

Review Article

Computational Challenges in miRNA Target Predictions: To Be or Not to Be a True Target?

**Christian Barbato,¹ Ivan Arisi,² Marcos E. Frizzo,³ Rossella Brandi,²
Letizia Da Sacco,⁴ and Andrea Masotti⁴**

¹European Brain Research Institute-Fondazione EBRI-Rita Levi-Montalcini, Via del Fosso di Fiorano, 64/65, 00143 Roma, Italy

²Neurogenomics Facility, European Brain Research Institute-Fondazione EBRI-Rita Levi-Montalcini, Via del Fosso di Fiorano, 64/65, 00143 Roma, Italy

³Departamento de Ciências Morfológicas, ICBS, UFRGS, Rua Sarmiento Leite 500, Porto Alegre, RS, CEP 90050-170, Brazil

⁴Gene Expression - Microarrays Laboratory, Bambino Gesù Children's Hospital, P.zza S. Onofrio 4, 00165 Roma, Italy

Correspondence should be addressed to Andrea Masotti, andrea.masotti@opbg.net

Received 2 January 2009; Accepted 20 March 2009

Recommended by Zhumur Ghosh

All microRNA (miRNA) target—finder algorithms return lists of candidate target genes. How valid is that output in a biological setting? Transcriptome analysis has proven to be a useful approach to determine mRNA targets. Time course mRNA microarray experiments may reliably identify downregulated genes in response to overexpression of specific miRNA. The approach may miss some miRNA targets that are principally downregulated at the protein level. However, the high-throughput capacity of the assay makes it an effective tool to rapidly identify a large number of promising miRNA targets. Finally, loss and gain of function miRNA genetics have the clear potential of being critical in evaluating the biological relevance of thousands of target genes predicted by bioinformatic studies and to test the degree to which miRNA-mediated regulation of any “validated” target functionally matters to the animal or plant.

Copyright © 2009 Christian Barbato et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

The microRNA- (miRNA-) guided “RNA” silencing pathway is a recently discovered process that is able to regulate gene expression by acting on messenger RNA (mRNA) at posttranscriptional level. miRNA biogenesis is mediated by Dicer which catalyzes the processing of double-stranded RNAs (dsRNAs) into ≈ 22 nt-long small miRNAs. The initial transcript, or “primary miRNA” (pri-miRNA), can be hundreds to thousands nucleotides long and, like any other Pol II transcript, undergoes capping and polyadenylation. The mature miRNA is part of a 60 to 80-nucleotide stem-loop structure contained within the pri-miRNA. The first step in miRNA biogenesis occurs in the nucleus and requires the excision of this hairpin structure. The excised hairpin, called pre-miRNA, is exported to the cytoplasm, and the pre-miRNA is then processed by another RNase III enzyme called Dicer. This endonuclease removes the loop region of

the hairpin, releasing the mature miRNA:miRNA* duplex. During the assembly of the RNA-induced silencing complex (RISC) with the miRNA, only one strand of the duplex is loaded, whereas the complementary miRNA* strand is removed and degraded. The mature miRNA is now ready to direct its activity on a target mRNA by binding miRNA responsive elements usually located in the 3' untranslated region (3'UTR) of the transcript. This association may result in either cleavage or translational repression of the target mRNA, depending on the degree of base-pairing between the miRNA and the responsive element. Perfect complementarity generally results in cleavage, whereas imperfect base-pairing leads to translational repression. These alternative effects might also reflect differences in the biochemical composition of the RISC complex associated to each specific miRNA:mRNA duplex. The proteins in the Argonaute (AGO) family are very tightly bound to small single-stranded RNAs within RISC, as the RNA-protein interaction persists

even under high-salt conditions. The PAZ domain of Ago has been implicated in RNA binding, and the PIWI domain seems to furnish RISC with effector-nuclease function [1]. The wide range of molecular weights reported for RISC complex (between 140 and 500 kDa) represents several different versions of the complex that contain other factors in addition to AGO. Because the other components of RISC are not required for slicing, they may have a role in other aspects of RISC activity, for example, substrate turnover and/or RISC subcellular localization. This variation may also represent species differences or may reflect developmental- or tissue-specific variations in RISC composition. The exact composition of the RISC complex is currently unknown [2].

miRNA genes represent about 1%-2% of the known eukaryotic genomes and constitute an important class of fine-tuning regulators that are involved in several physiological or disease-associated cellular processes. miRNAs are conserved throughout the evolution, and their expression may be constitutive or spatially and temporally regulated. Even in viral infections these small non-coding RNAs can contribute to the repertoire of host-pathogen interactions. The resources needed to study in details such interactions or to investigate their therapeutic implications have been recently reviewed [3]. Increasing efforts have been made to identify the specific targets of miRNAs, leading to speculation that miRNAs may regulate at least 30% of human genes. Computational predictions suggest that each miRNA can target more than 200 transcripts and that a single mRNA may be regulated by multiple miRNAs [4]. This entails that miRNAs and their targets are part of complex regulatory network and outline the widespread impact of miRNAs on both the expression and evolution of protein-coding genes [5].

