

# miRTar2GO: a novel rule-based model learning method for cell line specific microRNA target prediction that integrates Ago2 CLIP-Seq and validated microRNA–target interaction data

Alireza Ahadi<sup>1,2</sup>, Gaurav Sablok<sup>3,\*</sup> and Gyorgy Hutvagner<sup>2,\*</sup>

<sup>1</sup>Faculty of Engineering and Information Technology, School of Software, University of Technology Sydney, PO Box 123, Broadway, Sydney, NSW 2007, Australia, <sup>2</sup>Faculty of Engineering and Information Technology, Centre of Health Technologies, University of Technology Sydney, PO Box 123, Broadway, Sydney, NSW 2007, Australia and <sup>3</sup>Plant Functional Biology and Climate Change Cluster (C3), University of Technology Sydney, PO Box 123, Broadway, Sydney, NSW 2007, Australia

Received June 15, 2016; Revised November 13, 2016; Editorial Decision November 15, 2016; Accepted November 16, 2016

## ABSTRACT

**MicroRNAs (miRNAs) are ~19–22 nucleotides (nt) long regulatory RNAs that regulate gene expression by recognizing and binding to complementary sequences on mRNAs. The key step in revealing the function of a miRNA, is the identification of miRNA target genes. Recent biochemical advances including PAR-CLIP and HITS-CLIP allow for improved miRNA target predictions and are widely used to validate miRNA targets. Here, we present miRTar2GO, which is a model, trained on the common rules of miRNA–target interactions, Argonaute (Ago) CLIP-Seq data and experimentally validated miRNA target interactions. miRTar2GO is designed to predict miRNA target sites using more relaxed miRNA–target binding characteristics. More importantly, miRTar2GO allows for the prediction of cell-type specific miRNA targets. We have evaluated miRTar2GO against other widely used miRNA target prediction algorithms and demonstrated that miRTar2GO produced significantly higher  $F_1$  and  $G$  scores. Target predictions, binding specifications, results of the pathway analysis and gene ontology enrichment of miRNA targets are freely available at <http://www.mirtar2go.org>.**

## INTRODUCTION

MicroRNAs (miRNAs) are small non-coding RNAs (ncRNA) with lengths ranging between 19 and 22 nucleotides. They play an important role as post-

transcriptional regulators of gene expression (1). The most recent estimates suggest that approximately 60% of the mRNA repertoire are under the post-transcriptional control of miRNAs (2), and they play fundamental roles in the regulation of most biological processes including diseases such as cancer. In animals, mature miRNAs are incorporated into one member of the Argonaute (Ago) protein family of the RNA induced silencing complex (RISC) (3–7). RISC typically targets the 3' untranslated region (3'UTR) of the targeted messenger RNA (tmRNA) (8) leading to the inhibition of the translation of the corresponding mRNAs via various mechanisms (9–11). Binding site interactions of the miRNA–tmRNA depend on sequence complementarity; most importantly, on the short sequence homology between the miRNAs seed sequence (the second to seventh nucleotides of the miRNAs) and the targeted mRNA (12). Based on seed complementarity between miRNAs and tmRNAs, several computational methods have been developed to predict miRNA targets (12–18).

Base-pairing between the miRNA and its target is the most commonly used feature in miRNA target prediction tools (19,20). The majority of the prediction algorithms require, but are not necessarily limited to, the seed match between the miRNA and the tmRNA. Most miRNA target prediction tools use the extended seed match (complementary between the second and the eighth to ninth nts of the miRNAs and the corresponding tmRNAs) criterion. However, it has been shown that the majority of functional target sites are governed by less specific seed matches with a length of only six nucleotides (21). It was also demonstrated that narrowing the length of the seed match to six nucleotides increases the number of correct predictions for miRNA targets. However, this has also increased the number of in-

\*To whom correspondence should be addressed. Email: gyorgy.hutvagner@uts.edu.au  
Correspondence may also be addressed to Gaurav Sablok. Email: sablok@gmail.com

correctly identified targets as such short motifs occur frequently in the transcriptome and could produce high false positive ratio (FPR) (22,23). Additional factors such as target site accessibility (14) and evolutionary conservation of the binding site (24) have also been incorporated into prediction tools to reduce the high FPRs. However, these factors are context dependent and their contribution to define a functional miRNA binding site varies between species, tissues/cell types, developmental stages, and can also be modulated by physiological stress (25).

Cross-linking immunoprecipitation (CLIP) using Ago2 specific antibodies has been used to experimentally identify the Ago2 bound transcriptome, including transcripts possibly targeted by miRNAs (26,27). Here, we present miRTar2GO, which integrates information from Ago2 CLIP-Seq and experimentally verified miRNA–tmRNA interactions. MiRTar2GO uses a rule based learning approach to predict cell type specific miRNA targets. The core algorithm uses Ago2 CLIP-Seq data to identify short (6 nt) perfect seed matches between the 3' UTRs of mRNAs and miRNA seed regions and assigns a score to each miRNA–mRNA pair using two steps: first, it calculates the hybridization energy between the miRNA and its candidate binding sites. Second, it compares each predicted target site to the characteristics of all validated target sites derived from luciferase assays, expression profiling and cross-linking ligation and sequencing of hybrids (28) (CLASH) experiments of the given miRNA in order to rank the predictions. miRTar2GO further improves these prediction using Ago2 footprints shared between different cell types to identify common and cell specific miRNA–tmRNA interactions. The free online portal of miRTar2GO also provides information from external databases including functional annotation from KEGG (29) and hiPathDB (<http://hipathdb.kobic.re.kr>) (30). The current version of miRTar2GO allows the user to filter miRNA targets based on a probability score and also explore the target genes by performing functional enrichments of biological ontologies.

To evaluate our prediction model, we have compared the result of miRTar2GO to several other widely used miRNA target prediction algorithms. We demonstrated that miRTar2GO possess higher sensitivity than the enlisted miRNA target predictions by re-analyzing pSILAC (31). Additionally, using the interaction result of 3' LIFE (32) as well as the large scale experimental CLIP data (33) we have shown the highly predictive strength of miRTar2GO.

## MATERIALS AND METHODS

### Data preparation and the process of target recognition

MicroRNA target prediction is largely dependent on seed matching. A single miRNA can bind and inhibit hundreds of targets in one cell type and might target only a few number of transcripts in another cell type—e.g. cell specific targeting (34). miRTar2GO aims to address cell specific miRNA target prediction. To collect Ago2–mRNA interactions in different cell lines, we downloaded six publicly available Ago2 CLIP-Seq datasets from StarBase version 2 (35,36). This includes 273 934 Ago2 footprints from six human cell lines: HeLa (37) (169 346 CLIPed sites), lymphoma

