BIOINFORMATICS REVIEW Vol. 25 no. 23 2009, pages 3049–3055
doi:10.1093/bioinformatics/btp565

Gene expression

Lost in translation: an assessment and perspective for computational microRNA target identification

Panagiotis Alexiou[1,*], Manolis Maragkakis[1], Giorgos L. Papadopoulos[1], Martin Reczko[1,2] and Artemis G. Hatzigeorgiou[1,3]

[1]Institute of Molecular Oncology, Biomedical Sciences Research Center 'Alexander Fleming', 166 72 Varkiza, Greece, [2]Synaptic Ltd., 700 13 Heraklion, Greece and [3]Computer and Information Sciences, University of Pennsylvania, 19104-6391 Philadelphia, USA

Received on June 16, 2009; revised on September 24, 2009; accepted on September 27, 2009

Advance Access publication September 29, 2009

Associate Editor: Jonathan Wren

ABSTRACT

MicroRNAs (miRNAs) are a class of short endogenously expressed RNA molecules that regulate gene expression by binding directly to the messenger RNA of protein coding genes. They have been found to confer a novel layer of genetic regulation in a wide range of biological processes. Computational miRNA target prediction remains one of the key means used to decipher the role of miRNAs in development and disease. Here we introduce the basic idea behind the experimental identification of miRNA targets and present some of the most widely used computational miRNA target identification programs. The review includes an assessment of the prediction quality of these programs and their combinations.

Contact: p.alexiou@fleming.gr

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

It was only recently that the term microRNA (miRNA) was introduced to describe short RNA molecules that regulate gene expression by binding preferably to the 3′ untranslated region (3′UTR) of protein coding genes (Bartel, 2004). Although miRNAs were first identified in 1993 (Lee *et al*., 1993) via classical genetic techniques in *Caenorhabditis elegans*, in 2001 it was suggested that they are widespread and abundant in cells (Lagos-Quintana *et al*., 2001; Lau *et al*., 2001; Lee and Ambros, 2001). Each miRNA is 19–24 nucleotides in length and is processed from a longer transcript, referred to as the primary transcript (pri-miRNA), which can be up to thousands of nucleotides long. Primary transcripts are processed in the cell nucleus to short, ∼70 nucleotide long stem-loop structures known as pre-miRNAs. In animals, this processing is performed by a protein complex known as the Microprocessor complex, consisting of the nuclease Drosha and the double-stranded RNA binding protein Pasha (Denli *et al*., 2004). Pre-miRNAs are processed to mature miRNAs in the cytoplasm by interaction with the endonuclease Dicer which cleaves the pre-miRNA stem-loop into two complementary short RNA molecules. One of these molecules is integrated into the RNA-induced silencing complex (RISC) and guides it to the mRNA where it can inhibit translation or induce mRNA degradation (Fig. 1) (Liu *et al*., 2004). Generally, miRNA transcripts may be located within the introns of protein-coding genes, entirely outside of protein-coding genes ('intergenic') or more rarely in coding exons, untranslated regions (UTRs) or exons of non-coding transcripts. Frequently, pri-miRNA transcripts code for more than one miRNAs which are transcribed together and are referred to as a miRNA cluster.

Since their initial identification, miRNAs have been found to confer a novel layer of genetic regulation in a wide range of biological processes. Their involvement in cellular commitment and cell cycle regulation gives an important role to the miRNA class of regulatory modules in animal development and human diseases. Specifically, miRNAs have been found to regulate various developmental stages in animals such as *C.elegans* (Lau *et al*., 2001; Lee and Ambros, 2001; Lee *et al*., 1993; Reinhart *et al*., 2000), *Danio Rerio* (Wienholds *et al*., 2005), *Drosophila melanogaster* (Aravin *et al*., 2003), *Mus musculus* (Baroukh *et al*., 2007), *Homo sapiens* (Chen *et al*., 2004; Lu *et al*., 2007; Yi *et al*., 2006) and in plants (Kidner and Martienssen, 2005). miRNA-mediated regulation of pathways involved in human disease is currently a very active field and miRNAs have been linked to several human pathologies such

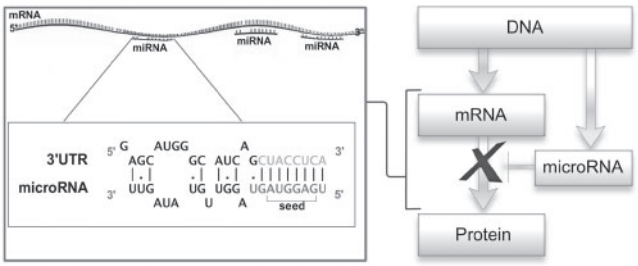

Fig. 1. The binding of a miRNA to a miTG. Multiple miRNAs may bind on the 3′UTR of a miTG. The seed sequence corresponds to six nucleotides at positions 2–7 of the miRNA sequence. The position where a miRNA binds to a miTG is called the MRE. miRNAs are transcribed mostly through Pol II from DNA. Protein coding genes are transcribed into mRNA molecules which then are translated to proteins. miRNAs integrate into the RISC complex and by binding to mRNA molecules they inhibit translation or induce mRNA degradation.

