# **Chapter 23**

# A Guide for miRNA Target Prediction and Analysis Using Web-Based Applications

Ana Lúcia Leitão, Marina C. Costa, and Francisco J. Enguita

#### **Abstract**

MiRNAs are small noncoding RNAs which act by binding to the 3'UTR of mRNA transcripts to exert a negative regulatory effect. The miRNA binding to its target follows rules based on the base complementarity of the seed sequence (2–9 first nucleotides of the miRNA sequence). Several algorithms have been developed to predict miRNA binding to genomic targets and its physiological consequences. This chapter will describe several practical aspects for the use of miRNA target prediction algorithms taking advantage of their web interfaces as well as how to produce integrative results in a graphical manner.

Key words Target predictor, miRNA, Seed sequence, Web server

#### 1 Introduction

Micro-RNAs (miRNAs) are an abundant class of small noncoding RNAs generated from specific transcription units in complex eukaryotic genomes [1]. These tiny RNAs (19–23 nucleotides) are negative posttranscriptional regulators that exert their effect by binding to the 3'UTR of a mRNA transcript mediated by a family of RNA-binding proteins designed as argonautes [2]. The protein complex responsible for the negative regulatory activity over mRNA transcripts is known as RISC (RNA Induced Silencing Complex) and contains Ago2, an argonaute protein, as a main effector. The binding of a miRNA to its cognate target is driven by Watson-Crick complementarity of a small stretch of its sequence called the "seed," which comprises the nucleotides 2–9 of the 5' end. The size of the seed sequence is directly dependent on the molecular dimensions of the substrate binding pocket of Ago2 protein [3]. In addition to Ago2, other argonaute proteins appear to be involved in miRNA function [4].

## 2 Target Prediction Algorithms

Target prediction algorithms are essential tools for discovery and characterization of miRNA function because experimental data is still limited and always derived from a particular biological process. A particular miRNA target can be recognized by a single miRNA in a set of different ways (Fig. 1). Canonical binding sites are characterized by a direct base pairing between the miRNA seed sequence and the mRNA target region comprising seven or eight nucleotides (7-mer and 8-mer target recognition) [5]. Noncanonical mechanisms for target recognition are based on the imperfect base pairing at the seed sequence, which is at least partially compensated by a base pairing at the 3'-end of the miRNA sequence [6].

Target prediction remains an important challenge in computational biology because the rules of this process are still not well understood and the size of miRNAs can lead to potential off-target predictions in big genomes. Typically, the majority of target prediction algorithms are restricted to the 3'UTR segment of mRNAs; however some of them are also designed to search for targets in the 5'UTR region and coding sequence [7]. Indeed, the biological relevance of miRNA binding to regions distinct from the 3'UTR is still a matter of discussion [8].



**Fig. 1** Types of miRNA target recognition. Canonical sites with 7 or 8 mer are based on the base pairing of the miRNA seed sequence with the mRNA target in a region flanked by a 3' adenosine nucleotide. Noncanonical sites present either an imperfect base pairing in the seed sequence, which is compensated by binding of the miRNA 3'-end to the target, or just supplementary binding of the 3'-end of the miRNA. Noncanonical binding sites also have the characteristic of not requiring the presence of an adenosine nucleotide in the targeted mRNA at the beginning of the seed sequence binding region