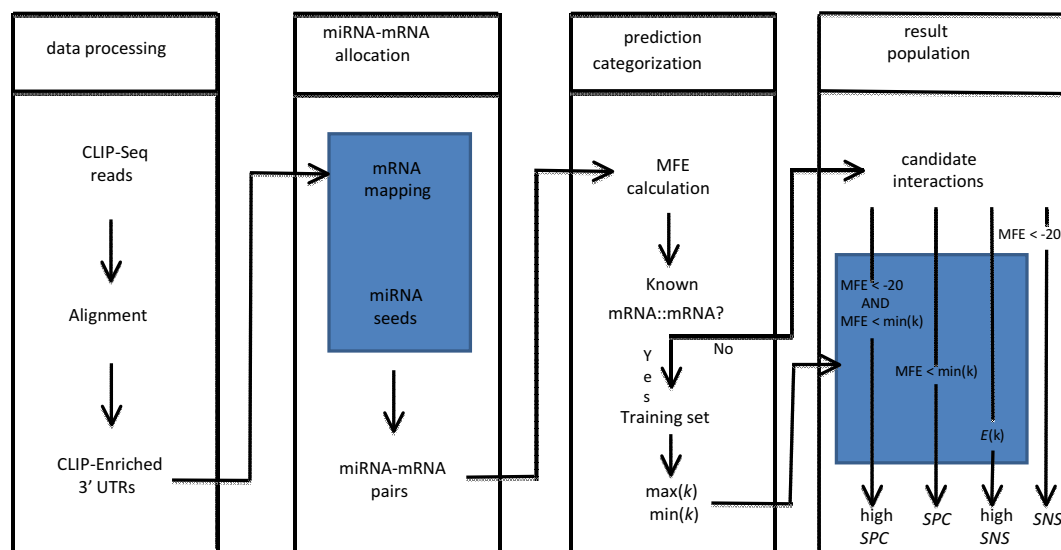
cell line (BC-1) (38) (43 997 CLIPed sites), lymphoblastoid cell line (39) (24 608 CLIPed sites), Human Embryonic Stem Cells (40) (9169 CLIPed sites), HEK 293 cell line (41) (24 041 CLIPed sites), and human lymphoblastic cell line (42) (2773 CLIPed sites). PAR-CLIP and HITS-CLIP have only small accuracy differences in identifying binding sites of Ago2 (43). These collected CLIPed sites were then aligned to genome reference consortium GRCh37. Intronic sequences (~14%) as well as those exons which mapped to the 5' UTR (~5% of all exons) or coding regions (~48% of all exons) of mRNAs were discarded. All chromosome conservation score files were downloaded from the UCSC genome browser website, which are based on PhastCons's (44) multiple alignment of 100 vertebrates to the human genome (hg19). Information on the conservation of miRNAs was obtained from TargetScan (45). All non-RefSeq accession IDs were mapped to RefSeq transcripts by applying mapping tables provided by either Ensembl or UCSC. The miRNA sequences were downloaded from miRBase release 20 (46), and mRNA sequences were downloaded from UCSC ftp website at <http://hgdownload.cse.ucsc.edu/goldenPath/hg19/>.

### MicroRNA–target mRNA allocation

Enrichment analysis of all possible 7-mers within the Ago2 cross-link-centered regions (CCRs) identified by PAR-CLIP suggests that the most significantly enriched 7-mers correspond to the reverse complementary strand of the seed region of abundant miRNAs (26). These enriched motifs are frequently positioned one to two nts downstream of the predominant cross-linked site within the CCRs where a window size of 41 nt centered on the predominant cross-linked position includes the bona fide miRNA binding sites. The specificity of the long seed matches (length 7 and 8) has been previously shown in CLIP based studies of miRNA targetome (21). Since the majority of the functional target sites are formed by 6 nt long seed matches (21), we focused on seed matches with a similar criterion. We aligned the 6-mer seed regions of all human miRNAs to the Ago2 CLIPed sites on the 3' UTR of mRNAs to identify all potential binding sites for each miRNA with a perfect 6-mer seed match. No mismatch or gap was allowed in the seed match alignment as only a small fraction of the miRNA target sites contain bulge or loops in the seed region (<6.6% in HEK293 based on PAR-CLIP data (26) and <15% in mouse brain HITS-CLIP data (47)). Similarly, a recent study demonstrates that miRNA–tmRNA interactions, which are governed by non-canonical seed matches, do not mediate repression and thus are not functional (48). Although, G:U wobbles were allowed as it is shown that the seed region of many heteroduplexes comprises G:U base-pairings (49,50). Applying these filtering criteria resulted in the set of miRNA–tmRNAs pairs in which the miRNA of each pair has a perfect seed region match in the Ago2 interacting region of the 3' UTR of the associated mRNA.

### Categorization of the predictions

We calculated minimum free energy (MFE) of hybridization to identify favorable miRNA–Ago2 CLIPed sites. MFE has



**Figure 1.** Pipeline of miRTar2GO. In the data processing step, the genomic coordinates of the Ago2 CLIP-Seq reads in different cell lines are mapped to the mRNAs to identify 3' UTRs which are enriched in Ago2 interaction sites. In the miRNA-mRNA allocation step, the 6mer seed region of all known miRNAs are aligned to the reverse complementary sequence of Ago2 CLIPed sites in the 3' UTRs. At the prediction categorization step, the whole miRNA sequence of each miRNA-mRNA pair is folded upon its allocated Ago2 CLIPed sequence to calculate hybridization energy. At this step, interactions with a MFE  $> -15$  kcal/mol are discarded. At the result population step, four different sets of results are generated. The interactions in the unseen set with a MFE  $< -20$  kcal/mol are the result of miRTar2GO sensitive. Interaction scores are calculated for all pairs of the unseen sets introducing the highly sensitive miRTar2GO result. The interactions of the unseen set that have hybridization energy  $< -20$  kcal/mol and less than the maximum MFE value associated to the miRNA are shown as miRTar2GO highly specific results. The interactions of the unseen set that have hybridization energy of less than the maximum energy associated with their corresponding miRNA are defined as miRTar2GO specific results. The result of this step is a set of miRNA target candidates, where the miRNA seed region of each candidate has a perfect complementary to the associated mRNA. The miRNA-mRNA interactions that are not experimentally verified are selected to generate the unseen set. The miRNA-mRNA interactions that are experimentally verified are used to calculate hybridization energy values for each miRNA including the minimum and the maximum of MFEs.

previously been shown to be an effective predictor of functional miRNA target sites (51). Since we were dealing with a large number of candidate interactions ( $N \sim 4.2$  millions), we selected the RNAcofold (52) program since it provides fast hybridization simulation of the miRNA-Ago2 CLIPed sequences in a large scale. For each miRNA-mRNA pair a sequence with a length of 40 nts on the CLIPed region starting from the first nucleotide before the seed match was extracted, and the thermodynamic properties of the miRNA-target RNA duplex was calculated with RNAcofold program. We used the length of 40 nts, because our analysis of the experimentally validated miRNA-target interactions showed that the majority of binding sites are shorter than 41 nts. The mean free energy for experimentally verified miRNA-target interactions with a perfect 6-mer seed match was  $-15$  kcal/mol. A value of  $\sim -14$  kcal/mol was also reported by a study using miRTarBase predicted interactions (53). Hence, all candidate interactions with a MFE value greater than  $-15$  kcal/mol were discarded and the remaining interactions were considered for scoring and model building for miRTar2GO.

To generate the training set, we prepared a set of experimentally verified miRNA-target interactions using all interactions from the miRTarBase release 4.5 (54). miRTarBase contains miRNA target sites that have been previously experimentally validated using reporter assays, western blots, microarrays or CLASH experiments. A total number of 37 473 interactions for 577 distinct human miRNAs were identified. We mapped these verified interactions to the Ago2

CLIPed sites whose MFE were  $< -15$  kcal/mol. A total number of 28 546 non-overlapping Ago2 CLIPed containing short (6-mer) canonical seed matches were identified in the investigated cell lines. These miRNA-target interactions were later used in the scoring schema of miRTar2GO. The rest of the miRNA-Ago2 CLIPed candidates ( $N \sim 4$  millions), which were not found in the verified interactions of miRTarBase, were considered as candidate interactions.

