D360–D370 *Nucleic Acids Research, 2018, Vol. 46, Database issue* *Published online 29 November 2017*
doi: 10.1093/nar/gkx1144

mirDIP 4.1—integrative database of human microRNA target predictions

Tomas Tokar[1], Chiara Pastrello[1], Andrea E.M. Rossos[1], Mark Abovsky[1], Anne-Christin Hauschild[1], Mike Tsay[1], Richard Lu[1] and Igor Jurisica[1,2,3,4,*]

[1]Krembil Research Institute, University Health Network, Toronto, Ontario M5T 2S8, Canada, [2]Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 1L7, Canada, [3]Department of Computer Science, University of Toronto, Toronto, Ontario M5S 3G4, Canada and [4]Institute of Neuroimmunology, Slovak Academy of Sciences, Bratislava, 845 10, Slovakia

Received September 17, 2017; Revised October 27, 2017; Editorial Decision October 28, 2017; Accepted October 30, 2017

ABSTRACT

MicroRNAs are important regulators of gene expression, achieved by binding to the gene to be regulated. Even with modern high-throughput technologies, it is laborious and expensive to detect all possible microRNA targets. For this reason, several computational microRNA–target prediction tools have been developed, each with its own strengths and limitations. Integration of different tools has been a successful approach to minimize the shortcomings of individual databases. Here, we present mirDIP v4.1, providing nearly 152 million human microRNA–target predictions, which were collected across 30 different resources. We also introduce an integrative score, which was statistically inferred from the obtained predictions, and was assigned to each unique microRNA–target interaction to provide a unified measure of confidence. We demonstrate that integrating predictions across multiple resources does not cumulate prediction bias toward biological processes or pathways. mirDIP v4.1 is freely available at http://ophid.utoronto.ca/mirDIP/.

INTRODUCTION

MicroRNAs (miRNAs) are conserved short non-coding RNAs that serve as post-transcriptional regulators of gene expression (1). miRNAs exert their function in concert with associated proteins of the Argonaute family (AGO) (2). The complex can recognize target mRNAs through seed sequences complementary (partially or completely) to the miRNA, with the 5′ terminus of the miRNA and the 3'UTR of the target mRNA most frequently involved (3). The binding results in accelerated mRNA degradation, translational repression, mRNA deadenylation or mRNA destabilization (2), and can interfere with translation initiation, repress translation elongation and termination, or recruit cofactors involved in protein degradation and sequestration (4).

This gives miRNAs the capacity to regulate most protein-coding transcripts (5). miRNAs are involved in diverse biological processes, including development, cell growth and metabolism (2). Growing evidence implicates miRNAs as key players in major human pathologies, including cancer (6), autoimmune diseases (7) and mental disorders (8).

Accurate identification of miRNA target genes has been a focus of computational biology for many years. Its importance motivated the development of a wide variety of computational resources dedicated to predict which genes are targeted by particular miRNAs. Since 2006, about 60 such resources have been published, according to OMICtools (https://omictools.com/). Individual resources differ in methodologies they use, ranging from assessment of evolutionary conservation of the putative miRNA binding sites, to machine learning classification algorithms.

With growing numbers of miRNA target prediction resources emerged the need for their integration. One of the earliest attempts to address this was the microRNA Data Integration Portal (mirDIP) (9), comprising 9.5 million predictions obtained from seven different resources. Since its release, mirDIP acquired over 13 500 unique users from 86 countries (GoogleAnalytics, September 2017).

Here we present the newest version of the mirDIP database, comprising almost 152 million human miRNA–target predictions obtained from 30 independent resources. In contrast to its previous versions, mirDIP now provides an integrative score assigned to each unique miRNA–target interaction, statistically inferred using the predictions obtained from individual resources. We show that the integrative score provides more accurate predictions than those obtained from any of the individual resources we integrated.

A typical pipeline of miRNA research involves selection of miRNAs of interest (e.g. deregulated miRNAs), followed by identification of their targets using miRNA-prediction

*To whom correspondence should be addressed. Tel: +1 416 581 7437; Fax: +1 416 603 6274; Email: juris@ai.utoronto.ca

tools and subsequent pathway (or other) enrichment analysis performed to examine biological function associated with those miRNAs and their targets. MiRNA–target predictions tend to identify targets belonging to specific biological processes or pathways, which is referred to as prediction bias (10,11). Thus, a major challenge associated with miRNA–target prediction is to avoid or diminish these biases. Reducing bias is of special importance when integrating diverse predictions, as it may potentially lead to bias cumulation. We demonstrate that combining the predictions across methodologies using our integrative score causes no bias cumulation.

MATERIALS AND METHODS

Acquisition of the miRNA–target prediction data

We summarized the miRNA–target prediction resources that provide original predictions on human miRNA–target associations, and were published or updated between the years 2006 and 2017 (this also includes resources originally published before 2006, but updated after 2006). We considered only resources whose predictions are evaluated by any type of quantitative measure, representing resource-subjective confidence assigned to a given prediction (e.g. binding energy, statistical significance, etc.). Predictions from the individual resources were either retrieved from the website, supplementary materials of the corresponding publication, or generated *de novo* by executing the prediction algorithm locally.

Some resources provide multiple prediction sets, obtained by applying distinct methodologies, or using different parameterization of the underlying algorithms. In these cases, all the available prediction sets were pre-processed, normalized and benchmarked separately. However, as described later, for each of the resources, we eventually retained only the dataset which provided the most accurate predictions.

Resources whose predictions couldn't be retrieved from their websites were disregarded. Similarly, algorithms whose installation or execution failed, as well as those whose runtime was unreasonably protracted (expected runtime > 90 days), were not included. For more details on data acquisition, see Supplementary Materials.

Processing and standardization of the miRNA–target prediction datasets

All the prediction sets were first preprocessed to be converted into a common format, that is, each entry comprised a gene symbol and miRNA name constituting the predicted interaction, followed by the measure quantifying resource-subjective confidence in the given prediction (e.g. binding energy, statistical significance or some type of score) as obtained from the given resource.

For each prediction set, we then performed standardization of the gene symbols and miRNA names with respect to the most current nomenclatures. Gene symbols were standardized according to Hugo Gene Nomenclature Committee (HGNC) (April 2017), using R package `HGNChelper v.0.3.5`. miRNA names were standardized according to miRBase v.21, using R package `miRNAmeConverter v.1.4.0` (12).

When gene symbol standardization yielded more than one standard symbol from the original symbol, the original entry was replaced by multiple entries with the same prediction score. No miRNA names yielded duplicate standards. If standardization failed because no matching standard symbol or miRNA name was found, the entry was removed.

Normalization of the miRNA–target predictions

For each prediction set, we ranked all the predictions by assigning a value r from the interval $\langle 0, 1\rangle$, where 0 was assigned to the most confident prediction from the given dataset, and 1 was assigned to the least confident prediction in the given dataset. Here we denote r_{ij} as rank of the j-th miRNA-target pair, as obtained from the i-th resource.

Some of the prediction sets contained multiple predictions for the identical miRNA–target pair, often predicted with varying confidence. This redundancy typically originates from the fact that most of the genes contain multiple putative miRNA binding sites within their 3'UTR. Therefore, resources that evaluate each of the potential binding sites separately may generate redundant predictions.

In particular, mRNAs can act as competing endogenous RNAs (ceRNAs) (or miRNA sponges or decoys) to prevent miRNAs from binding to their authentic targets, achieving a fine level of miRNA regulation (13). ceRNAs are supposed to be more effective when containing a higher amount of miRNA binding sequences (14), to achieve a stronger effect in the finely tuned equilibrium of molecules in a cell.

In order to mitigate the effect of multiple binding sites in a target, in each dataset, we replaced all the redundant predictions by a single prediction whose rank was calculated as a product of the three lowest ranks of the individual predictions (corresponding to the three most confident predictions). This way, we eliminated the effect of ceRNAs, yet we keep favoring genes containing multiple binding sites, as more likely targets of the given miRNA.