*To whom correspondence should be addressed.

as cardiovascular and neurodegenerative diseases (Hebert and De Strooper, 2007; Hebert *et al*., 2008; Zhang, 2008) as well as in human malignancies (Croce and Calin, 2005; Esquela-Kerscher and Slack, 2006; Fabbri *et al*., 2007; Gartel and Kandel, 2008; Garzon *et al*., 2006; Slack and Weidhaas, 2006). In particular, miRNAs are believed to be involved in many stages of cancer progression by both promoting and/or suppressing oncogenesis (He *et al*., 2005; Ivanovska *et al*., 2008; Lee and Dutta, 2007; Tagawa *et al*., 2007), tumor growth (Johnson *et al*., 2007; Si *et al*., 2007), invasion and metastasis (Asangani *et al*., 2008; Huang *et al*., 2008; Ma *et al*., 2007; Tavazoie *et al*., 2008; Zhu *et al*., 2008).

For many years, researchers have been analyzing microarray expression data of protein coding genes in different cancer types in order to identify specific expression signatures. The limited number of miRNAs, makes them an ideal candidate for this type of analysis. Currently, there are ∼700 human miRNAs registered in miRBase (Griffiths-Jones *et al*., 2008), and according to estimates their number may reach 1000 (Fig. 2). Analyzing their expression, several miRNA signatures have already been successfully associated with human cancers (Calin and Croce, 2006) such as leukemias (Calin and Croce, 2007; Landais *et al*., 2007), thyroid carcinomas (He *et al*., 2005), breast (Iorio *et al*., 2005), lung (Yanaihara *et al*., 2006) and pancreatic cancer (Lee *et al*., 2007).

2 EXPERIMENTAL IDENTIFICATION OF miRNA TARGETS

In order to analyze miRNA function, a large number of studies have been published that attempt to validate miRNA:mRNA interactions, using direct and indirect experimental methods. Direct methods allow the validation of specific miRNA:mRNA interactions, while indirect methods, based on high-throughput experiments such as microarrays and protein quantification experiments, provide an overview of changes in a larger number of gene products.

Direct validation of miRNA target genes is often based on the quantification of a reporter construct [e.g. Luciferase or Green Fluorescent Protein (GFP)] carrying the 3′UTR of the putative target gene after the introduction of a miRNA to the cell (Kiriakidou *et al*., 2004). Alternatively, quantitative RT–PCR can be used to monitor changes in mRNA levels after a miRNA has been introduced in a cell. Even though such methods can validate the miRNA:mRNA interaction, they fail to identify the specific miRNA recognition elements (MREs) responsible for the interaction. Such MREs can be identified using an integration of the reporter gene assay with site directed mutagenesis and/or by restoring the complementarity by mutating the miRNA sequence.

High-throughput techniques can provide information about global miRNA effects in cells and are based on measuring differential gene expression in the presence or absence of a miRNA in the cell. For the overexpression of a miRNA (Lim *et al*., 2005), expression constructs can be engineered using the mature miRNA, the precursor (hairpin) miRNA, or the pre-miRNA sequence for transfection *in vitro* or *in vivo*. Silencing of a miRNA can be accomplished by introducing chemically modified oligonucleotides perfectly complementary to the mature miRNA (Krutzfeldt *et al*., 2005) or by knocking down a miRNA gene. Until recently such gene expression levels changes have been monitored through gene expression microarrays (Krutzfeldt *et al*., 2005; Lim *et al*., 2005). These methods give significant information for miRNA targets where gene expression repression is caused by mRNA degradation (see also Supplementary Material), but is missing the targets where expression repression is caused by translation repression. Such targets were only recently identified using high-throughput proteomics methods (Baek *et al*., 2008; Selbach *et al*., 2008). In these studies, stable isotope labeling with amino acids in cell culture (SILAC) was applied and the protein expression levels for thousands of genes were measured. It should be noted that both methods provide indirect validation of targets. Recently, immunoprecipitation of RISC components has been used to identify mRNAs targeted by miRNAs (Beitzinger *et al*., 2007; Easow *et al*., 2007; Zhang *et al*., 2007). Moreover, high-throughput sequencing of RNAs isolated by crosslinking immunoprecipitation (HITS-CLIP) has been used (Chi *et al*., 2009) in order to identify and sequence specific miRNA binding sites on targeted mRNAs. Methods based on the measurement of differential expression of genes (microarrays, pSILAC), may contain many secondary and nonspecific effects and therefore the identified group of target genes does not constitute a comprehensive list of miRNA targets. Such results should be rather treated as enriched in direct miRNA targets of a specific miRNA. HITS-CLIP on the other hand might also identify non-functional binding sites of RISC. Summarizing, high-throughput methods can provide a broad set of miRNA targets in a cell that are hard to identify using direct verification methods but are not as specific as direct validation methods.

The rapid development in the methods of the experimental validation of miRNA targets and the increased interest of many labs for the function of miRNAs has caused a dramatic increase of miRNA target genes (miTGs) with experimental evidence (Fig. 2). An up to date collection of such targets including information for both the validated interaction and the methods used can be found in TarBase (Papadopoulos *et al*., 2009), a manually curated database with currently more than 1300 miRNA:mRNA interactions in several species.