Target prediction algorithms are mainly based in the complementarity of the seed sequence of a miRNA with its cognate target. However, because of the limited size of miRNA sequences, other factors are added to filter the results such as interspecies conservation of the binding site, free energy of the RNA-RNA hybrid, secondary structure of the target that may prevent the binding of a miRNA, and possible existence of compensatory binding in the 3' end of the miRNA [9]. Characteristics, advantages, and drawbacks for the main algorithms are summarized in Table 1. Initially developed algorithms such as PicTar (http://pictar. mdc-berlin.de/) and TargetScan (http://www.targetscan.org/) are very stringent only allowing a perfect match between the seed sequence and the target, taking into account not only the free energy of the hybrid but also evolutionary conservation of the seed region [10, 11]. Other algorithms such as miRanda (http://www. microrna.org/), PITA (http://genie.weizmann.ac.il/pubs/mir07), and RNA22 (http://cbcsrv.watson.ibm.com/rna22.html) are more inclusive and relaxed, taking into account the possibility of compensatory binding of the 3' end of the miRNA, but also showing a higher rate of false positive predictions. MiRanda algorithm is a part of the microRNA.org web resource, and has the advantage of working also as a local computer application that can be used to tackle miRNA targets in genomes not available in general purpose databases [12]. MiRanda and PITA algorithms take into account the free energy of the miRNA-mRNA duplex in a different way. While miRanda only uses  $\Delta G$  Gibbs function to score the more relevant targets, PITA relies on the presence of potential secondary structures in the target mRNA that may prevent the miRNA binding [13]. Composite predictors such as Diana micro-T use a combination of assumptions derived from the classic algorithms, conservation of interspecies and free energy of the RNA-RNA hybrid. This web server is a component of a whole suite for miRNA analysis which includes some utilities for putative function determination and pathway integration [14, 15]. More accurate predictors based on the neural network architecture and machine learning algorithms were also recently described such as MTar [16], Diana micro-T-ANN [17], and miRmap [18, 19]. Some other recent predictors such as miRTar Hunter [20] claimed to be more successful than the others by using proprietary methods based on a dynamic programming algorithm that incorporates more sequencespecific features and reflects the already known properties of various types of target sites.

High-throughput analysis of miRNA-target interactions based on selective immunoprecipitation of RISC complex components followed by deep-sequencing analysis showed the existence of a relative enrichment of targets comprising exact miRNA seed matches [21]. Moreover, recent studies showed that noncanonical miRNA-target recognition is very frequent, reaching up to 60 % of the total miRNA targets [21]. Target prediction methods frequently

Table 1 Characteristics, strengths, and drawbacks of the main individual miRNA target prediction applications

Target	Characteristics		Quality of the prediction	
algorithm	Final score based on	Cross-species conservation	Robustness	Drawbacks
miRanda	Complementarity and free energy binding	Interspecies conservation filter	Beneficial for prediction sites with imperfect binding within seed region	Low precision, too many false positives. Medium performance in noncanonical targets
TargetScan	Seed match, 3' complementarity local AU content and position contribution 5	The application scores each result by conservation	Many parameters included in target scoring. The final score correlates with protein downregulation	Sites with poor seed pairing are omitted. Low performance in noncanonical target prediction
TargetScanS	Seed match type	Only conservative sites are considered	Simple tool for search of conserved sites with stringent seed pairing	Underestimate miRNAs with multiple target sites
PicTar	Binding energy, complementarity, and conservation	Required pairing at conserved positions	MiRNAs with multiple alignments are favored	Does not predict non-conservative sites. Low performance in noncanonical sites
DIANA-microT	Free energy binding and seed complementarity	Dataset of conserved UTRs among human and mouse is used	Probability given for each target site. Possibility of using own miRNA sequence as an input	Some miRNAs with multiple target sites may be omitted
PITA	Target site accessibility energy	User-defined	The secondary structure of 3'UTR is considered for miRNA interaction	Low efficiency compared to other algorithms, mainly in noncanonical target prediction
Rna22	Pattern recognition and folding energy	Not available	Allows identifying sites targeted by yet-undiscovered miRNAs	Low efficiency compared to other algorithms

Adapted from Witkos et al. [31]