Benchmarking of the miRNA–target predictions

Depending on the underlying methodology, individual resources quantify confidence of their predictions by various measures. These may include binding energy, measure of evolutionary conservation, statistical measures such as *P*-value, or more abstract quantities resulting from the method applied. Here we aimed at mapping the separate measures into a universal measure, allowing direct quantitative comparison of the confidence with which the individual resources predict given interaction.

This was done by applying an approach similar (yet substantially adjusted) to one previously used by Junge *et al.* (15). First, each dataset was reduced to a subset comprising only miRNA–target pairs composed of miRNAs or genes (or both) covered by the experimentally validated interactions from the benchmarking dataset (described later). In other words, any miRNA–target pair where the miRNA and gene are both absent among the benchmarking dataset was excluded. A window of size $\Delta r = 0.05$, was slid down

from $r = 0$ to $r = 1 - \Delta r$, with step size $\delta r = 0.01$. For each prediction set, all predictions whose rank fell into the window were selected, and precision was calculated, defined as a number of predicted miRNA–gene pairs present in the benchmarking dataset (true positives) relative to the total number of predictions selected. The resulting precision was assigned to the median rank calculated across the selected predictions.

For each prediction set we thus obtained a set of ranks and associated precisions, as calculated with respect to the benchmarking dataset. Ranks and associated precisions were log-transformed and their relationship was fitted by the quadratic function, imposing the rank as an independent variable. The resulting function, specific for each prediction set, was then applied to interpolate the precision of all the predictions from the given prediction set. The obtained values we refer to as confidence scores, and we denote s_{ij} as a confidence score of the j-th miRNA–target pair, obtained from the i-th resource.

When multiple sets of predictions were retrieved from the given resource, we considered only the most precise set (evaluated by geometric mean of the resulting confidence scores) while the remaining prediction sets were disregarded. This way, we reduced total number of prediction sets to 30, matching the number of resources.

To provide a more intuitive rating of the individual predictions, we categorized them into 4 distinct confidence classes, labeled as *'very high'*, *'high'*, *'medium'* and *'low'* confidence, corresponding to ranks among top 1%, top 5% (excluding top 1%), top 1/3 (excluding top 5%) and remaining predictions, according to their confidence score.

Calculation of the integrative score

Finally, the integrative score (S) assigned to each miRNA–target interaction, was derived by applying noisy-or model using confidence scores (which by nature are statistical precisions) from the predictions obtained across all the datasets predicting a given interaction.

$$S_j = 1 - \prod_i (1 - s_{ij}), \quad (1)$$

where, s_{ij} denotes confidence score of the j-th miRNA–target interaction obtained from the i-th resource, with the product taken across all the resources predicting the given interaction.

Collection of experimentally validated miRNA–target interactions

We collated two sets of experimentally validated miRNA–gene interactions, obtained from TarBase v.7.0 (16) and NPinter v.3.0 (17). miRNA names and gene symbols were standardized as described above. We considered only miRNA–target interactions supported by wet-lab experimental evidence, excluding interactions supported by only computational results. Finally, we divided the collated data into benchmarking and validation set. For the validation set we selected all the miRNA–target interactions supported by at least one reporter assay, while for the benchmarking dataset we considered all the remaining interactions if supported by at least two independent experiments. For more details about experimental evidence types, please see Supplementary Material. The resulting benchmarking dataset contains 59 105 unique interactions between 7292 unique genes and 828 unique miRNAs; the validation set contains 1359 unique interactions between 762 unique genes and 259 unique miRNAs.

Validation of the miRNA–target predictions

There are two main challenges associated with proper evaluation of accuracy of the miRNA–target prediction resources. The first challenge stems from the lack of sufficient number of true negatives, i.e. interactions that we know do not occur under normal conditions. This lack prevents the use of standard measures for predictive performance evaluation, such as statistical accuracy, or area under the receiver operating characteristic curve. Although some authors rely on negative evidence provided by curated resources such as TarBase v.7.0 (16), we believe that the negative evidence may be circumstantial, affected by the various experimental factors, and should be perceived as absence of evidence, but not as evidence of absence.

Another major challenge stems from the fact that some of the resources provide predictions that were filtered by the authors, while others provide the full set of predictions obtained. Consequently, predictive performance of the individual resources cannot be compared by simply assessing prediction sets in their fullest extent, since in this way, obtained results will likely be biased, favoring predictions that were pre-filtered by authors.

To address these challenges, we developed the following approach for assessing the predictive performance of the the individual resources. We first defined the set of values H, taken from the interval $\langle 10^3; 10^6 \rangle$ with exponential sampling (step size equal to 0.25). For each resource and for each $h \in H$, we then selected subset of top h predictions according to their rank, and calculated the balanced F-score (also known as F_1-score) (18,19) with respect to the validation dataset. The maximum F-scores obtained across H, served as a measure quantifying the best achievable predictive performance of the given resource (analogous to optimized accuracy of the binary classification). We also use area under the curve delineated by the $F^i(h)$ dependence, as a measure quantifying overall predictive capacity of the given resource.

Evaluation of overlaps between predictions from the individual resources

To evaluate the overlap between predictions obtained from any two resources, we used Jaccard index, calculated across subset of top h predictions from each of these two resources, according to their rank. Value of h was selected from previously defined set of values H, in order to maximize the resulting Jaccard index.

Evaluation of prediction bias of the individual resources

For each resource, we took a subset comprising the top

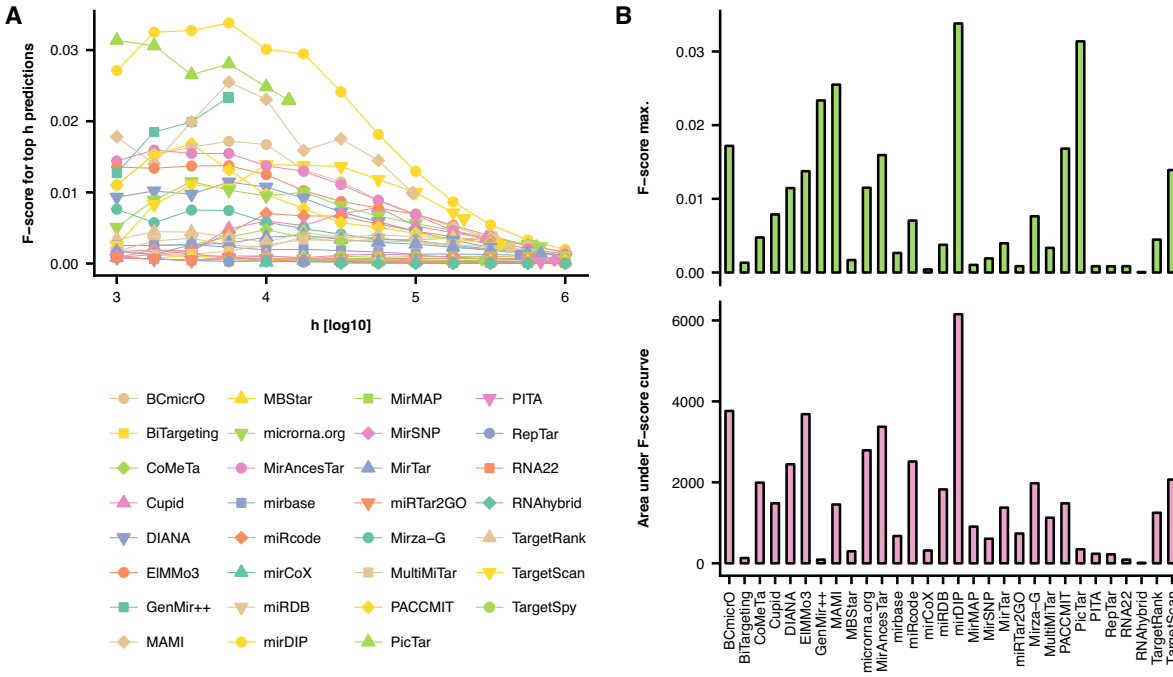

Figure 1. Validation of the predictions obtained across resources, along those derived from mirDIP's integrative score (denoted as mirDIP): (**A**) F-scores calculated using top *h* predictions from individual resources, with respected to validation set of experimental evidence. (**B**) Barplots depicting maximum values of F-score as obtained across expanding value of *h* (top) and area under F-score curves (see (A)), as calculated for individual resources.