#### Defining cell type specific miRNA expression and functional explorations of the predicted targets in miRTar2GO

Tissue specific miRNA expression values of the cell lines included in this study were collected from microRNA.org (55) and miRmine (<http://guanlab.cmb.med.umich.edu/mirmine/>). For each tissue, the expression value of each miRNA is defined as the mean of the given miRNA expression values measured in the different experiments. All miRNA expression values were normalized using  $z$ -score transformation. It is worth mentioning that the miRNA expression values are not used in the target identification process or the scoring schema of miRTar2GO. For the predicted targets of each miRNA, the GO analysis was performed using GStats (56) package in R version 3.2.0. The pathway information for predicted genes generated by four publicly available sources NCI-Nature PID (57), Reactome (58) version 32, BioCarta and KEGG (29) were obtained from hiPathDB (30) and mapped to the miRTar2GO's search engine. The conservation score of each CLIPed site was defined as the total conservation score of

all the nucleotides of the site divided by the length of the site.

### Statistical analysis

For testing statistical significance difference of miRTar2GO's performance on 3' LIFE data, Barnard's unconditional test was used in R version 3.2.0. The z-score transformation of miRNA expression values was performed in SPSS software package.

## RESULT AND DISCUSSION

### Classifications of the predictions of miRTar2GO

The prediction model of miRTar2GO is shown in Figure 1. miRTar2GO ranks the candidate interactions based on MFE of hybridization as a primary parameter. The current version of miRTar2GO offers four sets of predictions: highly specific, specific, sensitive and highly sensitive based on the prediction parameters as defined below.

#### miRTar2GO highly sensitive and miRTar2GO sensitive

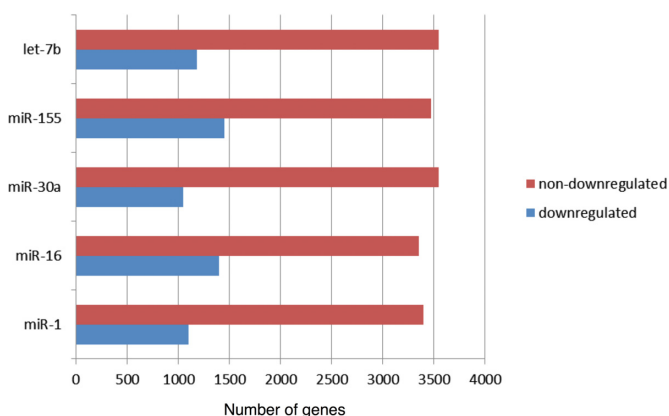
The goal of designing miRTar2GO was to identify as many miRNA targets as possible while keeping the false positive ratio at minimum. The miRTar2GO highly sensitive represents the core prediction result of miRTar2GO: those predictions with a MFE value of  $< -15$  kcal/mol (for details see Materials and Methods). The miRTar2GO highly sensitive scores the predictions based on the binding error of each predicted miRNA–target, relative to all experimentally validated interactions of the given miRNA as follows:

$$E_k = \begin{cases} 1 & X_k \geq \max(X) \\ 1 + \frac{X_k - \max(X)}{\max(X) - \min(X)} & \min(X) < X_k < \max(X) \\ 0 & X_k \leq \min(X) \end{cases}$$

where  $k$  is the predicted miRNA–tmRNA,  $X_k$  is the hybridization energy between the miRNA and the candidate miRNA binding site,  $X$  represents the set of all hybridizations energy values between the given miRNA and its experimentally validated target sites, and  $\max(X)$  and  $\min(X)$  are the biggest and the smallest hybridization values in  $X$  for the given miRNA. The average error of the total predicted interactions for each miRNA is also calculated as follows:

$$\text{Error}_k = \sqrt{\frac{\sum_{i=1}^n (E_k - 1)^2}{n}}$$

The miRTar2GO sensitive classification takes into account the MFE between the miRNA and the targeted 3' UTR. The hybridization energy value between a miRNA sequence and a given 3' UTR has a negative correlation with the length of the binding site and is dependent on the length of the 3' UTR (59). We previously determined that the length of the candidate binding sites was 40 nt. Therefore, we conclude that the MFE of the given miRNA sequence and the 40 nt strand of Ago2 CLIPed sequence is ideal, e.g. when compared to the MFE obtained from the full strand 3' UTR and miRNA. miRTar2GO sensitive only reports those candidates which have a MFE value of  $< -20$  kcal/mol (for details see Materials and Methods).



**Figure 2.** The number of downregulated and non downregulated transcripts in a pSILAC experiment generated by overexpressing miRNAs. *Let-7b*, miR-155, miR-30a, miR-16 and miR-1 were overexpressed in HeLa cells and the changes in protein levels were quantified by pSILAC. Each blue bar represents the number of transcripts which are downregulated with a  $\log_2$ -fold change  $< -0.1$  as the corresponding miRNA is introduced in the cell. Each red bar represents the number of non downregulated transcripts in each experiment.

#### miRTar2GO specific and miRTar2GO highly specific

To predict more specific targets for each miRNA, we further filtered miRNA–tmRNA interactions with MFE greater than the smallest MFE between the given miRNA and its all experimentally verified targets. This stringency resulted in the set of  $\sim 370$  K predicted miRNA–tmRNA interactions which is represented in the miRTar2GO specific. miRTar2GO highly specific limits the result of miRTar2GO specific by applying the previously used filtering criterion of ' $< -20$  kcal/mol' thus discarding those matches which do not pass the RNAhybrid (13,60) algorithm's recommended energy cut-off as well resulting in a set of highly specific potential targets for the given miRNAs.

#### Prediction efficiency and evaluation of miRTar2GO

*Evaluation of miRTar2GO using pulsed Stable Isotope Labeling with Amino Acids in Culture (pSILAC) based miRNA target identification.* Proteome analysis allows experimental identification of miRNA targets in a large scale (61,62) by measuring the changes in protein levels upon introduction of specific miRNAs. We used data from previously published proteomics experiments to evaluate and compare miRTar2GO to other miRNA target prediction tools (31). We have considered a mRNA the target of a miRNA if the protein corresponding to the mRNA is downregulated with a  $\log_2$  fold change of greater than  $-0.1$  when the miRNA is transfected to the cells. This work provides a background dataset, which introduces a total number of 23 379 samples (including mRNAs corresponding to down and non downregulated proteins) in the expression level of proteins followed by the transfection of five miRNAs individually. The mRNAs corresponding to the proteins obtained from these experiments are labeled either as downregulated or not downregulated. This includes 6191 downregulated transcripts as well as 17 188 not downregulated transcripts (Figure 2). Among the 23 379 measured changes in protein ex-

pression (74% non downregulated and 26% downregulated) reported by pSILAC in HeLa cell, miRTar2GO highly sensitive was able to correctly classify 15 115 of them (accuracy of 0.64, TPR = 0.32, TNR = 0.76). The breakdown of these numbers is available in Table 1.