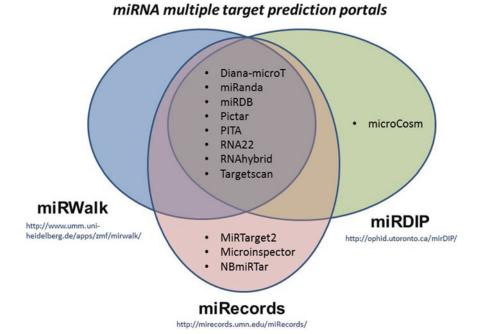
fail when predicting these noncanonical and compensatory targets. In consequence, more powerful algorithms are needed, probably including more complex biological rules that we are starting to understand from biological data. Other drawbacks from the currently available algorithms are related with the variations in the annotations of the reference databases used to map the miRNA targets. Recently, Ritchie and coworkers [22] analyzed the overlap in the number of predicted targets when the same algorithm was applied using two reference databases (UCSC and Ensembl) revealing that this overlap was less than 50 %. Moreover, the majority of the target predictors do not take into account the splicing pattern of the target gene and its cell-type specificity [22].

# 3 Multiple Predictors

One possible strategy to circumvent the intrinsic variability of the target prediction algorithms is to use what we like to call "multiple predictors." These applications are essentially web portals that will query a series of miRNA target predictors and represent the results in a user-friendly manner, showing which target is predicted by one or more algorithms. Among others, the most widely used multiple predictors are miRWalk [23], miRecords [24], and mirDIP [25]. MiRWalk and miRecords are integrated databases which include information about predicted and also validated miRNA targets. On the other hand, mirDIP is only a portal for target prediction analysis; however it offers a higher degree of customization when compared with miRWalk and miRecords. The user can filter the final results by deciding which applications to use as well as the minimum number of applications able to predict a particular miRNA target. Figure 2 shows the different number of algorithms interrogated by the multiple predictors miRWalk, miRDIP, and miRecords. All these multiple predictors are able to present the result in several formats, including tabular data format and Excel-friendly spreadsheets.

# 4 Protocol for miRNA Target Prediction Analysis in a Multiple Regulatory Scenario

Multiple regulatory scenarios are common in miRNA regulatory effect and characteristic of higher eukaryotes. They consist in a group of miRNAs which are differentially expressed in a particular cellular process, regulating different targets at the same time. In complex cellular networks, miRNAs can control the genomic output regulating genes in a coordinate or divergent manner. Unfortunately, we are far to understand which are the factors that control the quantitative contribution of each miRNA over a particular network or pathway. Indeed, scientists always have to face the dilemma of how to proceed

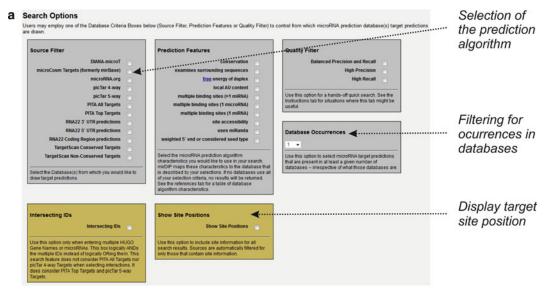


# Fig. 2 Venn diagram representing the cohort of single target prediction algorithms used by the multiple predictors mirDIP, miRWalk, and miRecords

with laboratory experiments after determining the miRNA expression profile in a particular biological situation (*see* **Note 1**).

Following the identification of misregulated miRNAs using methods such as qPCR or small RNA-seq, a coordinated target analysis can be performed as described here. The following step-by-step protocol takes advantage of the mirDIP web server for multiple target prediction [25] and the Navigator software [26] for results analysis and representation, but can be easily adapted to other softwares such as Cytoscape (http://www.cytoscape.org/).

- 1. For target prediction, access mirDIP data integration portal at the web address: http://ophid.utoronto.ca/mirDIP/.
- 2. In the upper menu, select the option "search mirDIP."
- 3. Once open the search window, introduce the identification of the misregulated miRNAs following the rules of miRBase v.18.0 separated by "enter." It is possible to enter as many miRNAs as the user like; however, if the total number of predicted targets exceeds 50,000, the software will prompt an error message suggesting the use of filters to significantly reduce the number of predictions. Species identification should be added at the beginning of the miRNA name (i.e., "hsa" for human miR-122). MirDIP is exclusive for human miRNAs; miRWalk allows the user to search for human, mouse, and rat miRNAs; and miRecords contains information for human, mouse, rat, zebrafish, fly, chicken, and dog (see Notes 2 and 3).