The mechanism of miRNA-mediated gene regulation remains controversial. However, artificial tethering of AGO proteins to the 3'UTR of a reporter mRNA is sufficient to induce its translational repression. This evidence suggests that miRNAs may act to guide the deposition of the RISC complex onto a specific site of the target mRNA [6].

To date, the computational identification of miRNA targets and the validation of miRNA-target interactions represent fundamental steps in disclosing the contribution of miRNAs toward cell functions. The prediction of miRNA targets by computational approaches is based mainly on miRNAs complementarity to their target mRNAs, and several web-based or stand-alone computer softwares are used to predict miRNA targets [4]. Among them, *TargetScanS*, *PicTar*, and *miRanda* are the most common target prediction programs while *miRBase*, *Argonaute*, *miRNAMap*, and *miRGen* are databases combining the compilation of miRNAs with target prediction modules.

Here, we summarize and discuss the most recent *in silico* and biological approaches aimed to unravelling the functional interactions between miRNAs and their targets with a special emphasis to combined methods for more accurate miRNA target gene prediction.

2. Combining mRNA and miRNA Expression Profiles for an Accurate Target Prediction

It is now well established that the formation of a double-stranded RNA duplex through the binding of miRNA to mRNA in the RNA-induced silencing complex (RISC) triggers either the degradation of the mRNA transcript or the inhibition of protein translation. However, experimental identification of miRNA targets is not straightforward, and in the last few years, many computational methods and algorithms have been developed to predict miRNA targets [7]. Even though target prediction criteria may vary widely, most often they include: (1) strong Watson-Crick basepairing of the 5' seed (i.e., positions 2–8) of the miRNA to a complementary site in the 3'UTR of the mRNA, (2) conservation of the miRNA binding site, and (3) a local miRNA-mRNA interaction with a positive balance of minimum free energy (MFE). These requirements should be accompanied by a good structural accessibility of the surrounding mRNA sequence. However, it is likely that other important parameters for functional miRNA-target interactions remain to be identified.

The first step in the prediction procedure requires the identification of potential miRNA binding sites in the mRNA 3'UTR according to specific base-pairing rules. The second step involves the implementation of cross-species conservation requirements [8]. Among the most popular prediction algorithms, we recall *PicTar* [9], *TargetScan* [10], and *miRanda* [11]. Each algorithm has a definite rate of both false positive and false negative predictions [7]. In common practice, more than one algorithm is used to make reliable predictions about a particular gene or a specific miRNA.

Surprisingly, different algorithms provide different predictions, and the degree of overlap between different lists of predicted targets is sometimes poor or null [8].

It has been predicted that up to 30% of mammalian genes are regulated by miRNAs [11–13], and many regulatory patterns are likely to be regulated by them [14]. However, when the number of genes under study is on the order of several hundreds or thousands (like in microarray experiments), a gene-by-gene search of miRNA targets of interest becomes impractical. Furthermore, when dealing with such a number of genes that may be coregulated, the evaluation of groups of genes with common binding sites for one or specific miRNAs or families of miRNAs is surely more informative. This goal may be reached using classical enrichment statistics, testing over-representation of the miRNA target predictions within the selected set of genes (see also next paragraph): the statistical methods are similar to those used for the Gene Ontology annotation (<http://www.geneontology.org/GO.tools.html>).

However, few prediction algorithms able to clarify miRNA function or integrate data coming from different experimental high-throughput techniques are currently available. Therefore, there is the need to develop accurate computational methods for the identification of functional miRNA-target interactions. Undoubtedly, a computational method able to efficiently combine gene expression studies

TABLE 1: Common softwares for “-omics” data analysis allowing in-depth analysis of high-throughput data.

Method name	Reference	Brief description	Computer platform	Web interface	Availability	URL
Babelomics	Al-Shahrour et al. 2006	Web-based tools for genomic data analysis. Gene annotations include predicted microRNA	Any platform, web browser	yes	Free access	http://www.babelomics.org/
M@ia	Le Behec et al. 2008	Modular tools for genomic data analysis. Gene annotations include predicted microRNA	Linux, MacOS, Windows. PHP language, Apache web server and MySQL database required	no	Open-source	http://maia.genouest.org/
TIGR Multiexperiment Viewer (MeV)		Integrated environment for -omics data analysis. Gene annotations include predicted microRNA	Windows, MacOS; Java required.	no	Free executable	http://www.tm4.org/mev.html
BRB-ArrayTools		Tools for -omics data analysis. The working environment is Microsoft Excel, an R engine is providing to Excel through and add-in module. Gene annotations include predicted microRNA	Windows. Java, Excel and R language required	no	Free executable	http://linus.nci.nih.gov/~brb/download.html
GeneSpring GX		Integrated environment for -omics data analysis. Gene annotations include predicted microRNA	Windows, Java required	no	Commercial from Agilent Technologies, free trial	http://www.silicongenetics.com/
Ingenuity Pathway Analysis		Integrated environment for -omics data analysis. Gene annotations include predicted microRNA. Functional annotation and analysis of biological networks.	Windows, Java required	no	Commercial from Ingenuity Systems Inc., free trial	http://www.ingenuity.com/index.html
R Bionconductor		A common open source environment for -omics data analysis and statistics. It includes tools for microRNA analysis and annotation.	Linux, MacOS, Windows.	no	Open Source	http://www.bioconductor.org/