10^4 predictions, from which we extracted the list of miRNAs involved. We then randomly picked 10 unique miRNAs, and from the selected predictions we extracted the list of their target genes. Using functions `enrichGO` and `enrichKEGG` from the Bioconductor package `clustProfiler v.3.4.4` (20), we calculated significance of over-representation (enrichment) of the extracted genes across biological processes as defined by Gene Ontology (GO) and biological pathways as defined by KEGG. For each resource we repeated the above steps 10^3 times, each time starting with an independent random pick of the 10 unique miRNAs. The bias of the given resource toward biological processes was measured by the total number of processes which were found to be significantly enriched ($P < 0.01$) more than 10 times (0.01×1000). Analogously, we quantified the bias of the resources toward biological pathways.

RESULTS

Characteristics of the mirDIP data collection

We summarized 75 resources providing computational predictions on human miRNA–target interactions (see Table 1). Although individual resources differ largely by the methods they employ for predictions, we recognized seven basic features that can characterize individual approaches. These include: (i) assessment of evolutionary conservation of putative binding region; (ii) target sequence analysis—referring to methods considering explicit properties of putative binding region, such as miRNA sequence complementarity, G–C content, accessibility to RISC complex, etc.; (iii) calculation of binding energy between miRNA and its putative target sequence; (iv) use of miRNA/mRNA expression profiles; (v) use of cross-linking immunoprecipitation (CLIP) data; (vi) use of machine learning methods; (vii) integrativeness—use of prior predictions obtained from other miRNA–target prediction tools.

From 75 resources we reviewed, we selected 30, from which we were able to retrieve 45 different sets of predictions. Each of the prediction sets was first pre-processed, then standardized, normalized and finally benchmarked. As described in the 'Materials and Methods' section, we then reduced the number of prediction sets to 30, so for each resource we retained only the set giving the most precise predictions. In total we compiled 151 869 821 predictions for 48 657 133 unique miRNA–target interactions, comprising 2586 unique miRNAs and 27 667 unique genes. The total number of predictions from the individual resources along the number of miRNAs and genes covered are summarized in Table 2.

We investigated the overlap of predictions across individual resources. We found that individual resources overlap only mildly, as even the greatest overlap detected (BCmicro–TargetRank), results in a Jaccard index equal to only 0.3, and its average is only 0.06 (Supplementary Figure 1). Clustering the resources using the Jaccard distance (1-Jaccard index) shows that even resources using very similar methodologies (as characterized by seven selected features) differ substantially. This low overlap indicates that miRNA–target predictions are heavily dependent on the underlying data used by the given methodology, its parametrization, or other details, and stresses the need for integrative approaches such as the one we present here.

Table 1. Table summarizing publicly available human miRNA–target prediction resources

Name	Publication Year	Last Updated	Evolutionary Conservation	Sequence Analysis	Binding energy	Expression	CLIP	Machine learning	Integrative	Ref.*
Avishkar	2015	2017	yes	yes	yes		yes	yes		(21)
BCmicrO	2012							yes	yes	(22)
BiTargeting	2010			yes	yes					(23)
ChemiRs	2016								yes	(24)
chimiRic	2016			yes	yes		yes	yes		(25)
CoMeTa	2012					yes			yes	(26)
comiR	2013	2015				yes		yes	yes	(27)
CUDA-miRanda	2014		yes	yes	yes					(28)
Cupid	2011	2015				yes		yes	yes	(29)
DIANA	2012		yes	yes	yes	yes	yes	yes	yes	(30)
doRiNA	2014						yes		yes	(31)
ElMMo3	2007		yes							(32)
GenMir++	2007					yes			yes	(33)
Hoctar	2011					yes			yes	(34)
HomoloMTI		2011	yes						yes	homolomti.mbc.nctu.edu.tw
HomoTarget	2012			yes	yes			yes		(35)
MAMI		2006							yes	mami.med.harvard.edu/
MBStar	2015						yes	yes		(36)
mESAdb	2011					yes			yes	(37)
MicroInspector	2005			yes	yes					(38)
microrna.org	2008	2010	yes	yes	yes	yes		yes	yes	(39)
MicroTar	2006	2008		yes	yes					(40)
mimiRNA	2009					yes			yes	(41)
MirAncesTar	2017		yes					yes	yes	(42)
mirbase	2006	2014		yes	yes					(43)
miRcode	2012		yes	yes						(44)
mirConnX	2011					yes			yes	(45)
mirCoX	2013					yes			yes	(46)
miRDB	2015	2016		yes	yes		yes	yes		(47)
mirDIP	2011	2017							yes	(9)
miRecords	2009	2013							yes	(48)
miREE	2011			yes	yes	yes	yes	yes	yes	(49)
miRGate	2015								yes	(50)
miRGator	2008	2013				yes			yes	(51)
MirMAP	2012	2013	yes	yes	yes		yes			(52)
miRNALasso	2015			yes	yes	yes				(53)
miRNAmap	2006	2007				yes			yes	(54)
miRNA_targets	2012								yes	(55)
miRó	2009								yes	(56)
MiRonTop	2010					yes			yes	(57)
miRror	2010								yes	(58)
MirSNP	2012			yes	yes					(59)
miRSystem	2012	2016							yes	(60)
MirTar	2008	2014		yes	yes	yes		yes		(61)
miRTar2GO	2016			yes	yes		yes	yes		(62)
mirTarPri	2013								yes	(63)
miRTarVis	2015					yes		yes	yes	(64)
mirWalk	2011	2017							yes	(65)
Mirza-G	2015		yes	yes	yes	yes		yes		(66)
miSTAR	2016			yes				yes		(67)
miTarget	2006			yes	yes			yes	yes	(68)
MMIA	2009					yes			yes	(69)
MultiMiTar	2011	2014		yes				yes		(70)
NbmiRTar	2007							yes	yes	(71)
PACCMIT	2012	2013	yes	yes	yes					(72)
PicTar	2005	2007		yes	yes					(73)
PITA	2007	2008		yes	yes					(74)
RAIN	2017							yes	yes	(15)
RepTar	2011			yes	yes					(75)
RegNetwork	2015	2017							yes	(76)
RNA22	2006	2015		yes						(77)
RNAhybrid	2004	2006		yes	yes					(78)
STarMir	2014						yes	yes	yes	(79)
SVMicro	2010			yes				yes		(80)
Talasso	2012					yes		yes	yes	(81)
TargetExpress	2016					yes		yes	yes	(82)
targetHub	2013								yes	(83)
TargetMiner	2009	2012	yes	yes	yes	yes		yes	yes	(84)
TargetRank	2007		yes	yes		yes				(85)
TargetScan	2003	2015	yes	yes				yes		(86)
TargetScore	2014	2017				yes		yes	yes	(87)
TargetSpy	2010		yes	yes				yes		(88)
TargetThermo	2011				yes				yes	(89)
Tools4miRs	2016	2017							yes	(90)
ToppMiR	2014							yes	yes	(91)

*We cite the publication as indicated by the resource website. If not available, we cite the most relevant publication for the resource.
Resources integrated within mirDIP are highlighted in bold.