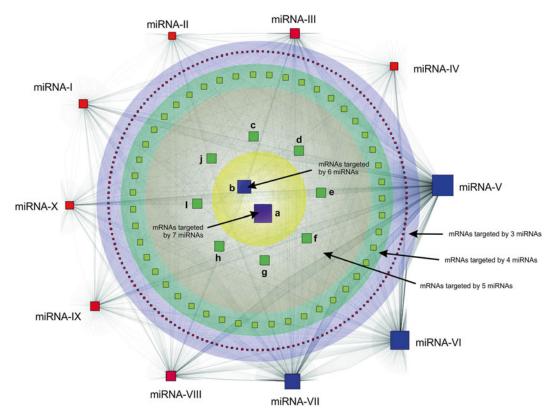
#### b Results (14):

Gene Symbol	Links ¢	microRNA \$	Source ¢	Score (orig) \$	Score (std) \$	Rank ¢
TRAF6	i2D GC	hsa-mir-146a	PITA Top Targets	-11.0	55.11	Top Third
TRAF6	i2D GC	hsa-mir-146a	DIANA-microT	8.0	2.03105	Top Third
TRAF6	i2D GC	hsa-mir-146a	microRNA.org	63.0	41.26984127	Bottom Third
TRAF6	12D GC	hsa-mir-146a	microCosm Targets (formerly mirBase)	18.1423	50.89943614	Top Third
TRAF6	i2D GC	hsa-mir-146a	RNA22 3' UTR predictions	-29.9	16.89655172	Top Third
TRAF6	2D GC	hsa-mir-146a	TargetScan Conserved Targets	-0.415	73.49382244	Top Third
TRAF6	IZD GC	hsa-mir-146a	TargetScan Non-Conserved Targets	-0.101	35.39042821	Mid Third
ZBTB2	2D GC	hsa-mir-146a	PITA Top Targets	-12.7	58.17	Top Third
ZBTB2	20 GC	hsa-mir-146a	DIANA-microT	5.0	1.1606	Top Third
ZBTB2	2D GC	hsa-mir-146a	microRNA.org	66.0	46.03174603	Bottom Third
ZBTB2	i2D GC	hsa-mir-146a	microCosm Targets (formerly mirBase)	15.9983	27.42541195	Mid Third
ZBTB2	2D GC	hsa-mir-146a	RNA22 3' UTR predictions	-29.200001	14.48276207	Top Third
ZBTB2	i2D GC	hsa-mir-146a	TargetScan Conserved Targets	-0.457	78.84345943	Top Third
ZBTB2	i2D GC	hsa-mir-146a	TargetScan Non-Conserved Targets	-0.222	50.62972292	Top Third

**Fig. 3** mirDIP interface features. (a) Input screen with the most useful customization options (predictor selection, display of target positions, and filtering for the number of applications that predict the same target); (b) mirDIP tabular output for the miRNA target prediction. Data includes the direct score generated by each individual predictor and a rank classification to assess the quality of the prediction. Data can be exported in tab- and comma-delimited format. Other multiple predictors such as miRWalk and miRecords have similar outputs

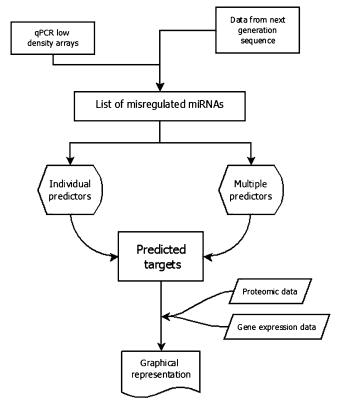
4. In order to customize the search, several filters can be applied in the "Search option" of the query page (Fig. 3a). In our hands, mirDIP software showed good performance in target prediction using the default conditions; however we strongly advise the user to modify the "Database occurrence" parameter to restrict the number of targets which are predicted by a precise number of applications. As a rule of thumb, to get a reliable prediction at least 50 % of the applications must predict a particular target. The interpretation of the results could be very difficult and, therefore, further filtering is advised to reduce the number of targets to a maximum of 100–150 genes

