In Silico Analysis of MiRNA Promoters

Fernando Martins[1], Ângela Inácio[2], Francisco J. Enguita[1]

[1]IMM, Faculdade de Medicina da Universidade de Lisboa, 1649-028 Lisboa, Portugal
[2]CBA, Faculdade de Ciências da Universidade de Lisboa, 1749-016 Lisboa, Portugal

Abstract. MicroRNAs are an abundant class of eukaryotic non-coding RNAs, and they are involved in the negative post-transcriptional regulation of gene expression.

Recently, the DNA sequences of more than 500 human miRNA promoters have been characterized by chromatin-immunoprecipitation. The present work has the main objective of performing an *in silico* characterization of all these promoters, studying the possible transcription factors controlling miRNA expression. We are looking for transcription factors regulating miRNA expression and being simultaneously the target protein-coding gene of that same miRNA.

The purpose of this work is to assemble and characterize a catalogue of such mixed transcription factor/miRNA regulation loops in humans. All data was processed and stored in a relational database. Furthermore, a web platform was developed in order to enable further investigations. This platform is available at http://mirnatools.eu/TFmiRNA/loops.html.

Keywords: microRNAs, Transcription Factors, Regulation Loops, TFBS Databases, Targets Databases

1 Introduction

MicroRNAs (miRNAs) are small (≈22 nucleotides), non-protein-coding RNA molecules known to regulate the expression of genes by binding to the 3'-untranslated regions (3'-UTR) of mRNAs. These non-protein-coding RNA molecules are master molecular regulators that have been found to be involved in cellular processes ranging from differentiation, cell division, signal transduction and cancer.

MicroRNAs expression appears to have a tissue specific signature in which specific miRNAs are expressed preferentially in some tissues or organs. It remains unclear which are the main factors that control this tissue-specificity, however several authors have postulated the existence of a regulatory feedback loop between transcription factors controlling miRNA expression and the regulatory control exerted by miRNA over the transcription factor expression [1] [2] [3] [4] [5]. In 2008, the DNA sequences of 550 human miRNA promoters have been characterized by chromatin-immunoprecipitation [6], and this work has the main objective of performing an *in silico* characterization of all these promoters.

1.1 MicroRNA biogenesis

The first microRNA molecules, lin-4 and let-7, were identified in 1993 [7] and, since then, there has been a rapid progress in identifying more miRNAs and understanding their biogenesis, functionality and their target gene regulation.

The majority of the miRNAs identified in the first 10 years were located in the noncoding regions between genes and transcribed by unidentified promoters. These miRNAs that are produced from their own genes are also known as intergenic miRNAs. In 2003, Ambros *et al* [8] also discovered some tiny noncoding RNAs derived from the intron regions of gene transcripts; these are intronic miRNAs, i.e., miRNAs produced from introns.

Transcription factors (TFs) are proteins that either activate or repress genes transcription by binding to short cis-regulatory elements called transcription-factor binding sites. These binding sites are located in the upstream region of genes – the promoter region, which is located around the transcription start site (TSS). Post-transcriptionally, microRNAs repress mRNA translation by binding to partially complementary sites, called miRNA binding sites, in their target mRNAs. In animals, miRNA-mediated repression is often relatively weak, whereas transcription-factor-mediated repression can be much stronger [9].

Similarly to TFs, a single miRNA can regulate multiple genes, and a single gene can be regulated by multiple miRNAs. Thus, it seems quite natural to think that both miRNAs and TFs may cooperate in regulating the same target genes at the transcriptional and post-transcriptional levels. However, the molecular mechanism and nature of this interaction has not yet been understood.

TFs are essential for transcription by binding to transcription-factor binding sites. The resulting transcript is capped with a specially-modified nucleotide at the 5' end, and polyadenylated with multiple adenosines - a poly(A) tail, at the 3' end [10]. In the case of the miRNAs, this initial transcript, also known as primary miRNA (pri-miRNA), can be hundreds to thousands of nucleotides long and may contain several miRNA precursors. Each one is a hairpin structure composed by 60 to 80 nucleotides.

The hairpin RNA structure is then recognized by a nuclear protein known as DGCR8 or "Pasha". Pasha associates with the enzyme Drosha and orients this last one to excise the hairpin structure. The resulting hairpin, known as pre-miRNA, is exported from the nucleus to the cytoplasm in a process mediated by Exportin-5 protein. This transportat is energy-dependent, using GTP bound to the Ran protein [11].