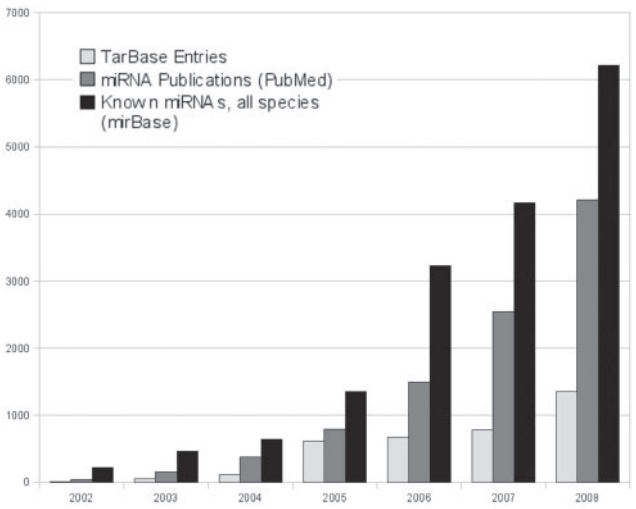

Fig. 2. The growth of known miRNA genes in miRBase database (black bars), the growth of miRNA related publications in PubMed (dark-gray bars) and the growth of the human experimentally determined miRNA target interactions in TarBase (light-gray bars).

3 OVERVIEW OF miRNA TARGET PREDICTION PROGRAMS

Despite the significant increase of experimentally validated miTGs the majority of miRNA targeted genes still remains unknown and computational target prediction programs remain the only source for a rapid identification of a putative miRNA target. Therefore, the development of computational target prediction programs goes hand in hand with the understanding of miRNA function. The first programs were developed back in 2003 shortly after it became evident that miRNAs are abundant in cells. Although a typical miRNA is ∼22 nucleotides (nt) long, several groups (Doench and Sharp, 2004; Kiriakidou *et al*., 2004) have shown experimentally that the nucleotides close to the 5′end of the miRNA are the most crucial for recognizing and binding to a target sequence. Additionally, a statistical analysis by Lewis *et al*. (2005) revealed that motifs in the 3′UTR of protein coding genes corresponding to nucleotides 2–7 of the miRNA are preferentially conserved in several species. These six nucleotides have been denoted as the 'seed' sequence of the miRNA (Fig. 1). However, later Krek *et al*. (2005) used seven nucleotides starting at position 1 or 2 of a miRNA to locate potential targets on the 3′UTR.

In the last years, several miRNA target prediction programs have been published (Sethupathy *et al*., 2006). The main prediction feature used in most of these programs is the sequence alignment of the miRNA seed to the 3′UTR of candidate target genes. Their specificity is usually increased by exploiting the evolutionary conservation of binding sites or by using additional features such as structural accessibility (Kertesz *et al*., 2007; Long *et al*., 2007), nucleotide composition (Grimson *et al*., 2007) or location of the binding sites within the 3′UTR (Baek *et al*., 2008; Gaidatzis *et al*., 2007; Grimson *et al*., 2007).

Here we summarize, in alphabetical order, eight of the most commonly used algorithms for miRNA target prediction for the human and mouse genome.

3.1 DIANA-microT 3.0

The DIANA-microT 3.0 (Maragkakis *et al*., 2009) algorithm is based on parameters calculated individually for each miRNA and each MRE depending on binding and conservation features. The prediction score of a miTG interaction is the weighted sum of the scores of conserved and non-conserved MREs on a gene. A signal to noise ratio (SNR) and a precision score are calculated for each interaction to provide an estimate of the false positive rate of each predicted miTG. Prediction data is available at http://microrna.gr/microT.

3.2 ElMMo

ElMMo (Gaidatzis *et al*., 2007) uses a general Bayesian method that scores the conservation of miRNA binding sites according to an evolutionary model that utilizes the assumed phylogenetic relationship among several species. Flat files of ElMMo target prediction data (v2, January 2008) are downloaded from http://www.mirz.unibas.ch/Computational_prediction_of_microRNA_targets_BULK.shtml. As suggested by the authors, a score threshold of 0.8 is used for high confidence in the comparisons.

3.3 miRanda

miRanda (John *et al*., 2004) uses a two-step approach for the identification of miRNA targets. First, the whole length of the miRNA is aligned against the 3′UTR sequence. Alignments that contain G:U wobble pairs are down-weighted accordingly. Second, for the highest scoring alignments, the thermodynamic stability of the complex is calculated and reported. Flat files of miRanda target prediction data are downloaded (January 2008) from: http://www.microrna.org/microrna/getDownloads.do.

3.4 miRBase

miRBase (Griffiths-Jones *et al*., 2008) uses the miRanda algorithm to identify potential binding sites for a given miRNA. Dynamic programming alignment is used to identify highly complementary sites. Strict complementarity at the 5′ seed region is demanded. Thermodynamic stability is estimated for each target site. For inclusion in the database, conservation of the target site at the exact same position in at least two species is needed. miRBase target prediction data is downloaded from http://microrna.sanger.ac.uk/cgi-bin/targets/v4/download.pl.