- per each analyzed miRNA. Some other interesting parameters can be used in the customization option such as the "multiple binding sites >1" found under the "prediction features" section, which will show only those miRNAs with more than one binding site per mRNA transcript. By default, the software will use 12 different algorithms for prediction but the user can specifically select them in the "source filter" option (*see* **Note 4**).
- 5. If you want to visualize the results on the screen or to download a file in CSV or comma delimited format for further use with other local applications, be sure to select the right format in the "Display/Download" option before activating the option "search" in the bottom part of the page. In our case, we will select the option "TAB." After executing the search, the server will prompt to save the file with the results.
- 6. Data output from the server has a tabular format (Fig. 3b), containing several columns that display the miRNA, its target genes, the application that predicts these targets, and two columns with original and standardized scores for each prediction. The last column of the table ranks the predictions in four groups depending on the standardized score: 1 % top, top third, medium third, and bottom third, being particularly useful in the assessment of each prediction. Tabular data can be imported in MS-Excel and filtered to select the targets according to a gene or a prediction score.
- 7. Graphical representation of the predictions can be achieved by specialized software such as Navigator (http://ophid.utoronto.ca/navigator/). At the time of the publication of this protocol, the Navigator version available was 2.2.1. The software must be downloaded and locally installed in a computer.
- 8. Once Navigator is installed in your operative system, execute the program and open the tabular data file generated by mirDIP selecting File>Open and the file name and format (\*.txt—TAB delimited format) of the data file. The following screens will allow the user to select the header of the file and also each interaction node. The software will construct a network on the basis of the relationships between two different nodes defined by columns in the table (miRNAs and targets).
- 9. Interpretation of raw networks is very difficult as they are produced by Navigator. However this software permits a customization of the network visualization using several parameters in the "appearance filters" section. For the particular case of miRNA target prediction from multiple miRNAs, we found extremely useful to filter the appearance of the network using the number of connections between nodes: number of miRNAs regulating a target and vice versa. The software will adjust



**Fig. 4** Layered circular representation of target prediction for multiple miRNAs in an idealized scenario. The working scenario hypothesizes the existence of ten differentially expressed miRNA (numbered in the graph using roman numerals) in a biological process. Previously to the construction of the graph, the predicted targets of each miRNA must be calculated using mirDIP, and all the data imported into Navigator or other visualization software. Symbols representing miRNAs and targets are size and color-coded to visualize the number of relationships between them. Representation of targets in layers taking into account the number of miRNAs regulating each of the targets simultaneously allows the reader to clearly visualize which target genes could be more relevant in the analyzed biological process. For instance, target genes "a" and "b" are putatively regulated simultaneously by 7 and 6 miRNAs respectively, and could be subject of further laboratory studies. These representations can be produced by software for network analysis such as Cytoscape [30] or Navigator [25] and can be complemented with further information such as protein interaction networks or relative expression values of each miRNA

- the color and size of the node symbols according to the number of connections (Fig. 4). Once filtering by the number of connections, the bigger symbols correspond to these genes or miRNAs with more connections in the network (*see* **Note 5**).
- 10. Other actions that may improve the observation of the overall relationships between miRNAs and predicted targets are the distribution of the nodes according to the levels of relationships. This can be achieved in different ways but we suggest using a circular distribution to generate regulatory layers (Fig. 4).



**Fig. 5** Flowchart for target prediction analysis using web-based applications and graphical representation

In principle, the nodes with more connections will be more relevant for the process, and susceptible of laboratory validation. For instance, this strategy has been successfully employed to pinpoint the combinatorial effect of miRNAs involved in the Myc oncogenic pathway [27].