**Evaluation of miRTar2GO using TarBase v7.** Since the primary goal of designing miRTar2GO was to identify less specific miRNA–tmRNA interactions, we set to extend the evaluation to a larger set of experimentally validated miRNA targets. Recently, the latest version of TarBase (63) was released (33) introducing the largest currently available dataset for miRNA–mRNA interactions with more than half a million entries. After processing the dataset and limiting the miRNA–tmRNA interactions to those miRNAs which have a prediction result in miRTar2GO highly sensitive (366 miRNAs), we further evaluated the TPR of miRTar2GO sensitive based on the validated interactions of four types of experiments including HITS-CLIP, PAR-CLIP, microarray and Biotin microarray (64) identifying approximately 190k, 78k, 14k and 6k verified interactions respectively. MiRTar2GO sensitive was able to identify ~70k, ~32k, 3k and ~2500 interactions of HITS-CLIP, PAR-CLIP, microarray and Biotin microarray correctly providing an overall TPR of 0.47 for all interactions provided in TarBase v7.

### Comparison of the performance of miRTar2GO with other miRNA target prediction tools

We collected the prediction results for miRNAs used in the pSILAC experiments (miR-1, miR-155, miR-16, miR-30a and *let-7b*) from eleven miRNA target prediction tools: TargetScan (45), TargetScanS (12), PITA (14), ComiR (65), DIANA-microT (66,67), mirTarget2 (68,69), MR-microT (70,71), PicTar (15), RNA22 (72), TargetMiner (73) and TargetSpy (16). We evaluated the prediction ability of the selected miRNA prediction tools using those metrics which best reflect the TPR. Hence, we selected the precision and recall metrics and evaluated the predictive strength of these predictions and compared them to the miRTar2GO highly sensitive output using the dataset derived from the pSILAC experiment. The precisions and recall were calculated:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

where for each tool: true positives refer to those mRNAs, which associate with significant decrease of protein level ( $\log_2$ -fold change  $< -0.1$ ) when a miRNA is overexpressed, and are correctly predicted as miRNA targets by the prediction tool. False positives refer to those mRNAs which do not coupled with significant protein level changes upon the overexpression of a miRNA in the pSILAC experiment, but are incorrectly predicted as miRNA targets by the prediction tool. False negatives are those mRNAs which are corresponded to proteins, whose level were significantly changed followed by the overexpression of a miRNA in the pSILAC experiment but are not recognized as miRNA targets by

the prediction tool. Each of the miRNA prediction methods varies in the number of the predictions based on the cut-off values to draw a line between miRNA targeted and non-targeted RNAs (Figure 4). Selecting a higher cut-off value in any predictive model results in a prediction with a higher number of true negatives. On the other hand, this may also generate less true positives. For instance, among the reviewed methods, MR-microT produces the highest number of true negatives with correctly predicting 17 116 transcripts as non-targets in the pSILAC dataset, while it recognizes only 186 putative miRNA targets from the actual 6191 downregulated proteins correctly (Figure 5A). On the contrary, selecting a low cut-off value results in a predictor with a higher number of true positives (Figure 4) may decrease the number of true negatives. For instance, among the reviewed methods, miRTar2GO has the highest number of true positives with a value of 1933, while its true negative ratio is relatively low compared to other tools (true negative ratio = 76.69%). Thus, precision and recall alone are not sufficient to compare the performance of these prediction methods (74). Based on precision and recall of each method, we obtained  $F_1$  score and  $G$  score for each tool, which are harmonic and geometric average of precision and recall, respectively (Figure 5B and C) and were calculated using the following equations:

$$F_1 = 2 * \text{precision} * \frac{\text{recall}}{\text{precision} + \text{recall}}$$

$$G = \sqrt{\text{precision} * \text{recall}}$$

Based on the estimated accuracy calls, miRTar2GO has the highest values for both  $F_1$  score and  $G$  score among the prediction tools used in this comparison.

miRTar2GO offers quantitative values for rating different aspects of the predicted interactions. For each prediction, it provides an interaction rate and two site rates. The interaction rate is based on the seed match location on the Ago2 cross-linked site. Most seed matches are frequently positioned 1–2 nt downstream of the predominant cross-linking site within the CCRs. This places the cross-linking site near the centre of the Ago2–miRNA–target ternary complex, which contributes directly to miRNA binding. To that end, for each seed match, the relative position of seed match on the Ago2 CLIPed site is considered. A seed match is given a higher rate if it is closer to the centre of the Ago2 CLIPed site. The first site rate is simply based on the cross-species conservation score of the Ago2 CLIPed region. The second site rate introduced by miRTar2GO is based on the degree of occurrence of the Ago2 CLIPed site among different cell lines. For each Ago2 CLIPed site, the number of cell lines that have an Ago2 interaction in the same genomic coordinate is obtained. These two scores could be helpful to make the result of prediction more context specific.

### Cell line specific miRNA target prediction by miRTar2GO

The Ago2 interactome could vary from one cell type to another. Since the predictions generated by miRTar2GO are based on Ago2 CLIP data, therefore it computes different set of targets for each miRNA in different cells (Table 2).

**Table 1.** Number of miRTar2GO predictions for the five miRNAs investigated in the pSILAC experiment in different cell lines

miRNA	Downregulated	True positives	Non downregulated	True negatives
miR-1	1126	333	3375	2552
miR-16	1389	404	3318	2699
miR-30a	1038	188	3524	3179
miR-155	1449	398	3426	2919
Let-7b	1189	610	3545	1833

**Table 2.** Total number of binding sites predicted to be functional by miRTar2GO for the five miRNAs used in the pSILAC experiment in different cell lines

miRNA	HeLa	BC1	EF3D	hESC	293s	Human_Lymphoblastic
miR-1	4845	2821	1754	664	921	497
miR-16	4028	1506	734	199	1349	488
miR-30a	2595	857	398	106	773	328
miR-155	3466	1496	1366	212	705	418
Let-7b	9912	5076	3506	1256	2663	1136
Total	24 846	11 756	7758	2437	6411	2867

The difference in the number of predictions for the same miRNA in different cell lines is directly linked to the varying number of Ago2 CLIPed sites in different cell lines.

**Table 3.** Number of correctly identified downregulated transcripts by miRNAs used in the pSILAC experiment carried out in HeLa cells which have overlapping Ago2 CLIPed binding sites in the presented cell lines