Table 2. Number of predictions, genes and miRNAs, as obtained from individual resources

Resource	Version/Date	Predictions	Genes	miRNAs
BCmicrO	March, 2017	10 682 301	18 418	580
BiTargeting	April, 2017	5 314 760	18 517	2582
CoMeTa	March, 2017	640 586	10 969	643
Cupid	March, 2017	298 163	8411	1181
DIANA	v5.0	7 112 061	18 529	1909
ElMMo3	March, 2017	2 837 861	18 179	997
GenMir++	March, 2017	5579	872	99
MAMI	March, 2017	95 408	14 285	309
MBStar	April, 2017	11 925 118	18 041	2031
microrna.org	January, 2008	684 192	18 424	241
MirAncesTar	March, 2017	36 116 591	18 532	2568
mirbase	March, 2017	498 128	17 913	684
miRcode	March, 2017	997 836	25 656	124
mirCoX	March, 2017	1 716 865	21 749	79
miRDB	v5.0	4 739 198	16 588	2571
MirMAP	v.1.1	11 392 502	18 574	2031
MirSNP	March, 2017	849 897	17 180	1909
MirTar	March, 2017	686 222	16 556	1897
miRTar2GO	March, 2017	1 164 371	10 890	366
Mirza-G	April, 2016	4 348 927	16 790	2564
MultiMiTar	March, 2017	429 258	10 986	473
PACCMIT	February, 2012	363 717	11 735	1905
PicTar	March, 2017	14 160	2430	114
PITA	v6.0	685 848	18 141	295
RepTar	March, 2017	2 996 265	17 280	1066
RNA22	v.2.0	3 127 672	1927	2584
RNAhybrid	v2.1.2	41 306 832	17 448	2584
TargetRank	March, 2017	342 703	14 241	525
Targetscan	v7.1	210 146	11 952	369
TargetSpy	April, 2016	286 654	15 485	356

We benchmarked all predictions by assigning a score and quantifying the confidence of the given prediction with respect to currently available experimental evidence. The assigned confidence score allows direct quantitative comparison of the predictions across resources. To provide more intuitive rating of the prediction confidence, predictions were subsequently categorized into four classes (referred to as 'confidence classes'), according to the assigned confidence score. Comparison of individual resources in terms of the frequency of predictions of the given confidence class (Supplementary Figure 2) revealed that methodologies relying solely on target sequence properties and binding energy (e.g., BiTargeting, RNAhybrid and RNA22), yield less confident predictions compared to more advanced tools.

By statistical inference, we then derived an integrative score assigned to each miRNA–target interaction, based on the individual predictions obtained across the resources. The resulting integrative score approximates a power-law distribution (Supplementary Figure 3), with mean 0.05 and median 0.02. Similarly, the number of predictions obtained for a given miRNA–target interaction across the resources approximates a power-law distribution (Supplementary Figure 2), with mean 3.1 and median 2.

Validation and assessment of the prediction bias

Next, we validated predictions obtained from individual resources against the collated experimental evidence. To do this, we used F-score as a measure of prediction performance, calculated across *h* top-ranked predictions from the given resource, where *h* was increasing exponentially from 10^3 to 10^6. For each resource we recorded the maximum F-score achieved along the *h* and the area under the resulting F-score dependence. Resulting F-scores were compared to the ones obtained from the predictions, as derived from mirDIP integrative score.

Resulting maximum F-scores as well as the area under the F-score curves (Figure 1) show that predictions derived from the integrative score (mirDIP) are more accurate compared to predictions from other resources we integrated here. Interestingly, some resources such as PicTar, or GenMir++ ranked among the top performing tools, as measured by maximum F-score, but ranked among the worst according to area under F-score curve. This is because these resources provide accurate predictions, but only for a small subset of genes and miRNAs.

We then examined the prediction bias of the individual resources, along the bias of the predictions derived from the integrative score (mirDIP). For each resource, we separately evaluated the bias toward GO biological processes and KEGG pathways, quantified as total number of falsely enriched processes and pathways, respectively (Figure 2). We did not keep record of which particular processes/pathways are subject to bias from individual resources. For each resource, we then calculated overall bias, as a sum of the two biases. Using linear regression we delineated a trend line to show dependence between resource predictive capacity and bias (log10-transformed).

We found significant positive relationship between the predictive capacity and the overall bias of the resource (Spearman's $\rho = 0.62$; $p = 2E\text{-}4$). As expected, tools adopting simpler methodologies, relying solely on sequence analysis and thermodynamic measures, such as PITA, RepTar

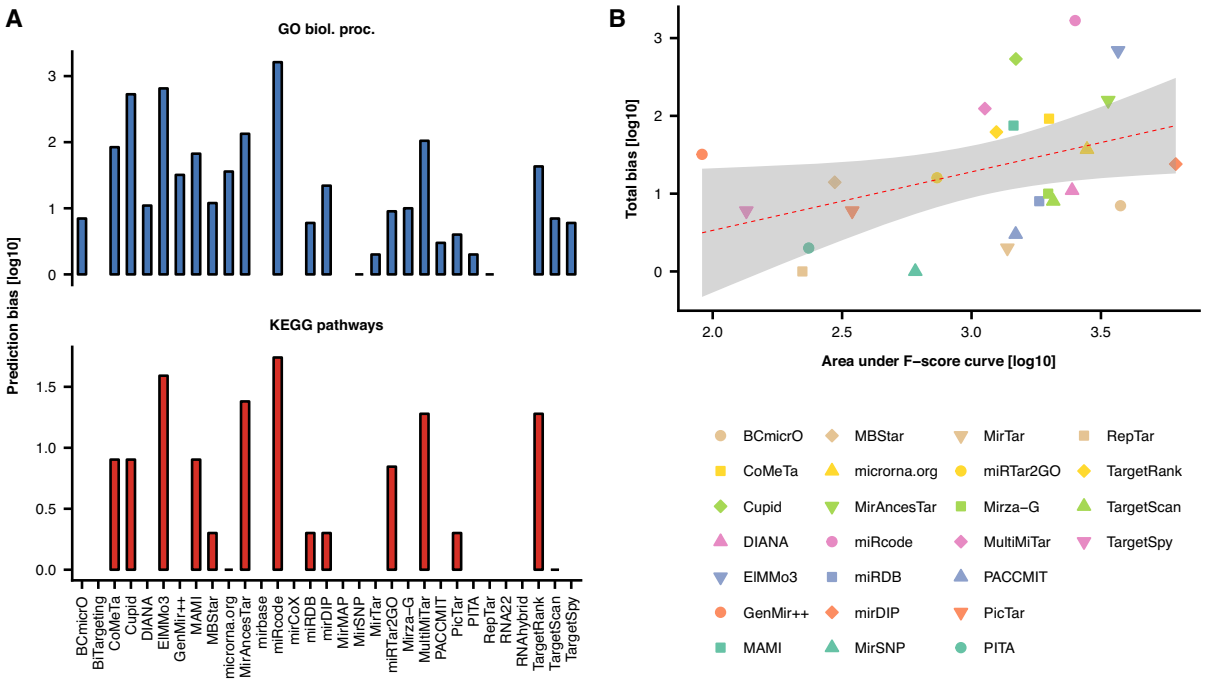

Figure 2. (**A**) Assessment of the prediction bias toward GO biological processes (top) and KEGG biological pathways (bottom). (**B**) Dependence between total bias and prediction accuracy quantified by area under F-score curves (see Figure 1). Regression line (red dashed line) denotes dependence of prediction bias on prediction accuracy and gray area highlights the 95% confidence interval of the regression.