- 11. Graphical representations could be always considered as a skeleton. Additional complexity layers can be included in the graph by using information obtained from gene expression, proteomic data, or specific databases. For instance, functional information about the relationships between the proteins encoded by the targeted mRNAs is very easy to include in the Navigator graph by overlaying another data group using the File>Import>Node features option (*see* Note 6). This information can be found in specific databases such as STRING (http://string-db.org/) [28].
- 12. The proposed flowchart for this protocol (Fig. 5) can be adapted also to other multiple predictors as miRecords and miRWalk by producing the corresponding table of miRNA–mRNA relationships and importing the data into Navigator.

### 5 Notes

- 1. MiRNA regulation of genomic output is often a consequence of the action of many miRNAs working simultaneously. The analysis of miRNA regulatory activity over a particular biological problem must be always oriented from the point of view of the systems biology. In higher eukaryotes like humans, no master miRNA is expected; in consequence, a complex network of regulatory events may exist.
- Target prediction algorithms are erratic, so the use of many applications in parallel is advised. Applications such as mirDIP, miRecords, or miRWalk, which employ many algorithms simultaneously (multiple predictors), are preferred over any single predictor.
- 3. Careful attention must be paid to the new miRNA nomenclature rules established from miRBase18.0 onwards. Current miRBase version 20 (released on June 2013) applies the same rules started in 2011 to identify the mature miRNAs coming from the same pre-miRNA by using their relative positions in the precursor transcript (5' and 3' ends) instead or their abundance within the cell [29]. Unfortunately, many of the target predictors run old versions of miRBase, a fact that should be taken into account.
- 4. It is advisable to use multiple predictors for target analysis in order to understand the effect of a group of miRNAs over a biological network. Those miRNAs putatively regulating a higher number of targets, which are regulated by more than one miRNA simultaneously, must be our focus for future laboratory experiments. This approach will direct our efforts in a more specific and intelligent way.
- 5. Graphical representation of miRNA-target interactions is always a plus to understand the complex regulatory networks. Initially, these graphical networks could be difficult to interpret but with proper filtering some interesting things may pop-out. The representation of targets in layers of complexity may help to the overall interpretation. Those targets putatively regulated by more miRNAs are probably the most relevant in our biological problem. This fact does not prevent the need for a biological validation and wet-lab studies.
- 6. Other layers of complexity can be added to the graphical representation of miRNA-mRNA interactions, including transcriptomic and/or proteomic data as well as functional data. In particular, information about protein-protein interactions will be extremely useful to understand the long-range effects of a miRNA over a metabolic pathway or network.

## **Acknowledgements**

M.C.C. was supported by a postdoctoral fellowship from Fundação para a Ciência e Tecnologia, Portugal (Ref. SFRH/BPD/65131/2009). The authors would like to acknowledge Francisco Enguita Jr. for his friendship and excellent technical support in the preparation of the manuscript.

#### References

- Ameres SL, Zamore PD (2013) Diversifying microRNA sequence and function. Nat Rev Mol Cell Biol. doi:10.1038/nrm3611
- Holcik M, Pestova TV (2007) Translation mechanism and regulation: old players, new concepts. Meeting on translational control and non-coding RNA. EMBO Rep 8:639–643
- 3. Elkayam E, Kuhn CD, Tocilj A et al (2012) The structure of human argonaute-2 in complex with miR-20a. Cell 150:100–110
- Turchinovich A, Burwinkel B (2012) Distinct AGO1 and AGO2 associated miRNA profiles in human cells and blood plasma. RNA Biol 9:1066–1075
- Brennecke J, Stark A, Russell RB et al (2005)
   Principles of microRNA-target recognition.
   PLoS Biol 3:e85
- Brodersen P, Voinnet O (2009) Revisiting the principles of microRNA target recognition and mode of action. Nat Rev Mol Cell Biol 10: 141–148
- Smalheiser NR, Torvik VI (2006) Complications in mammalian microRNA target prediction. Methods Mol Biol 342:115–127
- 8. Zhang Y, Verbeek FJ (2010) Comparison and integration of target prediction algorithms for microRNA studies. J Integr Bioinform 7. doi:10.2390/biecoll-jib-2010-127
- 9. Thomas M, Lieberman J, Lal A (2010) Desperately seeking microRNA targets. Nat Struct Mol Biol 17:1169–1174
- Friedman RC, Farh KK, Burge CB et al (2009) Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 19:92–105
- 11. Krek A, Grun D, Poy MN et al (2005) Combinatorial microRNA target predictions. Nat Genet 37:495–500
- 12. Betel D, Wilson M, Gabow A et al (2008) The microRNA.org resource: targets and expression. Nucleic Acids Res 36:D149–153