(mRNA profiles) with miRNAs expression profiles for a reliable prediction of miRNA target is essential. In fact, using the results of both miRNA and gene expression profiling, the prediction of miRNA-mRNA associations through the identification of anticorrelated pairs should be refined; based on the well-established knowledge of miRNA function, an upregulation of a specific miRNA will lead to lower expression of its mRNA targets, and a downregulation of a specific miRNA will lead to higher levels of its target genes. This effect is more clearly visible from in vitro studies where the system is perturbed either by the over-expression or by the silencing of a specific miRNA [15, 16]. Therefore, a ranking of downregulated (or upregulated) genes coupled to several mRNA predictions should allow the researcher to obtain a more reliable estimate of the “real” miRNA targets and finally their function [12, 13].

Unfortunately, so far this approach led to few examples, and the available software and algorithms will be briefly commented here. In contrast, a biological approach has led

to the development of several techniques that appear to be efficient alternatives to computational methods. These applications, briefly reviewed in this paper, are able to solve, at least in part, the problem of high-throughput validation of miRNA targets in vivo.

2.1. Gene Expression Analysis. Several software for the analysis of “-omics” data are commercially available or free for nonprofit organizations (Table 1). These systems are usually general purpose environments in which small databases of experimental samples can be built; the data can be filtered and normalized and also analyzed in depth using a number of statistical techniques such as analysis of variance (ANOVA), hierarchical clustering, Principal Component Analysis (PCA), among others. The same systems also offer annotation instruments such as enrichment statistics for a set of reference databases, including lists of miRNAs targeting all the known genes. The predictions come usually from

TABLE 2: Algorithms and software tools specifically developed for functional interpretation of miRNA expression data, inference of miRNA gene regulation from mRNA transcriptomic profiles, combination of parallel mRNA and miRNA expression data.

Method name	Reference	Brief description	Computer platform	Web interface	Availability	URL
miRGator	Nam et al. [17]	A web-based system to analyze microRNA expression data and to integrate parallel microRNA, mRNA, and protein profiles	Any platform, web browser	yes	Free access	http://genome.ewha.ac.kr/miRGator/
SigTerms	Creighton et al. [18]	Series of Microsoft Excel macros that compute an enrichment statistic for over-representation of predicted microRNA targets within the analyzed gene set. The software supports PicTar, TargetScan, and miRanda prediction algorithms.	Windows, Excel required	no	free source code	http://sigterms.sourceforge.net/
TopKCEMC	Lin and Ding [19]	Integration of different analysis results of the same data, each represented by a ranked list of entities. The algorithm finds the optimal list combining all the input ones. This system can be applied to the output lists of different microRNA target predictors as well as to different differentially expressed gene lists.	Linux, MacOS, Windows. R language	no	Open Source	http://www.stat.osu.edu/~statgen/SOFTWARE/TopKCEMC/
GenMIR++	Huang et al. [20]	Using a Bayesian learning network, the algorithm accounts for patterns of mRNA gene expression using miRNA expression data and a set of predicted miRNA targets. A smaller set of high-confidence functional miRNA targets then obtained from the data using the algorithm.	Any platform, Matlab language	no	Free source code	http://www.psi.toronto.edu/genmir/
MIR	Cheng and Li [21]	This method infers the level of microRNA expression starting from the gene expression profile and a gene target prediction. It is similar to GSEA for the analysis of gene expression. Every microRNA has an enrichment score based on the differential expression of its targets, weighted by a binding energy matrix.	Windows, Linux	no	Free executable	http://leili-lab.cmb.usc.edu/yeastaging/projects/microrna

the most popular computational predictors (TargetScan, PicTar, Miranda) and are not validated by databases of experimental miRNA-mRNA interactions. Given any mRNA expression profile and a selected gene list, this approach allows a first investigation of the miRNAs likely to directly modulate, at least partially, the mRNA degradation rate or indirectly modulate the mRNA transcription and translation rates. These techniques are not specifically tailored to the problem of integrating parallel miRNA and mRNA gene profiles obtained within the same experiment but are useful in combining data within the same analytical environment.