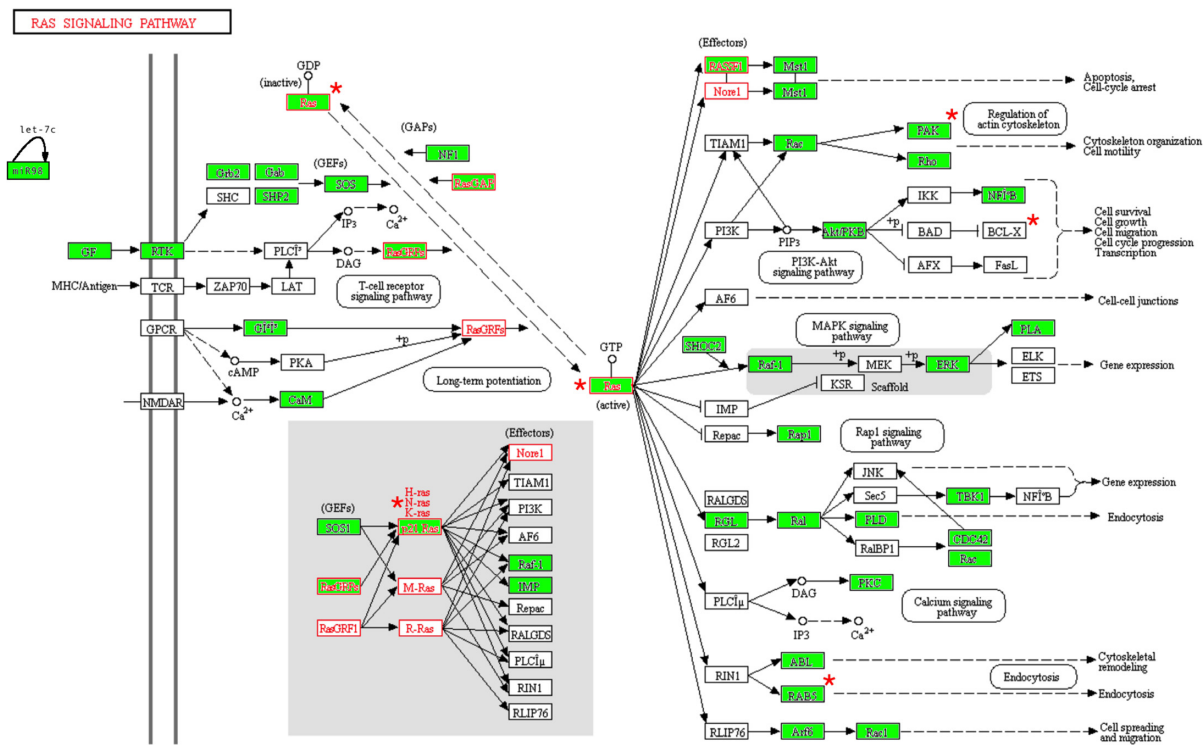
	miR-1	miR-16	miR-30a	miR-155	Let-7b	Total
293s	59 (0.22)	120 (0.46)	65 (0.4)	73 (0.24)	98 (0.27)	415 (0.31)
BC1	155 (0.60)	113 (0.43)	62 (0.82)	131 (0.44)	193 (0.53)	654 (0.56)
EF3D	97 (0.37)	61 (0.23)	23 (0.14)	126 (0.42)	101 (0.27)	408 (0.28)
Lymphoblastic	29 (0.11)	46 (0.17)	19 (0.11)	43 (0.14)	32 (0.08)	169 (0.12)
hESC	28 (0.10)	20 (0.07)	13 (0.08)	6 (0.02)	47 (0.12)	114 (0.07)

The first number in each column represents the frequency of the binding sites which are correctly identified by miRTar2GO in HeLa cell and have an overlapping binding site in the CLIP data of the corresponding cell type. Each number in parenthesis shows the percentage of the correctly identified downregulated target by miRTar2GO in HeLa cell which has an overlapping binding site in the associated cell line. The first number of the last column represents the total number of downregulated transcripts in HeLa cells which are identified by miRTar2GO and have an overlapping Ago2-mRNA footprint in the associated cell line. The second number in the last column represents the percentage of correctly identified downregulated target by miRTar2GO in HeLa cells.

Depending on the interaction of Ago2 and the 3' UTR in each cell line, miRTar2GO determines if the prediction is common to all cell lines or specific to only one or a few. For five miRNAs investigated in this study, HeLa cells have the highest number of predictions, followed by BC1, EF3D, 293s, Human Lymphoblastic and finally hESC. Not only does each cell have a different number of potentially functional miRNA binding sites, but also the degree of overlap for predicted binding sites for each miRNA varies from one cell line to another (75) (Supplementary Figure S1A–E). Table 3 shows the frequency and percentage of miRTar2GO identified downregulated transcripts identified in the pSILAC experiment in five different cell lines, compared to HeLa cells. BC1 cells have the highest degree of overlap with HeLa cells with a value of 56%. On the other hand, the overlap ratio for lymphoblastic cells and hESC cells are relatively low, only 0.12 and 0.07, respectively. We further evaluated the degree of overlap between number of correctly identified downregulated transcripts in pSILAC by miRTar2GO for each miRNA in different cell types (Supplementary Figure S2A–E). For each miRNA, there are sets of target sites which are cell type specific. For example, there are 88 *let-7b* targets, which have predicted binding sites in only, BC1 and HeLa cells. These results may strengthen the notion of context dependency of miRNA targeting.

### Pathway analysis based evaluation

Recently, Wolter and colleagues (32) carried out a large scale dual-luciferase reporter assay to verify miRNA targets called 3'LIFE. In this experiment they identified two different sets of *let-7c* targets. The first set included 19 genes, which have a miRNA repression index (RI) value of less than the mean standard error of the assay ( $RI < 15\%$ ). This index is defined as the average of the normalized miRNA repression values and is used to rank the 3' UTRs of the dual luciferase experiments. The second set includes 37 high-confidence targets with an RI value lower than 0.80 and a *P*-value lower than 0.05. We compared these findings to the prediction of miRTar2GO for *let-7c*, and we have found that 63% of the genes of the first set and 46% of the genes of the second set are accurately identified by miRTar2GO respectively. Wolter *et al.* (32) showed that 3'LIFE identified top targets for *let-7c* such as ID1, HSF1, CRK, DNMT1, ARID3A, EZH2, RhoB and RhoC function within the RAS signaling pathway. Among these genes, only ARID3A and EZH2 were bioinformatically predicted before. In contrast, miRTar2GO was able to identify half of these genes (ARID3A, CRK, EZH2 and RHOB). Moreover, miRTar2GO further predicts that 27% of RAS signaling pathway genes are targeted by *let-7c* including potentially novel targeted genes (Figure 3). miRTar2GO also

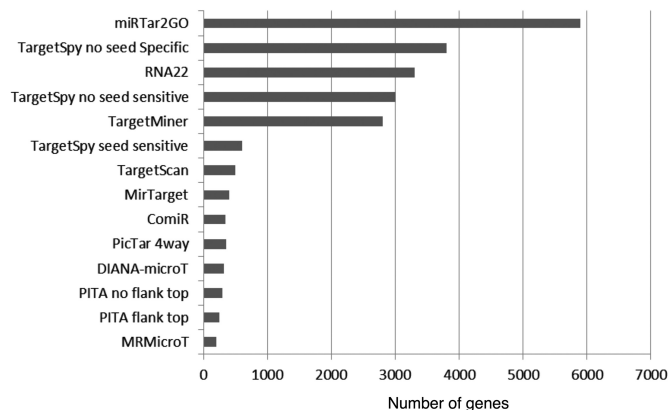


**Figure 3.** *let-7c* is predicted to target the RAS signaling pathway and miR-98. The figure shows the modified diagram of the RAS signaling pathway generated by hiPathDB. \* indicates experimentally validated *let-7c* targets in RAS signaling pathway used in training of miRTar2GO. The Green colored boxes indicate *let-7c* targets in RAS signaling pathway predicted by miRTar2GO. miR-98, a member of the *let-7* miRNA family, is added to the figure.

predicts that *let-7c* binds to pri-miR-98 with a probability of 0.99 (a miRNA in the *let-7* family). It has been shown that *let-7* in worm and human cells are the part of a positive feedback loop in which the Ago2 associated mature *let-7* recognizes complementary sequences on its own pri-miRNA that don't correspond to the miRNA hairpin sequence and stimulate pri-miRNA processing (76). The predicted *let-7c* target site on miR-98 also matches region of miR-98 that is outside of the hairpin sequence. To our knowledge no target prediction approach has considered such miRNA-pri-miRNA interaction before.