- 13. Kertesz M, Iovino N, Unnerstall U et al (2007) The role of site accessibility in microRNA target recognition. Nat Genet 39:1278–1284
- 14. Maragkakis M, Vergoulis T, Alexiou P et al (2011) DIANA-microT Web server upgrade supports Fly and Worm miRNA target prediction and bibliographic miRNA to disease association. Nucleic Acids Res 39:W145–148
- Maragkakis M, Reczko M, Simossis VA et al (2009) DIANA-microT web server: elucidating microRNA functions through target prediction. Nucleic Acids Res 37:W273–276
- Chandra V, Girijadevi R, Nair AS et al (2010) MTar: a computational microRNA target prediction architecture for human transcriptome. BMC Bioinformatics 11(Suppl 1):S2
- 17. Reczko M, Maragkakis M, Alexiou P et al (2011) Accurate microRNA target prediction using detailed binding site accessibility and machine learning on proteomics data. Front Genet 2:103
- Vejnar CE, Blum M, Zdobnov EM (2013) miRmap web: comprehensive microRNA target prediction online. Nucleic Acids Res 41: W165–W168
- Vejnar CE, Zdobnov EM (2012) MiRmap: comprehensive prediction of microRNA target repression strength. Nucleic Acids Res 40: 11673–11683
- Park K, Kim KB (2013) miRTar Hunter: a prediction system for identifying human microRNA target sites. Mol Cells 35: 195–201
- Helwak A, Kudla G, Dudnakova T et al (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell 153:654–665
- 22. Ritchie W, Flamant S, Rasko JE (2009) Predicting microRNA targets and functions: traps for the unwary. Nat Methods 6:397–398

- 23. Dweep H, Sticht C, Pandey P et al (2011) miRWalk–database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. J Biomed Inform 44:839–847
- 24. Xiao F, Zuo Z, Cai G et al (2009) miRecords: an integrated resource for microRNA-target interactions. Nucleic Acids Res 37:D105–110
- 25. Shirdel EA, Xie W, Mak TW et al (2011) NAViGaTing the micronome–using multiple microRNA prediction databases to identify signalling pathway-associated microRNAs. PLoS One 6:e17429
- Brown KR, Otasek D, Ali M et al (2009) NAViGaTOR: Network Analysis, Visualization and Graphing Toronto. Bioinformatics 25: 3327–3329
- 27. Bueno MJ, Gomez de Cedron M, Gomez-Lopez G et al (2011) Combinatorial effects of

- microRNAs to suppress the Myc oncogenic pathway. Blood 117:6255-6266
- Franceschini A, Szklarczyk D, Frankild S et al (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41: D808–815
- 29. Kozomara A, Griffiths-Jones S (2011) miR-Base: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 39: D152–157
- Zinovyev A, Viara E, Calzone L et al (2008) BiNoM: a Cytoscape plugin for manipulating and analyzing biological networks. Bioinformatics 24:876–877
- 31. Witkos TM, Koscianska E, Krzyzosiak WJ (2011) Practical aspects of microRNA target prediction. Curr Mol Med 11:93–109