Of these tools, only Babelomics is available via web. Algorithms for functional annotation, such as FatiGO, have been integrated into a single and user friendly interface. The software GeneSpring is a commercial package that offers, together with a wide range of standard and advanced

statistical analysis methods, other enrichment statistics for functional annotations. This last feature is further developed in the Ingenuity Pathway Analysis system, specifically designed for functional and pathway analysis. Other analysis software such as the popular Bioconductor package and the MeV from the TIGR institute, are open source projects that undergo constant updates. Bioconductor works within the R language environment, which enables it to be directly integrated with several other R libraries such as the TopKCEMC reported in Table 2.

2.2. Integration and Analysis of mRNA and miRNA Data. The usefulness of bioinformatic integration of mRNA and miRNA expression data into an interaction database (Transcriptome Interaction Database) [22] was emphasized by Chen et al. [23]. However, the functional significance of

many miRNAs is still largely unknown due to the difficulty in identifying target genes and the lack of genome wide expression data combining miRNA results.

In Table 2 there is a list of some recent algorithms or tools developed to investigate the effect of miRNAs on mRNA expression profiles, to better predict miRNA targets and to integrate different data sources.

SigTerms is a novel software package (a set of Microsoft Excel macros) that has been recently developed: for a given target prediction database, it retrieves all miRNA-mRNA functional pairs represented by an input set of genes [18]. For each miRNA, the software computes an enrichment statistic for over-representation of predicted targets within the gene set. This could help to define roles of specific miRNAs and miRNA-regulated genes in the system under study. In the hands of researchers, SigTerms is a powerful tool that allows rates of false positive and false negative responses to be minimized. One method to decrease the incidence of false positive predictions and to narrow down the list of putative miRNA targets is to compare the *in silico* target predictions to the genes that are differentially expressed in the biological system of interest. SigTerms can support this type of analytical approach allowing the user to manipulate, filter, and extract different output from miRNA-mRNA sets.

Another recently reported application is miRGator [17] that integrates target predictions, functional analyses, gene expression data and genome annotations. Since the function of miRNA is mostly unknown, diverse experimental and computational approaches have been applied to elucidate their role [24, 25]. In this context, miRGator provides a utility for statistical enrichment tests of target genes, performed for gene ontology (GO) function, GenMAPP and KEGG pathways, and for various diseases. Expression correlation between miRNA and target mRNA/proteins is evaluated, and their expression patterns can be readily compared with a user friendly interface. At present, miRGator supports only human and mouse genomes.

Another major task facing researchers studying complex biological systems is the integration of data from high-throughput “-omics” platforms such as DNA variations, transcriptome profiles, and RNAomics. Recently, some miRNA-bioinformatic aspects like the biological and therapeutic repertoire of miRNAs, the *in silico* prediction of miRNA genes and their targets, and the bioinformatic challenges lying ahead have been reviewed [26]. Combined modeling of multiple raw datasets can be extremely challenging due to their enormous differences, while rankings from each dataset might provide a common base for integration. Aggregation of miRNA targets, predicted from different computational algorithms is one of these problems. Another challenging issue is the integration of results from multiple mRNA studies based on different platforms. However, one of the methods recently proposed in the literature makes use of a global optimization technique, the so-called Cross Entropy Monte Carlo (CEMC) [19]. This algorithm, called TopKCEMC, searches iteratively for the optimal list that minimizes the sum of weighted distances between the candidate (aggregate) list and each of the input-ranked lists. The distance between two ranked lists is measured

using both the modified Kendall’s tau measure and the Spearman’s footrule [27]. The application of this technique in the field of miRNA seems appropriate when the diverse predicted targets from different computational algorithms are combined together to give an aggregate list that is more informative for downstream experiments [12, 13]. This algorithm is a clear example of what we think may be well suited for combining mRNA and miRNA data to furnish a list of more reliable miRNA targets. In fact, the comparison should be made combining the “classical” list of miRNA targets (obtained from different prediction softwares) and a list of ranked downregulated (or upregulated) mRNAs.

Another proposed method of inferring the effective regulatory activities of miRNAs requires integrating microarray expression data with miRNA target predictions. As previously mentioned, the method is based on the idea that regulatory activity changes of miRNAs could be reflected by the expression changes of their target transcripts (measured by microarray techniques) [21]. To verify the hypothesis, this method has been applied to selected microarray data sets measuring gene expression changes in cell lines after transfection or inhibition of specific miRNAs. Results indicate that this method can detect activity enhancement of the transfected miRNAs as well as activity reduction of the inhibited miRNAs with high sensitivity and specificity. Furthermore, this inference is robust with respect to false positive predictions (i.e., nonspecific interactions when silencing a miRNA or when the gene downregulation is erroneously associated to a direct miRNA targeting) [15]. This method is a generalization of the gene set enrichment analysis (GSEA), which was proposed to identify gene sets associated with expression change profiles [28].