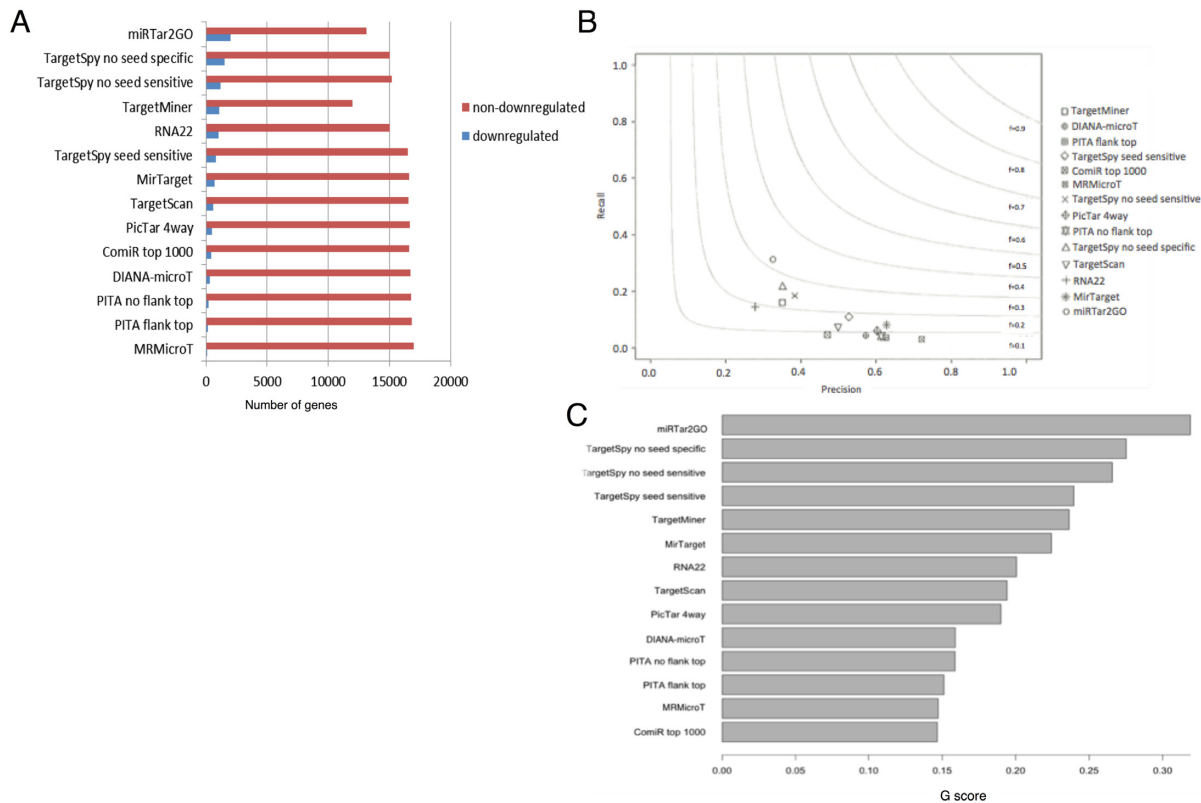
**Design and implemented algorithm layout of the miRTar2GO web service**

The design and implementation of miRTar2GO have been completed using MySQL as a back end database and PHP as a front end for visual interactions. The front view of the miRTar2GO allows the user to select the miRNA target predictions based on four criteria: highly specific, specific, sensitive and highly sensitive modes. The default scoring mechanism of the miRTar2GO highly sensitive is based on the similarity of the thermodynamic feature of the prediction and the training set. The default scoring feature for the other three computational modes is the hybridization energy of miRNA sequence and the predicted site. Based on the user selected mode, users will be re-directed to the respective pages, which allows to search for the target predictions based on type of miRNAs, type of target gene, miRNA specific sequence, chromosomal coordinates and type of cell lines. This will result in displaying cell specific



**Figure 4.** Comparison of different miRNA target prediction tools. The total number of transcripts predicted as miRNA target by popular miRNA target prediction tools from the downregulated transcripts identified by pSILAC followed by the overexpression of five miRNAs in HeLa cells.

miRNA targets. We emphasize that miRTar2GO does not introduce experimentally verified interactions in the result set as these interactions are used in the training set. Additional functionalities have been implemented to filter the results either by using hybridization energy, CCR conservation score, gene name, seed match distance to the CCR center, and the number of cell types in which the particular miRNA targeting are predicted. After submission, the results page displays curated information about the miRNA's predicted targets that are filtered accordingly to criteria se-



**Figure 5.** Evaluation of miRTar2GO. (A) The number of correctly identified miRNA targets by different target prediction tools using the pSILAC experiments (blue bars). The Red bars represent the correctly identified non tmRNAs. (B) Comparison of  $F_1$  scores of different miRNA target prediction methods using the pSILAC data set. The x-axis represents the precision and the y-axis represents the recall. Each spectrum in the figure represents one sub-range of the averaged value for the precision and the recall. (C) Comparison of  $G$  scores of different miRNA prediction tools using the pSILAC dataset.

lected by the user. In addition, the page also displays the corresponding biological pathways and the KEGG based assigned KO of the predicted targets. All the predicted targets are hyperlinked with corresponding hiPathDB links to allow visualization of the pathways with the embedded predicted miRNAs. Furthermore, upon user's request, the GO enrichment of the functional ontologies assigned to the predicted targets against the background ontologies is calculated in the background using Gostat (56) and the result is sent to the user's specified email address. The analysis of the miRNA binding sites in the user provided CLIP-Seq data is also available upon request.

## CONCLUSIONS

By combining short seed match criterion, thermodynamic characteristics of the experimentally validated miRNA–target interactions and Ago2 CLIP-Seq data, we have developed a new computational method that predicts canonical binding sites of miRNAs in a cell type specific manner. We have shown that our method (miRTar2GO highly sensitive) is able to identify more experimentally validated miRNA–target interactions compared to other miRNA target prediction tools, as it does not impose a high thermodynamic restriction. miRTar2GO is not dependent on long seed matches or on conservation score. The ranking system of miRTar2GO computes energy characteristics of each miRNA and its experimentally validated targets to

rank the predictions of each miRNA based on a training set. Due to these advantages, miRTar2GO is able to predict miRNA–target interactions that were not determined by other miRNA target prediction tools. miRTar2GO provides four different prediction modes specifically designed to adjust the sensitivity. The designed adjustment levels are based on thermodynamic features of the verified interactions, which is novel in its nature. These variations in sensitivity are designed to decrease the false-positive rate and increase the true-negative rate.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We would like to thank Dominik Beck for advising on the statistical analysis and for commenting on the manuscript. We also would like to thank Jom Sirisatien and Paul Kennedy for their help in designing the web server and Andrea Toth for proofreading the manuscript. We also would like to thank Vahid Behbood for his comments on the score engine design of miRTar2GO.



## FUNDING

ARC Future Fellowship (to G.H.). Funding for open access charge: UTS CHT core funding: 325630/2016010 and ARC FT: 110100455.

*Conflict of interest statement.* None declared.