3.5 Pictar

Pictar (Lall *et al*., 2006) identifies two types of miRNA:target interactions: (i) those with perfect complementarity between the seed region of the miRNA (7 nt starting at position 1 or 2 of the miRNA's 5′end) and the 3′UTR target site and (ii) those for which the perfect complementarity is interrupted by at most one nucleotide bulge, mismatch, or G:U wobble. In both instances, the algorithm requires that the binding stability of the putative miRNA:target interaction, as measured by thermodynamic binding energy, exceeds a specified threshold. Once individual miRNA:target interactions are identified, the algorithm labels highly conserved (among 4 or 5 species) target sites as 'anchors' and filters out those 3′UTRs that do not harbor a specified number of anchors. A hidden Markov model is then used to score the likelihood of a 3′UTR being targeted by miRNAs in a combinatorial manner. These scores are computed for a set of species and combined to compute the final score. Since the bulk download files for Pictar on the UCSC Genome Browser are outdated, the target results are downloaded from the Pictar web page (http://pictar.org/) following the link for 'Predictions in vertebrates, flies and nematodes' (Lall *et al*., 2006). The four species conservation is used.

3.6 PITA

PITA (Kertesz *et al*., 2007) considers the effect of target site accessibility on the strength of miRNA repression. Essentially, for each target site, an energy-based measure that represents the difference between the free energy gained by the binding of the miRNA to the target and the free energy lost by unpairing the nucleotides within the target site itself is calculated. The energy used to unpair additional nucleotides flanking the target sites is also taken into account. A flat file with target prediction data is downloaded from http://genie.weizmann.ac.il/pubs/mir07/mir07_data.html. The 'no 3_15' option in the PITA Targets Catalog version 5 (November 20, 2007) is used with the top targets identified as those with a score lower than −5.

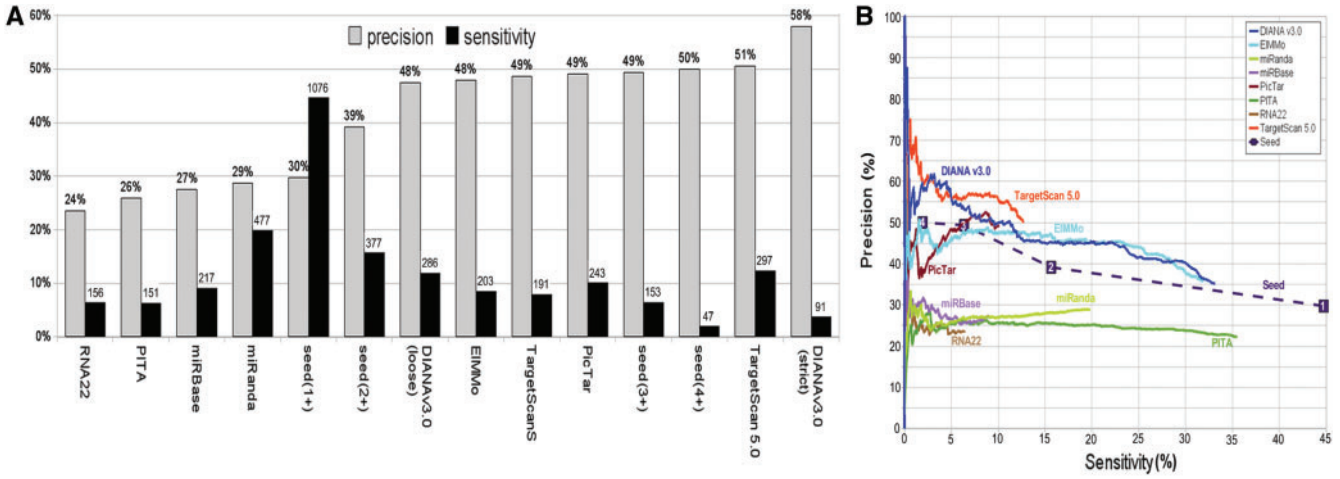

Fig. 3. Comparison of nine miRNA target prediction programs and the seed measure on the results provided by Selbach *et al.* (http://psilac.mdc-berlin.de). (**A**) The gray columns indicate precision (correctly predicted/total predicted) while the black columns show sensitivity (correctly predicted/total correct). The graph shows all targets above the score threshold of each program. A scatterplot of the same results is available in the Supplementary Materials. (**B**) A precision-receiver operating characteristic pROC (curve) showing the precision against the sensitivity of the miRNA target prediction programs. The seed measure has distinct values denoted as purple squares connected by a dotted line, where the numbers on the squares denote the minimum number of seeds per gene at each threshold. We annotate the four points having one to four seed matches.