The first example of a direct correlation between mRNA expression levels and the 3’UTR motif composition has been recently reported [29]. This algorithm, a novel application of REDUCE [30], has also led to the hypothesis that the number of vertebrate miRNA could be larger than previously estimated. The algorithm’s rationale is based on the assumption that motifs within 3’UTRs make a linear contribution to enhancing or inhibiting mRNA levels. The significant motifs are chosen by iteratively looking at the individual contribution that brings the greatest reduction in the difference between the model and the expression data. Motifs with a *P*-value lower than a defined threshold are retained and listed. This method was ultimately demonstrated to be more sensitive than the current target prediction algorithms not relying on cross-species comparisons.

The same approach has been followed in another recent paper [31]. Here, the authors demonstrated that the effect of a miRNA on its target mRNA levels can be measured within a single gene expression profile. This method, however, used a known public dataset of expression both for miRNA and mRNA, limiting the usefulness of the conclusions. However, the success of this approach has revealed the vast potential for extracting information about miRNA function from other gene expression profiles.

A novel Bayesian model and learning algorithm, GenMiR++ (Generative model for miRNA regulation), has also been proposed. GenMiR++ accounts for patterns of

TABLE 3: Other computational and experimental approaches capable of performing more reliable analysis by combining miRNA and mRNA expression data.

Reference	Brief description	Computer platform
Kort et al. [32]	Two signatures of differentially expressed mRNAs and microRNAs are used to cluster the data. Qualitative combination of mRNA and microRNA expression data.	Any platform, web browser, R language
Lanza et al. [33]	One signature of differentially expressed mRNAs and microRNAs in combination is used to correctly cluster the data. Qualitative combination of mRNA and microRNA expression data.	Any platform, GeneSpring software
Salter et al. [34]	Qualitative combining of mRNA profiling and microRNA expression, by clustering separately the data and analyzing differentially modulated pathways.	Any platform, GeneSpring software, R Language, GenePattern software
Nicolas et al. [15]	Experimental identification of real microRNA targets by overexpression or silencing of miR-140.	Any platform, web browser
Sood et al. [29]	A computational tool to directly correlate 3'UTR motifs with changes in mRNA levels upon miRNA overexpression or knockdown.	Linux, Cygwin (Windows), Mac OS X, SunOS platform. A web version is also available

gene expression using miRNA expression data and a set of candidate miRNA targets [20]. A set of high-confidence functional miRNA targets is obtained from the data using a Bayesian learning algorithm. With this model, the expression of a targeted mRNA transcript can be explained through the regulatory action of multiple miRNAs. GenMiR++ allows accurate identification of miRNA targets from both sequence and expression data and allows the recovery of a significant number of experimentally verified targets, many of which provide insight into miRNA regulation.

In Table 3 we summarize some research articles where the authors have combined expression data for miRNA and mRNA, using standard analytical techniques but without the use of specifically designed algorithms.

In a recent approach aimed at identifying miRNA targets, an experimental and analysis workflow was used to find a set of genes whose expression is modulated by miR-140 [15]. This method is based on the manipulation of a miRNA activity in mouse cell lines, where miR-140 is expressed at a moderate level, thus making it easier both to repress or enhance its activity. Expression of mRNAs repressed or enhanced upon miRNA overexpression and silencing, respectively, was profiled. Within the set obtained by the intersection of the up- and down regulated mRNAs measured by microarrays, the authors searched for complementary seed sequences in the 3'UTR section of transcripts: 21 out of 49 mRNAs were identified as candidate direct targets, while the others as potential indirect ones. Interestingly, none of the 21 identified candidates were computed by popular predictors such as TargetScan, MiRBase, and PiCTar, though one of these targets, Cxcl12, was validated by Northern Blot and Luciferase assay. This method suggests that the use of more cell lines would certainly increase the set of experimentally identified targets. In fact, since some of them were already found to have escaped the analysis, they were unaffected by the type of cell manipulation chosen in this approach. This method appears to be conservative and tends to find false negative

targets especially if they are not affected at the mRNA level.

A different type of combined analysis of mRNA and miRNA profiles is often used in the field of tumors: cancers may be classified into various subclasses or may respond differently to various chemotherapeutic procedures. To correctly distinguish two subtypes of carcinomas (i.e., the colorectal cancer that can be characterized by microsatellite pathway either stability or instability), the authors have identified two different gene signatures from the mRNA and miRNA expression profiles [33]. The two signatures were extracted by standard statistical techniques such as correct T-test, PAM (Prediction Analysis of Microarray) and SVM (support vector machine, provided by Gene Spring software, see Table 1). Then, their ability to classify the samples was tested through a hierarchical clustering, both separately and together. Results showed that the better performance was obtained when the two signatures were combined together in a single clustering tree, proving once more the well-assessed crucial role played by miRNAs in the genesis of cancers. Both mRNA and miRNA gene profiles coupled to hierarchical clustering techniques were recently used in obtaining a deeper understanding of the cancer biology of the Wilm's tumor [32].