In the cytoplasm, the pre-miRNA hairpin is recognized and cleaved by the Dicer enzyme, and its binding partners, TRBP protein included. This complex removes the loop region of the hairpin structure, releasing a miRNA duplex which is approximately 22 nucleotides long. The strand of the miRNA duplex that is less thermodynamically stable is preferentially loaded into the RNA-induced silencing complex (RISC) [12], which includes Dicer, TRBP and Argonaute proteins. The strand loaded into the RISC complex is called the guide strand and directs the RISC complex to its mRNA target. The other strand, the passenger strand, is subsequently degraded by an unknown mechanism [13].

The mature miRNA loaded onto to the RISC complex guides both to their mRNA target and usually binds to the 3'-UTR of the mRNA. This association may result in either cleavage or translational inhibition of the target mRNA, depending on the base

pair complementarity between the miRNA and the mRNA target region. Perfect complementarity usually results in mRNA cleavage by the RISC complex, whereas imperfect base pairing leads to translation repression [13].

1.2 Predicting transcription factor binding sites

The first step in the analysis of the transcription factor/microRNA regulation loops was to predict the transcription factor binding sites (TFBS) for all sequences of miRNA promoters published by Marson *et al* [6]. Given the miRNAs promoter sequences, it was necessary to know which TFs could bind to those promoters and regulate their transcription.

Currently, there are several programs available, including TFSEARCH 1.3 [14], MAPPER 2 [15], Match 1.0 [16], P-Match 1.0 [17], PROMO 3.0.2 [18] and TFBind [19]. Predicting TFBS using position weight matrices (PWM) is widely used and theoretically supported by Berg and von Hippel [20]. Each matrix relates a consensus sequence to the four bases and each score is proportional to the binding energy for the protein–DNA interaction [21]. Figure 1 illustrates this.

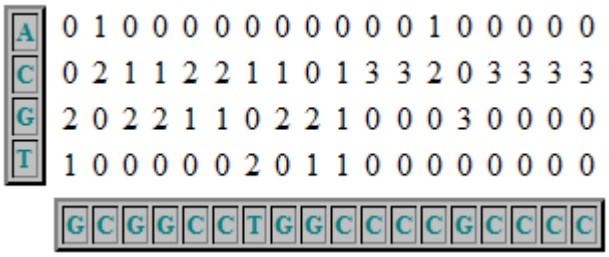

Fig. 1. Sp1 [T00757] Matrix on TRANSFAC 8.3

Matrices and TFBS have been collected into databases such as TRANSFAC [22] and JASPAR [23]. However not only all matrices have their own specificity, as prediction also requires the quantification of the similarity between the each weight matrix and a potential TFBS detected in the sequence.

In order to achieve a greater degree of accuracy, when comparing to the existing ones, several algorithms have been proposed in the last years. However, despite all efforts, these algorithms sometimes produce many false positives or false negatives. Thus, one of the major remaining problems is how to find the appropriate software. Consequently, investigators often use several of the existing programs.

1.3 Predicting microRNA targets

Nowadays it is evident that post-transcriptional processes play a much more important role in the regulation of gene expression than previously expected. So, a crucial step for the analysis of regulatory roles of miRNAs is the prediction of their targets. Although we do not know exactly the precise way how miRNAs play their role, it is known that, in animals, miRNAs are able to repress the translation of target genes by binding to a small region of nucleotides that are present at the 3'-UTR region of the regulated gene [24]. This region, called "seed", is located at positions 2-8 of the 5'

end of miRNAs and is known to contribute significantly to target recognition [25, 26]. That is why most existing algorithms start by trying to find regions of 3' UTR target gene that have strong Watson-Crick base pairing complementary to the miRNA seed sites.

Since this initial step usually results in thousands of potential target sites and many false positives, most algorithms also use other prediction criteria such as conservation of the miRNA target sites in homologous genes and local miRNA-mRNA interaction with a positive balance of minimum free energy [27]. However, several other features have been experimentally and computationally identified, considering an individual target site level as well a global mRNA level [28].

Currently, there are several programs available, e.g. Diana micro-T [29], miRanda [30], PicTar [31], PITA [32], RNA22 [33], TargetScan [34] and MicroCosm [35]. The several algorithms provide different predictions, and the degree of overlap between them is often poor or null [36]. Using GO (The Gene Ontology Consortium, 2000) has become a standard way to validate the functional coherence of genes in a target list. Nevertheless, this type of validation usually requires a statistical analysis to confirm statistical significance [37].

Additionally, databases such as miRWalk [38] and miRTarBase [39] have been published. These databases aggregate target predictions from several programs and/or also store experimentally validated targets.