3.7 RNA22

RNA22 (Miranda *et al*., 2006) is a miRNA target prediction program that incorporates identifying redundant patterns in mature miRNA sequences. A second-order Markov chain is implemented to estimate the statistical significance of the identified patterns. The reverse complement of all miRNA patterns are then identified within 3′UTR sequences. A 'Target Island' is an area where many such reverse complement hits accumulate. miRNAs are paired to target islands and the strength of the pairing is calculated based on the free energy and the number of nucleotides involved. The target prediction data is downloaded from http://cbcsrv.watson.ibm.com/rna22_download_content.html. The date of the precompiled predictions is November 11, 2006.

3.8 TargetScan 5.0

TargetScan (Friedman *et al*., 2009) predicts miRNA targets based on the identification of aligned seed matches and their conservation in several species. The overall scoring of a miRNA target site depends on the level of conservation, whether it binds to the miRNA on position 8 and/or whether it has an A at position 1, the distance of the target from the 3′UTR end and the AU composition of the flanking area. Data was downloaded from http://www.targetscan.org/cgi-bin/targetscan/data_download.cgi?db=vert_50.

3.9 Simple seed measure

In this approach, genes are identified and sorted according to the number of occurrences of the hexamer complementary to the seed (nucleotides 2–7) of the miRNA in the 3′UTR sequence. Unless stated otherwise, all genes containing at least one instance of the seed were used in comparisons. When multiple annotated 3′UTR sequences were available for a gene, the longest one was used.

The user interfaces of the miRNA target prediction programs described above offer a variety of options to the user and are summarized in the Supplementary Material. We would like to mention here that only a few programs (DIANA-microT 3.0, TargetScan 5.0) offer the option to predict targets for user defined novel miRNAs, and some programs offer the option of a meta analysis through information regarding miRNA and mRNA expression or/and Gene Ontology (ElMMo, miRBase). At this point, we would like to point out that programs are not always up-to-date regarding the number of miRNAs and genes used. This number ranges currently from 178 to 675. A table with the number of miRNAs for which each program gives predictions can be found in the Supplementary Materials.

4 COMPARISON OF miRNA TARGET PREDICTION PROGRAMS

In the two recently published works (Baek *et al*., 2008; Selbach *et al*., 2008) that measured changes of protein levels after overexpression or underexpression of a miRNA, several miRNA target prediction programs are evaluated. Similarly, we tested here all miRNA target prediction programs mentioned above against genes proposed as targeted in Selbach *et al.* (Material and Methods section in Supplementary Material) In Figure 3, the results for 5 miRNAs are summarized. Nearly half of the down-regulated genes contain at least one occurrence of a miRNA specific seed sequence (Fig. 3A). We notice that a group of five programs (DIANA-microT 3.0, TargetScan 5.0, TargetScanS, Pictar and ElMMo) has a precision of ∼50% with a sensitivity that ranges from 6 to 12%. All these programs rely heavily on the evolutionary conservation of the seed region or some small extensions of this region, and combine this information with other features that characterize miTGs.

Such features are detailed phylogenetic models to assess conservation, a miRNA specific SNR or a hidden Markov model to combine different MRE scores into a total miTG score (Fig. 1). Other programs include promising features like accessibility of the

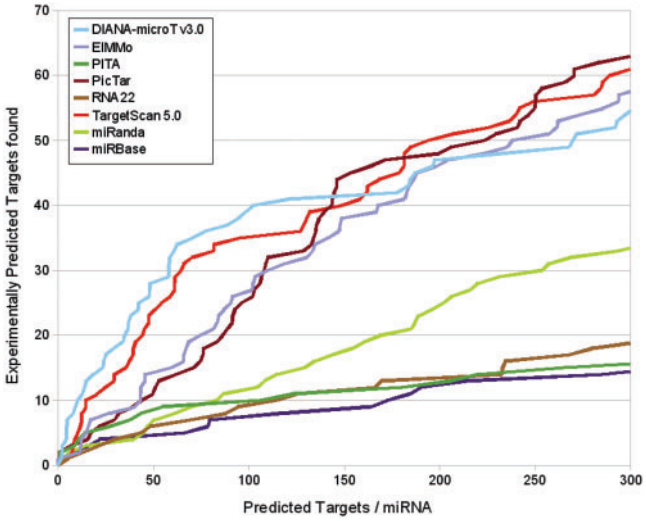

Fig. 4. Comparison of the miRNA target prediction programs on an experimentally supported miRNA target dataset. The number of correctly predicted targets is shown by different scores for increasing numbers of predicted targets per miRNA.

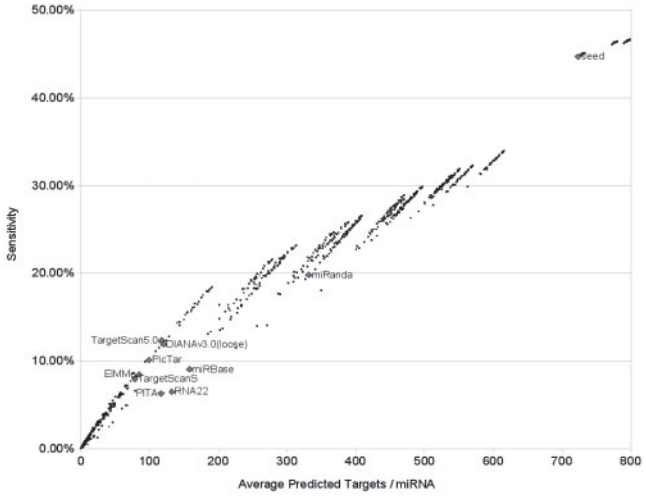

Fig. 5. Comparison of the combinations of several miRNA target prediction programs on the results provided by Selbach *et al.* The sensitivity of the prediction versus the number of predicted targets per miRNA is plotted. A larger version of this figure and an excel file with all sensitivity and specificity numbers can be found in the Supplementary Material.