A serious problem that affects the results of antineoplastic treatments is, together with a correct diagnosis and classification, the choice of the right chemotherapeutic agent [34]. Again, both mRNA and miRNA expression signatures of sensitive and resistant cell lines were used to predict patient response to a panel of commonly used chemotherapy agents. The signatures were first used to cluster analyze samples from real breast cancer patients, then also as predictors to separate patients into nonresponders/responders to each treatment. The miRNA profiles were also finally analyzed to investigate the biological mechanisms underlying the resistance/response to the agents used in the study, making use of the prior knowledge about the experimentally validated targets of the selected miRNAs.

3. Novel Biochemical Approaches for miRNA Target Characterization.

Finally, we would like to report a few examples that show how a biochemical approach may overcome all the difficulties encountered with the computational approach.

So far, the small number of available validated miRNA targets has hindered the evaluation of the accuracy of miRNA-target prediction software. Recently, the “mirWIP” method has been proposed for the capture of all known conserved miRNA-mRNA target relationships in *Caenorhabditis elegans*, with a lower false positive rate than other standard methods [35]. This quantitative miRNA target prediction method allows an accurate weighting of some immunoprecipitation-enriched parameters, finally optimizing sensitivity to verified miRNA-target interactions and specificity.

As indicative examples, two recent studies on *C. elegans* used immunoprecipitation of miRNA-containing ribonucleoprotein complexes and evaluated that only 30%–45% of miRNAs associated with these complexes contain perfectly matched, conserved seed elements in their 3'UTRs [36, 37]. Although these datasets have provided important insights into parameters associated with functional interactions, this approach is limited to the detection of miRNA-target interactions that result in transcript destabilization and does not identify stable, translationally repressed target mRNAs. Recently, immunoprecipitation of the RISC has been used to identify mRNAs that stably associate with the endogenous RISC [38]. This study recovered 3404 mRNA transcripts that specifically coprecipitate with the miRNA-induced silencing complex (miRISC) proteins AIN-1 and AIN-2. This “AIN-IP” set of mRNA transcripts provided a biologically derived estimate of how many genes are targeted by miRNAs: in this case, at least one-sixth of *C. elegans* genes. The authors used these features to develop the prediction algorithm mirWIP, which scores miRNA target sites by weighting site characteristics in proportion to their enrichment in the experimental AIN-IP set. MirWIP has improved overall performance compared to previous algorithms, in both recovery of the AIN-IP transcripts and correct identification of genetically verified miRNA-target relationships without a requirement for alignment of target sequences. MirWIP in its current form is supported by immunoprecipitation experiments that identify transcripts by their probable association with miRNAs, even if these experiments do not directly provide information about what particular miRNA (or set of miRNAs) is responsible for miRISC association.

Finally, because the miRISC immunoprecipitation approach may be biased toward the identification of stable miRNA-target complexes, miRNA-induced target destabilization can be screened using complementary datasets, such as microarray assays to identify mRNA transcripts that change in response to miRNA activity.

To overcome the above mentioned difficulties and since the identification of the downstream targets of miRNAs is essential to understand cellular regulatory networks, a direct biochemical method for miRNA target discovery has been

proposed that combines RISC purification with microarray analysis of bound mRNAs [39]. A biochemical method of identifying miRNA targets holds the promise of deepening the understanding of the determinants of miRNA-mediated regulation, particularly by revealing targets that are repressed without changes in mRNA levels. Identification of this class of targets will provide an opportunity to study sequences or structural features determining miRNAs regulatory fate. As a model, miR-124a has been used because its targets are well known and studied. This method consisted in the Ago2 co-immunoprecipitation of mRNA targets followed by microarray profiling of mRNAs. As a result, it has been proven that not only most of the immunoprecipitated mRNAs analyzed were direct miR-124a targets but also a significant subset was downregulated.

4. Conclusions

A novel sequencing era is going to dramatically change our view of studying gene expression, posttranscriptional modifications, DNA copy number variations, and SNPs. Novel high-throughput sequencing techniques are emerging at an impressive speed on the market and on the scientific community. In the near future, these novel approaches will surely help to elucidate the function of miRNAs and their role as fine regulators. One of the most important recently reported work is based on this approach [40]. Whereas conventional methods rely on computational prediction and subsequent experimental validation of target RNAs, the proposed method consists in the direct sequencing of more than 28 000 000 signatures from the 5' ends of polyadenylated products of miRNA-mediated mRNA decay. Briefly, by matching millions of 5' end sequences of RNA cleavage products back to their corresponding sequences in the genome, additional sequences flanking the potential cleavage sites were identified. These were used to identify matches to known or new potential miRNAs that could direct their cleavage. Even though this study was conducted on *Arabidopsis thaliana*, we expect that the proposed method will also be rapidly applied to other genomes for the understanding of the role and functions of miRNAs.

