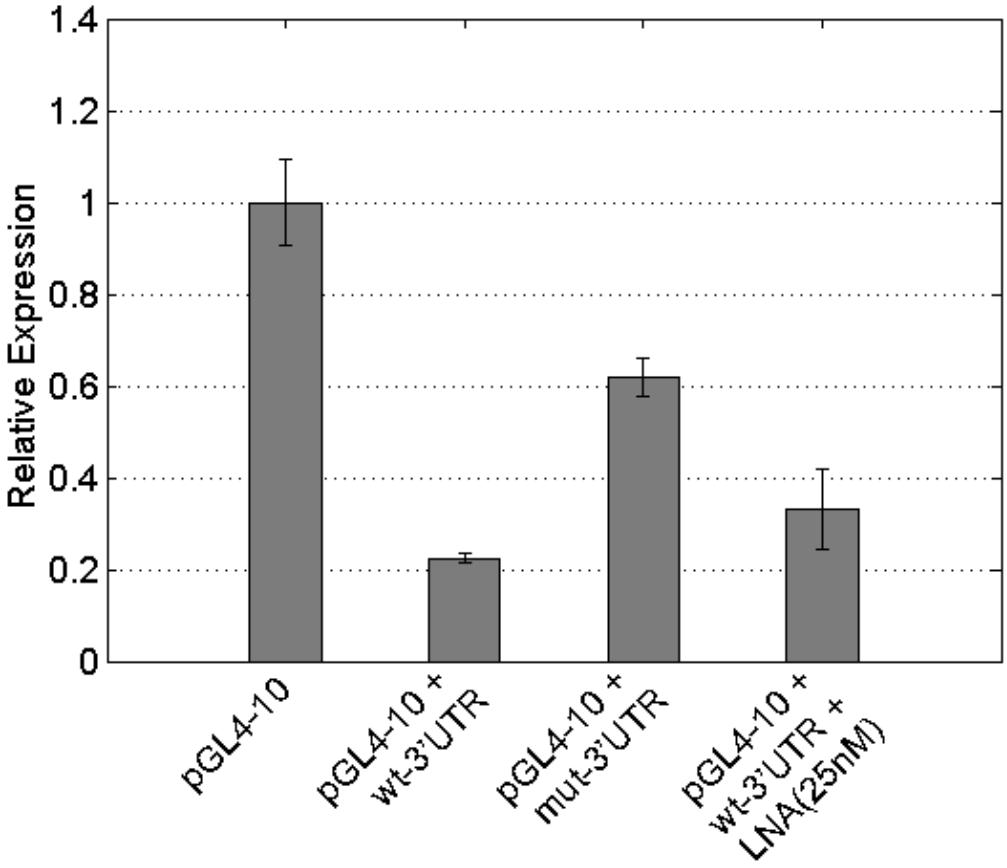


Figure S7. miRNA-sensor assay using luciferase expression as an indicator of miRNA activity after transfection of HeLa cells with various constructs.
Relative luciferase expression after transfection of HeLa cells with 3'UTR constructs: pGL4-10 – an empty pGL4-10 vector for standardization control, pGL4-10 + wt-3'UTR – vector containing a wild-type 3'UTR containing a single potential binding site for c-miR-Ch9, pGL4-10 + mut-3'UTR – a vector containing a single mutated potential binding site for c-miR-Ch9 and pGL4-10 + wt-3'UTR + LNA(25nM) - pGL4-10 + wt-3'UTR transfection was repeated with concurrent addition of anti-LNA for our c-miR-Ch9 (3 triplicates where performed for every condition).