binding site region, local concentration of redundant patterns of miRNA sequences, or thermodynamic stability but at the current stage they show lower predictive power. It has to be explored if these features in combination with other predictive methods can enhance target prediction.

We also investigate the very simple measure of counting the number of seed regions per gene. Nearly half of the down-regulated genes contain at least one occurrence of a miRNA specific seed sequence (Fig. 3A). Comparing the more sensitive prediction methods it can be noticed that the simple seed measure [Seed(1+) and Seed(2+)] outperforms other more complex computational methods, but fail when higher specificity is required [Seed(3+) and Seed(4+)]. Figure 3B presents the sensitivity and precision using different score cutoffs for all programs and the simple seed measure (see Supplementary Methods). The performance of the seed measure divides consistently the programs in two groups.

Further, we test the same programs with results obtained from overepxression of 2 miRNAs (hsa-mir-1 and hsa-mir-124) in HeLa cells and the subsequent measurement of mRNA levels using microarrays (Lim *et al.*, 2005) (see Supplementary Figs S3 and S4). For these data we compute the sensitivity measure at different levels for all programs. The results give a similar picture as discussed above (Supplementary Figs S3a and b, S4a and b).

A different test was performed for the same programs on a dataset of experimentally supported targets derived from TarBase (Papadopoulos *et al.*, 2009). This set includes 150 targets of 61 different miRNAs that were verified with direct experimental methods (available as Supplementary Material). The ranking of the prediction power of the tested programs shows the same order (Fig. 4).

To the non-expert, the choice of miRNA targets based on predictions by algorithms may seem like a daunting task. A natural inclination of a researcher is to assume that targets predicted by more than one algorithm are more accurate than other targets and thus leading to higher prediction precision. In a similar fashion, the union of different programs might improve the sensitivity. We test this by calculating all possible union and intersection combinations of the programs mentioned above (Fig. 5) for the high-throughput data provided for five miRNAs by Selbach *et al*. It can be observed that in most cases an accurate algorithm is better than a combination of predictions. Many of the combinations perform worse than the prediction of a single algorithm. The reason is that better specificity of a combination is achieved by a higher price for the sensitivity. Similar results are obtained using pairwise combinations of programs on the expression array data set (see Supplementary Fig. S3c).

5 FUTURE CHALLENGES OF miRNA TARGET IDENTIFICATION

The arrival of high-throughput proteomics analysis allows researchers to obtain a wider view of miRNA function in cells. Such data may help in the identification of new rules that govern miRNA function and also serve as training sets for applications based on machine learning approaches. As expression data is becoming increasingly available, it will be soon possible to train adaptive algorithms that will highlight additional rules for miRNA interactions with targeted genes. This notion is in line with the results provided in a recent publication that describes a miRNA target prediction method in *C.elegans*, mirWIP (Hammell *et al.*, 2008), which uses experimental data to define miTG prediction rules. Specifically, data from an immunoprecipitation experiment which identifies mRNAs targeted by the RISC were used and filters based on the structural accessibility of the target site, total energy of the miRNA-target hybridization as well as base pairing of the driver sequence were combined for the prediction of miTGs.

Another interesting field opening in miRNA target prediction, is the elucidation of the combinatorial effect of miRNAs. It is widely accepted that several miRNAs are co-regulated in miRNA gene clusters and are transcribed together. Additionally, levels of

several miRNAs may be correlated as markers for disease, indicating a co-regulation by more than one miRNAs. Therefore, two main questions may be asked: how do multiple miRNAs affect a single gene, and how do multiple miRNAs regulate a biological pathway or disease. High-throughput experiments involving the knock-out or overexpression of several miRNAs simultaneously as well as independently, could produce the data needed in order to tackle the first question. The second question requires more complex computational approaches that will precisely identify and predict miRNA regulatory networks and will model the interplay between miRNAs (Ivanovska and Cleary, 2008).

Traditionally, the 3′UTR has been thought of as the main region of miRNA binding. However, from as early as 2004 (Kloosterman *et al*., 2004), there have been reports that miRNA-binding sites could be functional even when artificially placed inside coding regions. In an important article laying basic rules for miRNA binding (Lewis *et al*., 2005), miRNA targeting was also detected in open reading frames of protein coding genes. More recently, the effect of introducing miRNA target sites into the 5′UTR of luciferase reporter mRNAs was extensively studied (Lytle *et al*., 2007) and naturally occurring miRNA targets in the amino acid coding sequence of mouse genes were experimentally identified (Tay *et al*., 2008). These findings indicate that miRNAs could target mRNAs by binding to positions outside the 3′UTR but it is still believed that these binding sites are scarce (Baek *et al*., 2008). However, it is possible that miRNAs act in these regions by different mechanisms and/or binding rules and therefore are hard to identify. Specifically, miRNA target prediction in coding regions would pose the difficulty of high background conservation and biased nucleotide composition.