In summary, we have addressed the issue of combining mRNA and miRNA expression data from different points of view. While biological validation of a predicted target is critical, failure to biologically validate the expression of a certain miRNA does not necessarily imply that the bioinformatic approach is incorrect. It is possible that the miRNA is not expressed in the examined tissues, the miRNA is expressed only in specific phase of cell cycle, or that the miRNA is expressed in low abundance, which escapes detection by the technique used. This latter cause is especially problematic for miRNA that shares a high degree of sequence homology with another miRNA. Expression of an abundant miRNA may therefore mask the expression of a rare one that is very similar in sequence, especially when using polymerase chain reaction amplification. While several methods already exist to predict miRNA targets, albeit with a heterogeneous and wide range of results, there are few

tools and algorithms or even only analysis workflow capable of elucidating the functional role of miRNAs. The wider availability of experimentally validated miRNA targets and their action mechanisms will certainly permit in the near future more reliable computational predictions.

Acknowledgment

The authors thank editors and reviewers for their useful suggestions and comments to improve the manuscript. This work was supported by grants of the Italian Ministry of Health (to A.M and L.D.S), partially funded by "Project EBRI-IIT" of the Italian Institute of Technology (IIT) and supported by "Fondazione Alazio Award 2007" (<http://www.fondazionealazio.org/> - Via Torquato Tasso 22, 90144 Palermo) (to C.B).

References

- [1] J.-J. Song, S. K. Smith, G. J. Hannon, and L. Joshua-Tor, "Crystal structure of argonaute and its implications for RISC slicer activity," *Science*, vol. 305, no. 5689, pp. 1434–1437, 2004.
- [2] V. N. Kim, J. Han, and M. C. Siomi, "Biogenesis of small RNAs in animals," *Nature Reviews Molecular Cell Biology*, vol. 10, no. 2, pp. 126–139, 2009.
- [3] Z. Ghosh, B. Mallick, and J. Chakrabarti, "Cellular versus viral microRNAs in host-virus interaction," *Nucleic Acids Research*, vol. 37, no. 4, pp. 1035–1048, 2009.
- [4] M. Lindow and J. Gorodkin, "Principles and limitations of computational microRNA gene and target finding," *DNA and Cell Biology*, vol. 26, no. 5, pp. 339–351, 2007.
- [5] D. P. Bartel, "MicroRNAs: target recognition and regulatory functions," *Cell*, vol. 136, no. 2, pp. 215–233, 2009.
- [6] R. S. Pillai, C. G. Artus, and W. Filipowicz, "Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis," *RNA*, vol. 10, no. 10, pp. 1518–1525, 2004.
- [7] N. Rajewsky, "microRNA target predictions in animals," *Nature Genetics*, vol. 38, supplement 1, pp. S8–S13, 2006.
- [8] P. Sethupathy, M. Megraw, and A. G. Hatzigeorgiou, "A guide through present computational approaches for the identification of mammalian microRNA targets," *Nature Methods*, vol. 3, no. 11, pp. 881–886, 2006.
- [9] A. Krek, D. Grün, M. N. Poy, et al., "Combinatorial microRNA target predictions," *Nature Genetics*, vol. 37, no. 5, pp. 495–500, 2005.
- [10] B. P. Lewis, C. B. Burge, and D. P. Bartel, "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets," *Cell*, vol. 120, no. 1, pp. 15–20, 2005.
- [11] B. John, A. J. Enright, A. Aravin, T. Tuschl, C. Sander, and D. S. Marks, "Human microRNA targets," *PLoS Biology*, vol. 2, no. 11, article e363, pp. 1–18, 2004.
- [12] B. P. Lewis, I.-H. Shih, M. W. Jones-Rhoades, D. P. Bartel, and C. B. Burge, "Prediction of mammalian microRNA targets," *Cell*, vol. 115, no. 7, pp. 787–798, 2003.
- [13] X. Xie, J. Lu, E. J. Kulbokas, et al., "Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals," *Nature*, vol. 434, no. 7031, pp. 338–345, 2005.
- [14] J. Li, G. Musso, and Z. Zhang, "Preferential regulation of duplicated genes by microRNAs in mammals," *Genome Biology*, vol. 9, no. 8, article R132, pp. 1–10, 2008.
- [15] F. E. Nicolas, H. Pais, F. Schwach, et al., "Experimental identification of microRNA-140 targets by silencing and overexpressing miR-140," *RNA*, vol. 14, no. 12, pp. 2513–2520, 2008.
- [16] S. S. Li, S. L. Yu, L. P. Kao, et al., "Target identification of microRNAs expressed highly in human embryonic stem cells," *Journal of Cellular Biochemistry*, vol. 106, no. 6, pp. 1020–1030, 2009.
- [17] S. Nam, B. Kim, S. Shin, and S. Lee, "miRgator: an integrated system for functional annotation of microRNAs," *Nucleic Acids Research*, vol. 36, database issue, pp. D159–D164, 2008.
- [18] C. J. Creighton, A. K. Nagaraja, S. M. Hanash, M. M. Matzuk, and P. H. Gunaratne, "A bioinformatics tool for linking gene expression profiling results with public databases of microRNA target predictions," *RNA*, vol. 14, no. 11, pp. 2290–2296, 2008.
- [19] S. Lin and J. Ding, "Integration of ranked lists via cross entropy Monte Carlo with applications to mRNA and microRNA studies," *Biometrics*, vol. 65, no. 1, pp. 9–18, 2009.
- [20] J. C. Huang, Q. D. Morris, and B. J. Frey, "Bayesian inference of microRNA targets from sequence and expression data," *Journal of Computational Biology*, vol. 14, no. 5, pp. 550–563, 2007.
- [21] C. Cheng and L. M. Li, "Inferring microRNA activities by combining gene expression with microRNA target prediction," *PLoS ONE*, vol. 3, no. 4, article e1989, pp. 1–9, 2008.
- [22] R. W. Georgantas III, R. Hildreth, S. Morisot, et al., "CD34⁺ hematopoietic stem-progenitor cell microRNA expression and function: a circuit diagram of differentiation control," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 8, pp. 2750–2755, 2007.
- [23] A. Chen, M. Luo, G. Yuan, et al., "Complementary analysis of microRNA and mRNA expression during phorbol 12-myristate 13-acetate (TPA)-induced differentiation of HL-60 cells," *Biotechnology Letters*, vol. 30, no. 12, pp. 2045–2052, 2008.
- [24] T. W. Nilsen, "Mechanisms of microRNA-mediated gene regulation in animal cells," *Trends in Genetics*, vol. 23, no. 5, pp. 243–249, 2007.
- [25] A. Rodriguez, E. Vigorito, S. Clare, et al., "Requirement of *bic/microRNA-155* for normal immune function," *Science*, vol. 316, no. 5824, pp. 608–611, 2007.
- [26] Z. Ghosh, J. Chakrabarti, and B. Mallick, "miRNomics—the bioinformatics of microRNA genes," *Biochemical and Biophysical Research Communications*, vol. 363, no. 1, pp. 6–11, 2007.
- [27] R. Fagin, R. Kumar, and D. Sivakumar, "Comparing top *k* lists," *SIAM Journal on Discrete Mathematics*, vol. 17, no. 1, pp. 134–160, 2003.
- [28] A. Subramanian, P. Tamayo, V. K. Mootha, et al., "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 43, pp. 15545–15550, 2005.
- [29] P. Sood, A. Krek, M. Zavolan, G. Macino, and N. Rajewsky, "Cell-type-specific signatures of microRNAs on target mRNA expression," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 8, pp. 2746–2751, 2006.