hsa-let-7a-5p	hsa-let-7b-5p	hsa-let-7c-5p	hsa-let-7d-5p	hsa-let-7e-5p	hsa-let-7f-5p	hsa-let-7g-5p	hsa-let-7i-5p	hsa-miR-98-5p	hsa-let-7a-3p	hsa-let-7b-3p	hsa-let-7g-3p	hsa-let-7i-3p	hsa-miR-98-3p
BZW1	BZW1	BZW1	BZW1	BZW1	BZW1	COL1A2	COL1A2	BZW1	YY1	PUM2	CDYL	BDNF	YY1
COL1A2	COL1A2	COL1A2	COL1A2	COL1A2	COL1A2	BZW1	BZW1	HMGA2	IER5	YY1	BACH2		
DUSP1	HMGA2	HMGA2	HMGA2	ZFYVE26	HMGA2	HMGA2	HMGA2	COL1A2	MAST4	PTCH1	FLRT3		
HMGA2	DUSP1	DUSP1	EIF4G2	DUSP1	DUSP1	ARID3B	ARID3B	ARID3B	RAB2A	RAB2A	CAMTA1		
ARID3B	ARID3B	ARID3B	BACH1	EIF4G2	ARID3B	DUSP1	MAP4K3	MAP4K3	NR3C1	JAG1	GRM7		
KPNA4	KPNA4	MAP4K3	SMARCAD1	TGFBR1	MAP4K3	SMARCAD1	BACH1	PBX3	NR5A2	IER5	KCNJ6		
EIF4G2	MAP4K3	KPNA4	USP38	MAP4K3	BACH1	LRIG2	TGFBR1	BACH1	PTCH1	NR5A2	STXBP5L		
MAP4K3	LRIG2	LRIG2	LRIG2	CDC34	LRIG2	BACH1	IGDCC3	LRIG2	NEUROG2	DOCK1	FMR1		
BACH1	BACH1	BACH1	DMD	LRIG2	EIF4G2	TGFBR1	DUSP1	TGFBR1	ARHGAP20	REV3L	NCOA6		
LRIG2	TGFBR1	EIF4G2	AKAP6	HMGA2	SMARCAD1	GALNT1	LRIG2	GALNT1	WWC1	NR3C1	MYCBP2		

Figure 3. Common targets of members of let-7 family. Shown are only top 10 targets (if available) for each miRNA according to mirDIP integrated score.

RNA22 and RNAhybrid, exhibited little to no bias. Importantly, the bias of the predictions derived from the integrative score is moderate compared to other resources, providing a favorable trade-off between precision and associated bias.

mirDIP description

We identified two major use-cases for computational miRNA–target predictions. The first identifies downstream gene targets (or upstream miRNA regulators) of selected miRNAs (or genes of interest). This mode is used to generate hypotheses and prioritize miRNAs or genes for further functional studies. The second is to provide confirmation of the miRNA–gene associations derived from prior experiments.

mirDIP v4.1 was designed to facilitate these two tasks, functioning in two distinct modes, which we refer to as 'unidirectional' and 'bidirectional' search, provided as separate tabs at http://ophid.utoronto.ca/mirDIP/.

Unidirectional mode:. In the unidirectional mode, the user is required to specify either miRNAs or genes of interest. mirDIP will search for all the predictions involving the input miRNAs, or input genes. The search is conducted as an exact, case sensitive string match, requiring precise miRNA names or gene symbols to be entered. The search can be restricted to only interactions whose integrative score exceeds the user-specified threshold. Results consist of target gene symbols, their uniprot IDs, miRNA names, integrated score, number of independent predictions supporting given interactions, and score class. Results are ordered by integrative score and can be downloaded either as a comma- or tab-separated file. In the unidirectional mode, mirDIP allows results to be summarized in a wide format in the miRNA–gene matrix tab, where miRNA–gene interactions are represented by an adjacency matrix, the input miRNAs/genes

being its columns, and resulting interaction partners being its rows.

Bidirectional mode:. In the bidirectional mode, mirDIP requires the user to specify both miRNAs and genes. mirDIP will search for all the predictions associating any of the input miRNAs with any of the input genes. The search can be restricted to interactions confirmed by at least k number of predictions, where k can be any value from 1 to 30, with 1 as default value. Similarly, the search can be restricted by confidence class, where 'Very high confidence' is set as default.

The resulting output is a table listing the standardized target gene symbols, their uniprot IDs, standardized upstream miRNA names, non-standardized gene symbols and miRNA names as originally obtained from the given resource, rank with which the given interaction was predicted by the given resource, name of the resource from which the given prediction was obtained, confidence score and confidence class. Results are ordered by gene symbols and miRNA names and can be downloaded either as a comma- or tab-separated file.

In the unidirectional mode mirDIP utilizes the integrative score to provide prioritization of the potential target genes or upstream miRNAs—which, as we showed, is more accurate than the use of any of the resources alone. In the bidirectional mode, mirDIP provides the summary of the computational predictions of interactions between the given miRNAs and genes, across 30 different resources.

All the underlying data are available for download as a flat file in a tab-delimited format.

Working example

To demonstrate mirDIP workflow, we used the unidirectional search, under 'very high confidence' filter, to identify targets of the let-7 miRNA family (results obtained, along with the list of the input miRNAs, are provided in Supplementary Data 1). Let-7 was one of the first miRNAs discovered in *Caenorhabditis elegans* and its members are highly conserved across various species (92). Their role has been well studied in the context of development and of several cancer types (93).

For further analyses, we considered only the top 10 targets for each miRNA. As we found, YY1 is targeted by 3 out of 5 -3p miRNAs; DUSP1, ARIDB3, MAP4K3 are common targets of 7 out of 9 -5p miRNAs, BACH1 is a common target of 8 out of 9 -5p miRNAs and BZW1, COL1A2, HMGA2 and LRIG2 are common targets of all -5p miRNAs (Figure 3). Interestingly, COL1A2, HMGA2, ARIDB3, MAP4K3 and BACH1 have been extensively validated as targets of let-7 family members (94–101).

As BZW1 is the top target for seven miRNAs, a researcher might be interested in how the predictions for pairs let-7x-5p–BZW1 are distributed across sources. For this reason we performed a bidirectional search applying 'very high confidence' filter (results of which are provided in Supplementary Data 2).

Biological validation of this finding is beyond the scope of this paper, but it is evident that the integration and prioritization performed in mirDIP highlight targets that are well supported by the literature.

DISCUSSION

mirDIP integrates human miRNA–target predictions across 30 resources. It stores 151.9 million predictions, while covering 2586 of the 2588 known human mature miRNAs (miRBase v 21.), and 81.3% of the 34 010 human genes (excluding genes of non-coding RNAs, HGNC April 2017). When compared to similar resources (Table 3), mirDIP integrates more than twice as many prediction tools than miRror and over 60 million more predictions than miRGate, the two largest integrative resources according to the number of integrated tools and predictions, respectively.

mirDIP supports the two most frequent miRNA-related workflows. The first is to summarize currently available predictions that support hypothesized interactions between specified lists of genes and miRNAs (bidirectional search). The second is to identify plausible miRNA targets, or gene's miRNA regulators (unidirectional search).

mirDIP utilizes the integrated score to prioritize predicted interactions of the input miRNAs/genes. The integrated score was statistically inferred from resource specific measures of prediction confidence. As we have shown here, mirDIP's integrated score provides more accurate predictions (as measured by F-score) of miRNA–target interactions than those obtained from the individual resources. Among the integrative resources, only miRror and RAIN provide integrative scores similar to mirDIP. In contrast to mirDIP, miRror does not provide one-to-one miRNA–target predictions and does not allow their bulk download, preventing any comparison with mirDIP. RAIN provides an integrative score that was inferred from benchmarked scores obtained from predictions across various resources using a similar methodology as used in mirDIP. However, RAIN's integrative score is heavily affected by integration of experimental data and curated literature, making it unsuitable for comparison with mirDIP's integrative score, which is derived solely from computational predictions.

MiRNA–target predictions tend to be biased toward certain biological processes and pathways (10,11). Since the integration of the predictions may lead to bias cumulation, we inspected the bias of the individual resources along the bias of predictions derived from the integrative score. Importantly, predictions derived from the integrative score are not overly biased; instead they provide a better trade-off between bias and prediction accuracy than most of the resources we tested.

Individual prediction resources approach the problem of miRNA–target prediction on different level of details. Similarly, resources summarizing miRNA–target experimental evidence typically report only the respective target molecules, without specifying the precise binding location (depending on the experimental type, this is specially applicable to luciferase reporter assay and other low-throughput techniques). In order to integrate large number of prediction resources, and to utilize available experimental evidence, we had to restrict mirDIP to report only about miRNA target molecules, but not their target sites.