6 CONCLUSION

Results produced by recently developed high throughput experimental techniques suggest that miRNAs have a broad impact on cellular processes. Moreover, the availability of such data allows for extensive benchmarking of existing target prediction algorithms. These benchmarks reveal that even the most sensitive programs fail to identify a large part of the targeted genes.

We believe that the dramatic progress in high throughput experimental methods will soon lead to significant qualitative and quantitative improvements in the characterization of miRNA regulation.

This will allow the development of more powerful algorithms from the statistical or machine learning field trained on such high throughput data. These methods will likely identify novel prediction rules and optimize those currently used, to create more accurate models of the underlying biological phenomena.

Closing we would like to apologize to the large number of groups working in this field whose work is not included in this review due to size limitations.

Conflict of Interest: none declared.

REFERENCES

Aravin,A.A. *et al*. (2003) The small RNA profile during Drosophila melanogaster development. *Dev. Cell*, **5**, 337–350.

Asangani,I.A. *et al*. (2008) MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Oncogene*, **27**, 2128–2136.

Baek,D. *et al*. (2008) The impact of microRNAs on protein output. *Nature*, **455**, 64–71.

Baroukh,N. *et al*. (2007) MicroRNA-124a regulates Foxa2 expression and intracellular signaling in pancreatic beta-cell lines. *J. Biol. Chem.*, **282**, 19575–19588.

Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.

Beitzinger,M. *et al*. (2007) Identification of human microRNA targets from isolated argonaute protein complexes. *RNA Biol.*, **4**, 76–84.

Calin,G.A. and Croce,C.M. (2006) MicroRNA signatures in human cancers. *Nat. Rev. Cancer*, **6**, 857–866.

Calin,G.A. and Croce,C.M. (2007) Investigation of microRNA alterations in leukemias and lymphomas. *Methods Enzymol.*, **427**, 193–213.

Chen,C.Z. *et al*. (2004) MicroRNAs modulate hematopoietic lineage differentiation. *Science*, **303**, 83–86.

Chi,S.W. *et al*. (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, **460**, 479–486.

Croce,C.M. and Calin,G.A. (2005) miRNAs, cancer, and stem cell division. *Cell*, **122**, 6–7.

Denli,A.M. *et al*. (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature*, **432**, 231–235.

Doench,J.G. and Sharp,P.A. (2004) Specificity of microRNA target selection in translational repression. *Genes Dev.*, **18**, 504–511.

Easow,G. *et al*. (2007) Isolation of microRNA targets by miRNP immunopurification. *RNA*, **13**, 1198–1204.

Esquela-Kerscher,A. and Slack,F.J. (2006) Oncomirs – microRNAs with a role in cancer. *Nat. Rev. Cancer*, **6**, 259–269.

Fabbri,M. *et al*. (2007) Regulatory mechanisms of microRNAs involvement in cancer. *Expert. Opin. Biol. Ther.*, **7**, 1009–1019.

Friedman,R.C. *et al*. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.

Gaidatzis,D. *et al*. (2007) Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics*, **8**, 69.

Gartel,A.L. and Kandel,E.S. (2008) miRNAs: little known mediators of oncogenesis. *Semin. Cancer Biol.*, **18**, 103–110.

Garzon,R. *et al*. (2006) MicroRNA expression and function in cancer. *Trends Mol. Med.*, **12**, 580–587.

Griffiths-Jones,S. *et al*. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, **36**, D154–D158.

Grimson,A. *et al*. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, **27**, 91–105.

Hammell,M. *et al*. (2008) mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nat. Methods*.

He,H. *et al*. (2005) The role of microRNA genes in papillary thyroid carcinoma. *Proc. Natl Acad. Sci. USA*, **102**, 19075–19080.

He,L. *et al*. (2005) A microRNA polycistron as a potential human oncogene. *Nature*, **435**, 828–833.

Hebert,S.S. and De Strooper,B. (2007) Molecular biology. miRNAs in neurodegeneration. *Science*, **317**, 1179–1180.

Hebert,S.S. *et al*. (2008) Loss of microRNA cluster miR-29a/b-1 in sporadic Alzheimer's disease correlates with increased BACE1/beta-secretase expression. *Proc. Natl Acad. Sci. USA*, **105**, 6415–6420.

Huang,Q. *et al*. (2008) The microRNAs miR-373 and miR-520c promote tumour invasion and metastasis. *Nat. Cell Biol.*, **10**, 202–210.

Iorio,M.V. *et al*. (2005) MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.*, **65**, 7065–7070.

Ivanovska,I. and Cleary,M.A. (2008) Combinatorial microRNAs: working together to make a difference. *Cell Cycle*, **7**, 3137–3142.