- [30] H. J. Bussemaker, H. Li, and E. D. Siggia, "Regulatory element detection using correlation with expression," *Nature Genetics*, vol. 27, no. 2, pp. 167–171, 2001.
- [31] A. Arora and D. A. C. Simpson, "Individual mRNA expression profiles reveal the effects of specific microRNAs," *Genome Biology*, vol. 9, no. 5, article R82, pp. 1–16, 2008.
- [32] E. J. Kort, L. Farber, M. Tretiakova, et al., "The E2F3-oncomir-1 axis is activated in Wilms' tumor," *Cancer Research*, vol. 68, no. 11, pp. 4034–4038, 2008.
- [33] G. Lanza, M. Ferracin, R. Gafà, et al., "mRNA/microRNA gene expression profile in microsatellite unstable colorectal cancer," *Molecular Cancer*, vol. 6, article 54, pp. 1–11, 2007.
- [34] K. H. Salter, C. R. Acharya, K. S. Walters, et al., "An integrated approach to the prediction of chemotherapeutic response in patients with breast cancer," *PLoS ONE*, vol. 3, no. 4, article e1908, pp. 1–8, 2008.
- [35] M. Hammell, D. Long, L. Zhang, et al., "mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts," *Nature Methods*, vol. 5, no. 9, pp. 813–819, 2008.
- [36] G. Easow, A. A. Teleman, and S. M. Cohen, "Isolation of microRNA targets by miRNP immunopurification," *RNA*, vol. 13, no. 8, pp. 1198–1204, 2007.
- [37] M. Beitzinger, L. Peters, J. Y. Zhu, E. Kremmer, and G. Meister, "Identification of human microRNA targets from isolated argonaute protein complexes," *RNA Biology*, vol. 4, no. 2, pp. 76–84, 2007.
- [38] L. Zhang, L. Ding, T. H. Cheung, et al., "Systematic identification of *C. elegans* miRISC proteins, miRNAs, and mRNA targets by their interactions with GW182 proteins AIN-1 and AIN-2," *Molecular Cell*, vol. 28, no. 4, pp. 598–613, 2007.
- [39] F. V. Karginov, C. Conaco, Z. Xuan, et al., "A biochemical approach to identifying microRNA targets," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 49, pp. 19291–19296, 2007.
- [40] M. A. German, M. Pillay, D.-H. Jeong, et al., "Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends," *Nature Biotechnology*, vol. 26, no. 8, pp. 941–946, 2008.