## REFERENCES

- Tran, N. and Hutvagner, G. (2013) Biogenesis and the regulation of the maturation of miRNAs. *Essays Biochem.*, **54**, 17–28.
- Friedman, R.C., Farh, K.K.-H., Burge, C.B. and Bartel, D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.
- Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H. and Siomi, M.C. (2005) Slicer function of Drosophila Argonautes and its involvement in RISC formation. *Genes Dev.*, **19**, 2837–2848.
- Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L. and Hannon, G.J. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science*, **305**, 1437–1441.
- Rivas, F. V., Tolia, N.H., Song, J.J., Aragon, J.P., Liu, J., Hannon, G.J. and Joshua-Tor, L. (2005) Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat. Struct. Mol. Biol.*, **12**, 340–349.
- Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G. and Tuschl, T. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell*, **15**, 185–197.
- Khvorova, A., Reynolds, A. and Jayasena, S.D. (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell*, **115**, 209–216.
- Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–33.
- Eulalio, A., Huntzinger, E. and Izaurralde, E. (2008) Getting to the root of miRNA-mediated gene silencing. *Cell*, **132**, 9–14.
- Mishima, Y., Fukao, A., Kishimoto, T., Sakamoto, H., Fujiwara, T. and Inoue, K. (2012) Translational inhibition by deadenylation-independent mechanisms is central to microRNA-mediated silencing in zebrafish. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 1104–1109.
- Kozak, M. (2008) Faulty old ideas about translational regulation paved the way for current confusion about how microRNAs function. *Gene*, **423**, 108–115.
- Lewis, B.P., Burge, C.B. and Bartel, D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.
- Rehmsmeier, M., Steffen, P., Höchsmann, M. and Ho, M. (2004) Fast and effective prediction of microRNA / target duplexes. 1507–1517.
- Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. and Segal, E. (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**, 1278–1284.
- Krek, A., Grün, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M. et al. (2005) Combinatorial microRNA target predictions. *Nat. Genet.*, **37**, 495–500.
- Sturm, M., Hackenberg, M., Langenberger, D. and Frishman, D. (2010) TargetSpy: a supervised machine learning approach for microRNA target prediction. *BMC Bioinformatics*, **11**, 292.
- Marín, R.M. and Vaníek, J. (2011) Efficient use of accessibility in microRNA target prediction. *Nucleic Acids Res.*, **39**, 19–29.
- Gaidatzis, D., van Nimwegen, E., Hausser, J. and Zavolan, M. (2007) Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics*, **8**, 69.
- Peterson, S.M., Thompson, J.A., Ufkin, M.L. and Sathyanarayana, P. (2014) Common features of microRNA target prediction tools, **5**, 1–10.
- Dweep, H., Sticht, C. and Gretz, N. (2013) In-Silico Algorithms for the Screening of Possible microRNA Binding Sites and Their Interactions. *Curr. Genomics*, **14**, 127–136.
- Ellwanger, D.C., Büttner, F., Mewes, H.-W. and Stümpflen, V. (2011) The sufficient minimal set of miRNA seed types. *Bioinformatics*, **27**, 1346–50.
- Zhang, Y. and Verbeek, F.J. (2010) Comparison and integration of target prediction algorithms for microRNA studies. *J. Integr. Bioinform.*, **7**, 1–13.
- Yue, D., Liu, H. and Huang, Y. (2009) Survey of computational algorithms for microRNA target prediction, 478–492.
- Thomas, M., Lieberman, J. and Lal, A. (2010) Desperately seeking microRNA targets. *Nat. Struct. Mol. Biol.*, **17**, 1169–1174.
- Erhard, F., Haas, J., Lieber, D., Malterer, G., Jaskiewicz, L., Zavolan, M., Dölken, L. and Zimmer, R. (2014) Widespread context dependency of microRNA-mediated regulation. *Genome Res.*, **24**, 906–919.
- Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M. Jr, Jungkamp, A.C., Munschauer, M. et al. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, **141**, 129–41.
- Chi, S.W., Zang, J.B., Mele, A. and Darnell, R.B. (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, **460**, 479–86.
- Helwak, A., Kudla, G., Dudnakova, T. and Tollervey, D. (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*, **153**, 654–665.
- Kanehisa, M. and Goto, S. (2000) KEGG: Kyoto Encyclopaedia of Genes and Genomes. *Nucleic Acids Res.*, **28**, 27–30.
- Yu, N., Seo, J., Rho, K., Jang, Y., Park, J., Kim, W.K. and Lee, S. (2012) hiPathDB: A human-integrated pathway database with facile visualization. *Nucleic Acids Res.*, **40**, D797–D802.
- Selbach, M. et al. (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature*, **455**, 58–63.
- Wolter, J.M., Kotagama, K., Pierre-Bez, A.C., Firago, M. and Mangone, M. (2014) 3' LIFE: a functional assay to detect miRNA targets in high-throughput. *Nucleic Acids Res.*, **42**, e132.
- Vlachos, I.S., Paraskevopoulou, M.D., Karagkouni, D., Georgakilas, G., Vergoulis, T., Kanellos, I., Anastasopoulos, I.L., Maniou, S., Karathanou, K., Kalfakakou, D. et al. (2015) DIANA-TarBase v7.0: Indexing more than half a million experimentally supported miRNA:mRNA interactions. *Nucleic Acids Res.*, **43**, D153–D159.
- Clark, P.M., Lohner, P., Quann, K., Brody, J., Londin, E.R. and Rigoutsos, I. (2014) Argonaute CLIP-Seq reveals miRNA targetome diversity across tissue types. *Sci. Rep.*, **4**, 5947.
- Yang, J.-H., Li, J.H., Shao, P., Zhou, H., Chen, Y.Q. and Qu, L.H. (2011) starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Res.*, **39**, D202–D209.
- Li, J.-H., Liu, S., Zhou, H., Qu, L.-H. and Yang, J.-H. (2014) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.*, **42**, D92–D97.
- Xue, Y., Ouyang, K., Huang, J., Zhou, Y., Ouyang, H., Li, H., Wang, G., Wu, Q., Wei, C., Bi, Y. et al. (2013) Direct conversion of fibroblasts to neurons by reprogramming PTB-regulated MicroRNA circuits. *Cell*, **152**, 82–96.
- Gottwein, E., Corcoran, D.L., Mukherjee, N., Skalsky, R.L., Hafner, M., Nusbaum, J.D., Shmulaitapam, P., Love, C.L., Dave, S.S., Tuschl, T. et al. (2011) Viral microRNA targetome of KSHV-infected primary effusion lymphoma cell lines. *Cell Host Microbe*, **10**, 515–526.
- Skalsky, R.L., Corcoran, D.L., Gottwein, E., Frank, C.L., Kang, D., Hafner, M., Nusbaum, J.D., Feederle, R., Delecluse, H.J., Luftig, M.A. et al. (2012) The viral and cellular microRNA targetome in lymphoblastoid cell lines. *PLoS Pathog.*, **8**, e1002484.
- Lipchina, I., Elkabetz, Y., Hafner, M., Sheridan, R., Mihailovic, A., Tuschl, T., Sander, C., Studer, L. and Betel, D. (2011) Genome-wide identification of microRNA targets in human ES cells reveals a role for miR-302 in modulating BMP response. *Genes Dev.*, **25**, 2173–2186.
- Karginov, F.V. and Hannon, G.J. (2013) Remodeling of Ago2-mRNA interactions upon cellular stress reflects miRNA complementarity and correlates with altered translation rates. *Genes Dev.*, **27**, 1624–1632.
- Riley, K.J., Rabinowitz, G.S., Yario, T.A., Luna, J.M., Darnell, R.B. and Steitz, J.A. (2012) EBV and human microRNAs co-target oncogenic and apoptotic viral and human genes during latency. *EMBO J.*, **31**, 2207–2221.
- Kishore, S., Jaskiewicz, L., Burger, L., Hausser, J., Khorshid, M. and Zavolan, M. (2011) A quantitative analysis of CLIP methods for