Ivanovska,I. *et al*. (2008) MicroRNAs in the miR-106b family regulate p21/CDKN1A and promote cell cycle progression. *Mol. Cell Biol.*, **28**, 2167–2174.

John,B. *et al*. (2004) Human microRNA targets. *PLoS Biol.*, **2**, e363.

Johnson,C.D. *et al*. (2007) The let-7 microRNA represses cell proliferation pathways in human cells. *Cancer Res.*, **67**, 7713–7722.

Kertesz,M. *et al*. (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**, 1278–1284.

Kidner,C.A. and Martienssen,R.A. (2005) The developmental role of microRNA in plants. *Curr. Opin. Plant Biol.*, **8**, 38–44.

Kiriakidou,M. *et al*. (2004) A combined computational-experimental approach predicts human microRNA targets. *Genes Dev.*, **18**, 1165–1178.

Krutzfeldt,J. *et al*. (2005) Silencing of microRNAs in vivo with 'antagomirs'. *Nature*, **438**, 685–689.

Lagos-Quintana,M. *et al*. (2001) Identification of novel genes coding for small expressed RNAs. *Science*, **294**, 853–858.
Lall,S. *et al*. (2006) A genome-wide map of conserved microRNA targets in C.elegans. *Curr. Biol.*, **16**, 460–471.
Landais,S. *et al*. (2007) Oncogenic potential of the miR-106-363 cluster and its implication in human T-cell leukemia. *Cancer Res.*, **67**, 5699–5707.
Lau,N.C. *et al*. (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science*, **294**, 858–862.
Lee,E.J. *et al*. (2007) Expression profiling identifies microRNA signature in pancreatic cancer. *Int. J. Cancer*, **120**, 1046–1054.
Lee,R.C. and Ambros,V. (2001) An extensive class of small RNAs in Caenorhabditis elegans. *Science*, **294**, 862–864.
Lee,R.C. *et al*. (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, **75**, 843–854.
Lee,Y.S. and Dutta,A. (2007) The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes Dev.*, **21**, 1025–1030.
Lewis,B.P. *et al*. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.
Lim,L.P. *et al*. (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, **433**, 769–773.
Liu,J. *et al*. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science*, **305**, 1437–1441.
Long,D. *et al*. (2007) Potent effect of target structure on microRNA function. *Nat. Struct. Mol. Biol.*, **14**, 287–294.
Lu,Y. *et al*. (2007) Transgenic over-expression of the microRNA miR-17-92 cluster promotes proliferation and inhibits differentiation of lung epithelial progenitor cells. *Dev. Biol.*, **310**, 442–453.
Lytle,J.R. *et al*. (2007) Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5′ UTR as in the 3′ UTR. *Proc. Natl Acad Sci. USA*, **104**, 9667–9672.
Ma,L. *et al*. (2007) Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature*, **449**, 682–688.
Maragkakis,M. *et al*. (2009) Accurate microRNA target prediction correlates with protein repression levels. *BMC Bioinformatics*, **10**, 295.
Miranda,K.C. *et al*. (2006) A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell*, **126**, 1203–1217.
Papadopoulos,G.L. *et al*. (2009) The database of experimentally supported targets: a functional update of TarBase. *Nucleic Acids Res.*, **37**, D155–D158.
Reinhart,B.J. *et al*. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature*, **403**, 901–906.
Selbach,M. *et al*. (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature*, **455**, 58–63.
Sethupathy,P. *et al*. (2006) A guide through present computational approaches for the identification of mammalian microRNA targets. *Nat. Methods*, **3**, 881–886.
Si,M.L. *et al*. (2007) miR-21-mediated tumor growth. *Oncogene*, **26**, 2799–2803.
Slack,F.J. and Weidhaas,J.B. (2006) MicroRNAs as a potential magic bullet in cancer. *Future Oncol.*, **2**, 73–82.
Tagawa,H. *et al*. (2007) Synergistic action of the microRNA-17 polycistron and Myc in aggressive cancer development. *Cancer Sci.*, **98**, 1482–1490.
Tavazoie,S.F. *et al*. (2008) Endogenous human microRNAs that suppress breast cancer metastasis. *Nature*, **451**, 147–152.
Tay,Y. *et al*. (2008) MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature*, **455**, 1124–1128.
Wienholds,E. *et al*. (2005) MicroRNA expression in zebrafish embryonic development. *Science*, **309**, 310–311.
Yanaihara,N. *et al*. (2006) Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell*, **9**, 189–198.
Yi,R. *et al*. (2006) Morphogenesis in skin is governed by discrete sets of differentially expressed microRNAs. *Nat. Genet.*, **38**, 356–362.
Zhang,C. (2008) MicroRNAs: role in cardiovascular biology and disease. *Clin. Sci.*, **114**, 699–706.
Zhang,L. *et al*. (2007) Systematic identification of C. elegans miRISC proteins, miRNAs, and mRNA targets by their interactions with GW182 proteins AIN-1 and AIN-2. *Mol. Cell*, **28**, 598–613.
Zhu,S. *et al*. (2008) MicroRNA-21 targets tumor suppressor genes in invasion and metastasis. *Cell Res.*, **18**, 350–359.