Currently, mirDIP requires inputs to be an exact match to the standardized miRNA names and gene symbols within the database. Due to the heterogeneity of nomenclature,

Table 3. Major integrative miRNA–target prediction resources

Name	Predictions	Interactions	Genes	miRNAs	Resources
chemiRs‡	NA	5 087 441	36 817	2588	9
mirDIP	151 869 821	48 657 133	27 667	2586	30
miRecords‡	NA	NA	NA	NA	11
miRGate‡	85 844 670	NA	20 805	2680	5
miRó‡	NA	NA	NA	NA	6
miRror‡	NA	NA	NA	NA	12
miRSystem‡	NA	2 128 551	NA	NA	6
mirWalk*‡	64 354 911	NA	45 727	2057	11
RAIN‡	NA	835 174	19 941	2571	5
RegNetwork‡	NA	197 331	19 719	1904	5
Tools4miRs	NA	NA	NA	NA	7

*Includes predictions for human, mouse and rat. ‡Includes experimental data.
Columns refer to the number of integrated predictions, unique miRNA–target interactions, genes, miRNA and number of integrated resources, respectively. NA indicates that the value is not available.

users may need to use non-standard terms; thus, we plan to provide a more advanced search, considering previous names/symbols, synonyms, as well as alternative identifiers, such as miRBase accession numbers, uniprot gene IDs, etc. We also intend to introduce tissue-specificity into miRNA–target predictions and extend mirDIP into a platform allowing the generation of *de novo* predictions for newly discovered miRNAs.

Altogether, mirDIP v4.1 provides a comprehensive, reliable and user-friendly resource for miRNA–target predictions, suitable for a wide range of users, even with minimal statistical or computational experience.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Krembil Foundation; Ontario Research Fund [#34876]; Canadian Cancer Society Research Institute [#704238]; Natural Sciences Research Council [#203475]; Canada Foundation for Innovation [#22540#30865]; Canada Research Chair Program (CRC) [#225404]; Ontario Research Fund [GL2-01-030]. Funding for open access charge: CRC Program [#225404].
Conflict of interest statement. None declared.

REFERENCES

1. Huntzinger,E. and Izaurralde,E. (2011) Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat. Rev. Genet.*, **12**, 99–110.
2. Jonas,S. and Izaurralde,E. (2015) Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.*, **16**, 421–433.
3. Gorski,S.A., Vogel,J. and Doudna,J.A. (2017) RNA-based recognition and targeting: sowing the seeds of specificity. *Nat. Rev. Mol. Cell Biol.*, **18**, 215–228.
4. Valinezhad Orang,A., Safaralizadeh,R. and Kazemzadeh-Bavili,M. (2014) Mechanisms of miRNA-mediated gene regulation from common downregulation to mRNA-specific upregulation. *Int. J. Genomics*, **2014**, 970607.
5. Ha,M. and Kim,V.N. (2014) Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.*, **15**, 509–524.
6. Di Leva,G., Garofalo,M. and Croce,C.M. (2014) MicroRNAs in cancer. *Annu. Rev. Pathol.*, **9**, 287–314.
7. Qu,Z., Li,W. and Fu,B. (2014) MicroRNAs in autoimmune diseases. *BioMed. Res. Int.*, **2014**, 527895.
8. Hommers,L.G., Domschke,K. and Deckert,J. (2015) Heterogeneity and individuality: microRNAs in mental disorders. *J. Neural Transm.*, **122**, 79–97.
9. Shirdel,E.A., Xie,W., Mak,T.W. and Jurisica,I. (2011) NAViGaTing the micronome–using multiple microRNA prediction databases to identify signalling pathway-associated microRNAs. *PLoS One*, **6**, e17429.
10. Bleazard,T., Lamb,J.A. and Griffiths-Jones,S. (2015) Bias in microRNA functional enrichment analysis. *Bioinformatics*, **31**, 1592–1598.
11. Godard,P. and van Eyll,J. (2015) Pathway analysis from lists of microRNAs: common pitfalls and alternative strategy. *Nucleic Acids Res.*, **43**, 3490–3497.
12. Haunsberger,S.J., Connolly,N.M. and Prehn,J.H. (2016) miRNAmeConverter: an R/bioconductor package for translating mature miRNA names to different miRBase versions. *Bioinformatics*, **33**, 592–593.
13. Zhang,J., Le,T.D., Liu,L. and Li,J. (2017) Identifying miRNA sponge modules using biclustering and regulatory scores. *BMC Bioinformatics*, **18**, 44.
14. Salmena,L., Poliseno,L., Tay,Y., Kats,L. and Pandolfi,P.P. (2011) A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*, **146**, 353–358.
15. Junge,A., Refsgaard,J.C., Garde,C., Pan,X., Santos,A., Alkan,F., Anthon,C., von Mering,C., Workman,C.T., Jensen,L.J. *et al.* (2017) RAIN: RNA–protein Association and Interaction Networks. *Database*, **2017**, baw167.
16. Sethupathy,P., Corda,B. and Hatzigeorgiou,A.G. (2006) TarBase: A comprehensive database of experimentally supported animal microRNA targets. *RNA*, **12**, 192–197.
17. Hao,Y., Wu,W., Li,H., Yuan,J., Luo,J., Zhao,Y. and Chen,R. (2016) NPInter v3. 0: an upgraded database of noncoding RNA-associated interactions. *Database*, **2016**, baw057.
18. Dice,L.R. (1945) Measures of the amount of ecologic association between species. *Ecology*, **26**, 297–302.
19. Sørensen,T. (1948) A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Biol. Skr.*, **5**, 1–34.
20. Yu,G., Wang,L.-G., Han,Y. and He,Q.-Y. (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics*, **16**, 284–287.
21. Ghoshal,A., Shankar,R., Bagchi,S., Grama,A. and Chaterji,S. (2015) MicroRNA target prediction using thermodynamic and sequence curves. *BMC Genomics*, **16**, 999.
22. Yue,D., Guo,M., Chen,Y. and Huang,Y. (2012) A Bayesian decision fusion approach for microRNA target prediction. *BMC Genomics*, **13**, S13.
23. Veksler-Lublinsky,I., Shemer-Avni,Y., Kedem,K. and Ziv-Ukelson,M. (2010) Gene bi-targeting by viral and human miRNAs. *BMC Bioinformatics*, **11**, 249.