- identifying binding sites of RNA-binding proteins. *Nat. Methods*, **8**, 559–564.
44. Gortmaker, S.L., Hosmer, D.W. and Lemeshow, S. (1994) Applied logistic regression. *Contemp. Sociol.*, **23**, 159.
  45. Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P. and Bartel, D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, **27**, 91–105.
  46. Kozomara, A. and Griffiths-Jones, S. (2014) MiRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**.
  47. Chi, S.W., Hannon, G.J. and Darnell, R.B. (2012) An alternative mode of microRNA target recognition. *Nat. Struct. Mol. Biol.*, **19**, 321–327.
  48. Agarwal, V., Bell, G.W., Nam, J.W. and Bartel, D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, **4**, doi:10.7554/eLife.05005.
  49. Clark, P.M., Lohar, P., Quann, K., Brody, J., Londin, E.R. and Rigoutsos, I. (2014) Argonaute CLIP-Seq reveals miRNA targetome diversity across tissue types. *Sci. Rep.*, **4**, 5947.
  50. Da Hsu, S., Lin, F.M., Wu, W.Y., Liang, C., Huang, W.C., Chan, W.L., Tsai, W.T., Chen, G.Z., Lee, C.J., Chiu, C.M. *et al.* (2011) MiRTarBase: a database curates experimentally validated microRNA–target interactions. *Nucleic Acids Res.*, **39**, 163–169.
  51. Lekprasert, P., Mayhew, M. and Ohler, U. (2011) Assessing the utility of thermodynamic features for microRNA target prediction under relaxed seed and no conservation requirements. *PLoS One*, **6**, e20622.
  52. Bernhart, S.H., Tafer, H., Mückstein, U., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2006) Partition function and base pairing probabilities of RNA heterodimers. *Algorithms Mol. Biol.*, **1**, 3.
  53. Hsu, S.-D., Lin, F.M., Wu, W.Y., Liang, C., Huang, W.C., Chan, W.L., Tsai, W.T., Chen, G.Z., Lee, C.J., Chiu, C.M. *et al.* (2011) miRTarBase: a database curates experimentally validated microRNA–target interactions. *Nucleic Acids Res.*, **39**, D163–D169.
  54. Hsu, S., Tseng, Y.T., Shrestha, S., Lin, Y.L., Khaleel, A., Chou, C.H., Chu, C.F., Huang, H.Y., Lin, C.M., Ho, S.Y. *et al.* (2014) miRTarBase update 2014: an information resource for experimentally validated miRNA–target interactions, **42**, 78–85.
  55. Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M. *et al.* (2007) A Mammalian microRNA Expression Atlas Based on Small RNA Library Sequencing. *Cell*, **129**, 1401–1414.
  56. Falcon, S. and Gentleman, R. (2007) Using GOstats to test gene lists for GO term association. *Bioinformatics*, **23**, 257–258.
  57. Schaefer, C.F., Anthony, K., Krupa, S., Buchoff, J., Day, M., Hannay, T. and Buetow, K.H. (2009) PID: the pathway interaction database. *Nucleic Acids Res.*, **37**, D674–D679.
  58. Croft, D., O’Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B. *et al.* (2011) Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.*, **39**, D691–D697.
  59. Rehmsmeier, M., Steffen, P., Hochsmann, M. and Giegerich, R. (2004) Fast and effective prediction of microRNA/target duplexes. *RNA*, **10**, 1507–1517.
  60. Krüger, J. and Rehmsmeier, M. (2006) RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.*, **34**, W451–W454.
  61. Witkos, T.M., Koscianska, T. and Krzyzosiak, W.J. (2011) Practical aspects of microRNA target prediction. *Curr. Mol. Med.*, **11**, 93–109.
  62. Baek, D., Villén, J., Shin, C., Camargo, F.D., Gygi, S.P. and Bartel, D.P. (2008) The impact of microRNAs on protein output. *Nature*, **455**, 64–71.
  63. Sethupathy, P., Corda, B. and Hatzigeorgiou, A.G. (2006) TarBase: a comprehensive database of experimentally supported animal microRNA targets. *RNA*, **12**, 192–197.
  64. Thomson, D.W., Bracken, C.P. and Goodall, G.J. (2011) Experimental strategies for microRNA target identification. *Nucleic Acids Res.*, **39**, 6845–6853.
  65. Coronnello, C. and Benos, P.V. (2013) ComiR: combinatorial microRNA target prediction tool. *Nucleic Acids Res.*, **41**, 1–6.
  66. Maragkakis, M., Alexiou, P., Papadopoulos, G.L., Reczko, M., Dalamagas, T., Giannopoulos, G., Goumas, G., Koukis, E., Kourtis, K., Simossis, V.A. *et al.* (2009) Accurate microRNA target prediction correlates with protein repression levels. *BMC Bioinformatics*, **10**, 295.
  67. Maragkakis, M., Reczko, M., Simossis, V.A., Alexiou, P., Papadopoulos, G.L., Dalamagas, T., Giannopoulos, G., Goumas, G., Koukis, E., Kourtis, K. *et al.* (2009) DIANA-microT web server: Elucidating microRNA functions through target prediction. *Nucleic Acids Res.*, **37**, 273–276.
  68. Wang, X. and El Naqa, I.M. (2008) Prediction of both conserved and nonconserved microRNA targets in animals. *Bioinformatics*, **24**, 325–332.
  69. Wang, X. (2008) miRDB: a microRNA target prediction and functional annotation database with a wiki interface. *RNA*, **14**, 1012–1017.
  70. Kanellos, I., Dalamagas, T., Hatzigeorgiou, A., Fleming, B. Al and Athena, R.C. (2014) MR-microT: a MapReduce-based microRNA target prediction method. *SSDBM*.
  71. Reczko, M., Maragkakis, M., Alexiou, P., Grosse, I. and Hatzigeorgiou, A.G. (2012) Functional microRNA targets in protein coding sequences. *Bioinformatics*, **28**, 771–776.
  72. 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.
  73. Bandyopadhyay, S. and Mitra, R. (2009) TargetMiner: microRNA target prediction with systematic identification of tissue-specific negative examples. *Bioinformatics*, **25**, 2625–2631.
  74. Powers, D. (2011) Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *J. Mach. Learn. Technol.*, **2**, 37–63.
  75. Connerty, P., Ahadi, A. and Hutvagner, G. (2015) RNA Binding Proteins in the miRNA pathway. *Int. J. Mol. Sci.*, **17**, e31.
  76. Zisoulis, D.G., Kai, Z.S., Chang, R.K. and Pasquinelli, A.E. (2012) Autoregulation of microRNA biogenesis by let-7 and Argonaute. *Nature*, **486**, 541–544.