24. Su,E. C.-Y., Chen,Y.-S., Tien,Y.-C., Liu,J., Ho,B.-C., Yu,S.-L. and Singh,S. (2016) ChemiRs: a web application for microRNAs and chemicals. *BMC Bioinformatics*, **17**, 167.
25. Lu,Y. and Leslie,C.S. (2016) Learning to predict miRNA-mRNA interactions from AGO CLIP sequencing and CLASH data. *PLoS Comput. Biol.*, **12**, e1005026.
26. Gennarino,V.A., D'Angelo,G., Dharmalingam,G., Fernandez,S., Russolillo,G., Sanges,R., Mutarelli,M., Belcastro,V., Ballabio,A., Verde,P. *et al.* (2012) Identification of microRNA-regulated gene networks by expression analysis of target genes. *Genome Res.*, **22**, 1163–1172.
27. Coronnello,C. and Benos,P.V. (2013) ComiR: combinatorial microRNA target prediction tool. *Nucleic Acids Res.*, **41**, W159–W164.
28. Wang,S., Kim,J., Jiang,X., Brunner,S.F. and Ohno-Machado,L. (2014) GAMUT: GPU accelerated microRNA analysis to uncover target genes through CUDA-miRanda. *BMC Medi. Genomics*, **7**, S9.
29. Chiu,H.-S., Llobet-Navas,D., Yang,X., Chung,W.-J., Ambesi-Impiombato,A., Iyer,A., Kim,H.R., Seviour,E.G., Luo,Z., Sehgal,V. *et al.* (2015) Cupid: simultaneous reconstruction of microRNA-target and ceRNA networks. *Genome Res.*, **25**, 257–267.
30. 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.
31. Blin,K., Dieterich,C., Wurmus,R., Rajewsky,N., Landthaler,M. and Akalin,A. (2015) DoRiNA 2.0–upgrading the doRiNA database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res.*, **43**, D160–D167.
32. 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.
33. Huang,J.C., Babak,T., Corson,T.W., Chua,G., Khan,S., Gallie,B.L., Hughes,T.R., Blencowe,B.J., Frey,B.J. and Morris,Q.D. (2007) Using expression profiling data to identify human microRNA targets. *Nat. Methods*, **4**, 1045–1049.
34. Gennarino,V.A., Sardiello,M., Mutarelli,M., Dharmalingam,G., Maselli,V., Lago,G. and Banfi,S. (2011) HOCTAR database: a unique resource for microRNA target prediction. *Gene*, **480**, 51–58.
35. Ahmadi,H., Ahmadi,A., Azimzadeh-Jamalkandi,S., Shoorehdeli,M.A., Salehzadeh-Yazdi,A., Bidkhori,G. and Masoudi-Nejad,A. (2013) HomoTarget: a new algorithm for prediction of microRNA targets in Homo sapiens. *Genomics*, **101**, 94–100.
36. Bandyopadhyay,S., Ghosh,D., Mitra,R. and Zhao,Z. (2015) MBSTAR: multiple instance learning for predicting specific functional binding sites in microRNA targets. *Sci. Rep.*, **5**, 8004.
37. Kaya,K.D., Karakülah,G., Yakıcıer,C.M., Acar,A.C. and Konu,Ö. (2011) mESAdb: microRNA expression and sequence analysis database. *Nucleic Acids Res.*, **39**(Suppl. 1), D170–D180.
38. Rusinov,V., Baev,V., Minkov,I.N. and Tabler,M. (2005) MicroInspector: a web tool for detection of miRNA binding sites in an RNA sequence. *Nucleic Acids Res.*, **33**(Suppl. 2), W696–W700.
39. Betel,D., Koppal,A., Agius,P., Sander,C. and Leslie,C. (2010) Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.*, **11**, R90.
40. Thadani,R. and Tammi,M.T. (2006) MicroTar: predicting microRNA targets from RNA duplexes. *BMC Bioinformatics*, **7**, S20.
41. Ritchie,W., Flamant,S. and Rasko,J.E.J. (2010) mimiRNA: a microRNA expression profiler and classification resource designed to identify functional correlations between microRNAs and their targets. *Bioinformatics*, **26**, 223–227.
42. Leclercq,M., Diallo,A.B. and Blanchette,M. (2017) Prediction of human miRNA target genes using computationally reconstructed ancestral mammalian sequences. *Nucleic Acids Res.*, **45**, 556–566.
43. Kozomara,A. and Griffiths-Jones,S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, D68–D73.
44. Jeggari,A., Marks,D.S. and Larsson,E. (2012) miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics*, **28**, 2062–2063.
45. Huang,G.T., Athanassiou,C. and Benos,P.V. (2011) mirConnX: condition-specific mRNA-microRNA network integrator. *Nucleic Acids Res.*, **39**(Suppl. 2), W416–W423.
46. Giles,C.B., Girija-Devi,R., Dozmorov,M.G. and Wren,J.D. (2013) mirCoX: a database of miRNA-mRNA expression correlations derived from RNA-seq meta-analysis. *BMC Bioinformatics*, **14**, S17.
47. Wong,N. and Wang,X. (2015) miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.*, **43**, D146–D152.
48. Xiao,F., Zuo,Z., Cai,G., Kang,S., Gao,X. and Li,T. (2008) miRecords: an integrated resource for microRNA–target interactions. *Nucleic Acids Res.*, **37**(Suppl. 1), D105–D110.
49. Reyes-Herrera,P.H., Ficarra,E., Acquaviva,A. and Macii,E. (2011) miREE: miRNA recognition elements ensemble. *BMC Bioinformatics*, **12**, 454.
50. Andrés-León,E., González Peña,D., Gómez-López,G. and Pisano,D.G. (2015) miRGate: a curated database of human, mouse and rat miRNA–mRNA targets. *Database*, **2015**, bav035.
51. Cho,S., Jang,I., Jun,Y., Yoon,S., Ko,M., Kwon,Y., Choi,I., Chang,H., Ryu,D., Lee,B. *et al.* (2013) MiRGator v3. 0: a microRNA portal for deep sequencing, expression profiling and mRNA targeting. *Nucleic Acids Res.*, **41**, D252–D257.
52. Vejnar,C.E. and Zdobnov,E.M. (2012) MiRmap: comprehensive prediction of microRNA target repression strength. *Nucleic Acids Res.*, **40**, 11673–11683.
53. Wang,Z., Xu,W. and Liu,Y. (2015) Integrating full spectrum of sequence features into predicting functional microRNA–mRNA interactions. *Bioinformatics*, **31**, 3529–3536.
54. Hsu,S.-D., Chu,C.-H., Tsou,A.-P., Chen,S.-J., Chen,H.-C., Hsu,P.W.-C., Wong,Y.-H., Chen,Y.-H., Chen,G.-H. and Huang,H.-D. (2007) miRNAMap 2.0: genomic maps of microRNAs in metazoan genomes. *Nucleic Acids Res.*, **36**(Suppl. 1), D165–D169.
55. Kumar,A., Wong,A. K.-L., Tizard,M.L., Moore,R.J. and Lefèvre,C. (2012) miRNA_Targets: a database for miRNA target predictions in coding and non-coding regions of mRNAs. *Genomics*, **100**, 352–356.
56. Laganà,A., Forte,S., Giudice,A., Arena,M., Puglisi,P.L., Giugno,R., Pulvirenti,A., Shasha,D. and Ferro,A. (2009) miRo: a miRNA knowledge base. *Database*, **2009**, bap008.
57. Le Brigand,K., Robbe-Sermesant,K., Mari,B. and Barbry,P. (2010) MiRonTop: mining microRNAs targets across large scale gene expression studies. *Bioinformatics*, **26**, 3131–3132.
58. Friedman,Y., Naamati,G. and Linial,M. (2010) MiRror: a combinatorial analysis web tool for ensembles of microRNAs and their targets. *Bioinformatics*, **26**, 1920–1921.
59. Liu,C., Zhang,F., Li,T., Lu,M., Wang,L., Yue,W. and Zhang,D. (2012) MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs. *BMC Genomics*, **13**, 661.
60. Lu,T.-P., Lee,C.-Y., Tsai,M.-H., Chiu,Y.-C., Hsiao,C.K., Lai,L.-C. and Chuang,E.Y. (2012) miRSystem: an integrated system for characterizing enriched functions and pathways of microRNA targets. *PLoS One*, **7**, e42390.
61. Hsu,J. B.-K., Chiu,C.-M., Hsu,S.-D., Huang,W.-Y., Chien,C.-H., Lee,T.-Y. and Huang,H.-D. (2011) miRTar: an integrated system for identifying miRNA-target interactions in human. *BMC Bioinformatics*, **12**, 300.
62. Ahadi,A., Sablok,G. and Hutvagner,G. (2017) 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. *Nucleic Acids Res.*, **45**, e42.
63. Wang,P., Ning,S., Wang,Q., Li,R., Ye,J., Zhao,Z., Li,Y., Huang,T. and Li,X. (2013) mirTarPri: improved prioritization of microRNA targets through incorporation of functional genomics data. *PLoS One*, **8**, e53685.
64. Jung,D., Kim,B., Freishtat,R.J., Giri,M., Hoffman,E. and Seo,J. (2015) miRTarVis: an interactive visual analysis tool for microRNA-mRNA expression profile data. *BMC Proc. BioMed Central*, **9**, S2.
65. Dweep,H. and Gretz,N. (2015) miRWalk2. 0: a comprehensive atlas of microRNA-target interactions. *Nat. Methods*, **12**, 697–697.
66. Gumienny,R. and Zavolan,M. (2015) Accurate transcriptome-wide prediction of microRNA targets and small interfering RNA off-targets with MIRZA-G. *Nucleic Acids Res.*, **43**, 1380–1391.

67. Van Peer,G., De Paepe,A., Stock,M., Anckaert,J., Volders,P.-J., Vandesompele,J., De Baets,B. and Waegeman,W. (2017) miSTAR: miRNA target prediction through modeling quantitative and qualitative miRNA binding site information in a stacked model structure. *Nucleic Acids Res.*, **45**, e51.
68. Kim,S.-K., Nam,J.-W., Rhee,J.-K., Lee,W.-J. and Zhang,B.-T. (2006) miTarget: microRNA target gene prediction using a support vector machine. *BMC Bioinformatics*, **7**, 411.
69. Nam,S., Li,M., Choi,K., Balch,C., Kim,S. and Nephew,K.P. (2009) MicroRNA and mRNA integrated analysis (MMIA): a web tool for examining biological functions of microRNA expression. *Nucleic Acids Res.*, **37**(Suppl. 2), W356–W362.
70. Mitra,R. and Bandyopadhyay,S. (2011) MultiMiTar: a novel multi objective optimization based miRNA-target prediction method. *PLoS One*, **6**, e24583.
71. Yousef,M., Jung,S., Kossenkov,A.V., Showe,L.C. and Showe,M.K. (2007) Naive Bayes for microRNA target predictions–machine learning for microRNA targets. *Bioinformatics*, **23**, 2987–2992.
72. Marín,R.M., Voellmy,F., von Erlach,T. and Vaníček,J. (2012) Analysis of the accessibility of CLIP bound sites reveals that nucleation of the miRNA: mRNA pairing occurs preferentially at the 3′-end of the seed match. *RNA*, **18**, 1760–1770.
73. 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.
74. 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.
75. Elefant,N., Berger,A., Shein,H., Hofree,M., Margalit,H. and Altuvia,Y. (2010) RepTar: a database of predicted cellular targets of host and viral miRNAs. *Nucleic Acids Res.*, **39**(Suppl. 1), D188–D194.
76. Liu,Z.-P., Wu,C., Miao,H. and Wu,H. (2015) RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database*, **2015**, bav095.
77. Loher,P. and Rigoutsos,I. (2012) Interactive exploration of RNA22 microRNA target predictions. *Bioinformatics*, **28**, 3322–3323.
78. Rehmsmeier,M., Steffen,P., Höchsmann,M. and Giegerich,R. (2004) Fast and effective prediction of microRNA/target duplexes. *RNA*, **10**, 1507–1517.
79. Rennie,W., Liu,C., Carmack,C.S., Wolenc,A., Kanoria,S., Lu,J., Long,D. and Ding,Y. (2014) STarMir: a web server for prediction of microRNA binding sites. *Nucleic Acids Res.*, **42**, W114–W118.
80. Liu,H., Yue,D., Chen,Y., Gao,S.-J. and Huang,Y. (2010) Improving performance of mammalian microRNA target prediction. *BMC Bioinformatics*, **11**, 476.
81. Muniategui,A., Nogales-Cadenas,R., Vázquez,M., Aranguren,X.L., Agirre,X., Luttun,A., Prosper,F., Pascual-Montano,A. and Rubio,A. (2012) Quantification of miRNA-mRNA interactions. *PLoS One*, **7**, e30766.
82. Ovando-Vázquez,C., Lepe-Soltero,D. and Abreu-Goodger,C. (2016) Improving microRNA target prediction with gene expression profiles. *BMC Genomics*, **17**, 364.
83. Manyam,G., Ivan,C., Calin,G.A. and Coombes,K.R. (2013) targetHub: a programmable interface for miRNA–gene interactions. *Bioinformatics*, **29**, 2657–2658.
84. Bandyopadhyay,S. and Mitra,R. (2009) TargetMiner: microRNA target prediction with systematic identification of tissue-specific negative examples. *Bioinformatics*, **25**, 2625–2631.
85. Nielsen,C.B., Shomron,N., Sandberg,R., Hornstein,E., Kitzman,J. and Burge,C.B. (2007) Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA*, **13**, 1894–1910.
86. Agarwal,V., Bell,G.W., Nam,J.-W. and Bartel,D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, **4**, e05005.
87. Li,Y., Goldenberg,A., Wong,K.-C. and Zhang,Z. (2014) A probabilistic approach to explore human miRNA targetome by integrating miRNA-overexpression data and sequence information. *Bioinformatics*, **30**, 621–628.
88. Sturm,M., Hackenberg,M., Langenberger,D. and Frishman,D. (2010) TargetSpy: a supervised machine learning approach for microRNA target prediction. *BMC Bioinformatics*, **11**, 292.
89. 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.
90. Lukasik,A., Wjcikowski,M. and Zielenkiewicz,P. (2016) Tools4miRs—one place to gather all the tools for miRNA analysis. *Bioinformatics*, **32**, 2722–2724.
91. Wu,C., Bardes,E.E., Jegga,A.G. and Aronow,B.J. (2014) ToppMiR: ranking microRNAs and their mRNA targets based on biological functions and context. *Nucleic Acids Res.*, **42**, W107–W113.
92. Lee,H., Han,S., Kwon,C.S. and Lee,D. (2016) Biogenesis and regulation of the let-7 miRNAs and their functional implications. *Protein Cell*, **7**, 100–113.
93. Boyerinas,B., Park,S.-M., Hau,A., Murmann,A.E. and Peter,M.E. (2010) The role of let-7 in cell differentiation and cancer. *Endocr. Relat. Cancer*, **17**, F19–F36.
94. Lee,Y.S. and Dutta,A. (2007) The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes Dev.*, **21**, 1025–1030.
95. Hou,W., Tian,Q., Steuerwald,N.M., Schrum,L.W. and Bonkovsky,H.L. (2012) The let-7 microRNA enhances heme oxygenase-1 by suppressing Bach1 and attenuates oxidant injury in human hepatocytes. *Biochim. Biophys. Acta*, **1819**, 1113–1122.
96. Døssing,K.B., Binderup,T., Kaczkowski,B., Jacobsen,A., Rossing,M., Winther,O., Federspiel,B., Knigge,U., Kjær,A. and Friis-Hansen,L. (2014) Down-regulation of miR-129-5p and the let-7 family in neuroendocrine tumors and metastases leads to up-regulation of their targets Egr1, G3bp1, Hmga2 and Bach1. *Genes*, **6**, 1–21.
97. Liu,Q., Liu,T., Zheng,S., Gao,X., Lu,M., Sheyhidin,I. and Lu,X. (2014) HMGA2 is down-regulated by microRNA let-7 and associated with epithelial–mesenchymal transition in oesophageal squamous cell carcinomas of Kazakhs. *Histopathology*, **65**, 408–417.
98. Park,J.T., Kato,M., Lanting,L., Castro,N., Nam,B.Y., Wang,M., Kang,S.-W. and Natarajan,R. (2014) Repression of let-7 by transforming growth factor-β 1-induced Lin28 upregulates collagen expression in glomerular mesangial cells under diabetic conditions. *Am. J. Physiol. Renal Physiol.*, **307**, F1390–F1403.
99. Zhao,B., Han,H., Chen,J., Zhang,Z., Li,S., Fang,F., Zheng,Q., Ma,Y., Zhang,J., Wu,N. *et al.* (2014) MicroRNA let-7c inhibits migration and invasion of human non-small cell lung cancer by targeting ITGB3 and MAP4K3. *Cancer Lett.*, **342**, 43–51.
100. Liao,T.-T., Hsu,W.-H., Ho,C.-H., Hwang,W.-L., Lan,H.-Y., Lo,T., Chang,C.-C., Tai,S.-K. and Yang,M.-H. (2016) let-7 modulates chromatin configuration and target gene repression through regulation of the ARID3B complex. *Cell Rep.*, **14**, 520–533.
101. Shi,W., Zhang,Z., Yang,B., Guo,H., Jing,L., Liu,T., Luo,Y., Liu,H., Li,Y. and Gao,Y. (2017) Overexpression of microRNA let-7 correlates with disease progression and poor prognosis in hepatocellular carcinoma. *Medicine*, **96**, e7764.