2 Materials and Methods

The characterization of the DNA sequences of miRNA promoters by chromatin-immunoprecipitation provided, among other data and information, a table with human miRNA promoters and associated proteins and genomic features (Supplementary table S7). All human coordinate information upon which this investigation is based it was downloaded in January 2005 from the UCSC Genome Browser (hg17, NCBI build 35). We started from these data and the first thing done was to collect all sequences from the indicated version of UCSC Genome Browser, according to the TSS positions of all 550 promoters. For that purpose it was necessary to write a small program. One of the sequences (hsa-mir-142) was later discarded because TFBS prediction tools are unable to deal with such a lengthy sequence (406435 nucleotides).

Having all these promoters' sequences, it was then necessary to predict TFBS for all of them. For that, we used seven programs, namely Mapper 2, Match 1.0, Patch 1.0, P-Match 1.0, PROMO 3.0.2, TFBind and TFSEARCH 1.3. Except for the input sequence and, when possible, matrices selection (HS or vertebrates), default parameters were used. However, each program has its own specificities and it was necessary to deal with that in order to harmonize both inputs and outputs.

Their first difference is the way how promoter sequences can be sent to them. MAPPER 2 is the only one that was able to process a FASTA file containing all promoter sequences. For TFSEARCH 1.3 we were able to download EZRetrieve. This free tool is a TFSEARCH viewer and also processed the complete FASTA file. For TFBind we conceived a tool similar to EZRetrieve. This program reads a FASTA file and sends each sequence to the TFBind tool that is available online. Then saves the HTML outputs that can be seen when we perform the online search.

For all the others, a previous registration on the sites where these tools are available is necessary. Therefore, it is necessary to login before starting to use these tools. Because of that, it would be much more difficult to conceive a tool to perform this search automatically. We splitted our FASTA file into several small files and submit each one of them to each one of these tools.

Having all these huge amount of data, it was then necessary to prepare it to be analyzed. EZRetrieve produced a table indicating the number of binding sites for each pair of predicted transcription factor and miRNA promoter sequence given to it as input. Since the number of binding sites is a good indicator for the probability of a TF to regulate the transcription of a miRNA promoter sequence, we decided to write a tool to parse all output files of each prediction program in order to count all binding sites for each pair transcription factor/miRNA promoter.

Besides the specificities of each output, this tool had to deal with the fact that we were only interested in results from *Homo sapiens* (HS) and some programs gave us more than that. Thus, when not indicated in the output result, it was necessary to test each matrix against matrices databases in order to verify if we were in the presence of a human matrix or not. Same verification was performed with gene symbols, when necessary. Applied these filters and totalized all binding sites, it became obvious that there are significant differences among all prediction programs.

Another issue related with these predictions is gene names. These outputs usually indicate a gene name and the identification of which matrix was used to get each prediction. However, gene names are not always compatible among the several databases, because most genes have more than one name. We downloaded all *Homo sapiens* genes registered in the NIH genetic sequence database GenBank from the NCBI site. Besides the official gene symbols, this file also contains their synonyms or aliases. After comparing gene names, we were able to identify most of the genes listed in the outputs of the TFBS applications. However, some of them remain unclassified and many others are not *Homo sapiens* genes.

The next step regards to the prediction of miRNA targets. Unlike, TFBS predictions, these databases can be downloaded as text files. Despite the huge size of some databases, it is quite simple to understand their layouts and, when necessary, to write data extraction programs. For this purpose, we started by using miRWalk target published predictions. Targets of all miRNAs analyzed by Marson *et al* [6] were compared with *Homo sapiens* genes predicted by all TFBS applications and with their synonyms as well. Then it was necessary to identify the matching predictions between mirWalk targets and all TFBS databases in order to find the predicted loops. It is important to clarify that predicted loops were based on matching predictions only.

This analysis was also performed using mirTarBase, a database with experimentally validated targets. These first two databases differ from the others because their results are published scientific articles, instead of predicted sites.

The next database used was Diana micro-T v3.0. This database has about 2.5 million records and targets are identified by Ensemble ID. Since GenBank also contains Ensemble IDs, we wrote a program in order to extract from Diana database all records in which the target gene is one of the genes predicted by the TFBS applications.

We also analyzed miRanda databases. There are four of them, combining good and non-good mirSVR scores with conserved and non-conserved miRNAs. However, we only analyzed good mirSVR scores databases. In these databases genes are identified

by GeneBank ID (NCBI Entrez ID) and we started by writing a program in order to extract from these databases all records in which the target gene is one of the genes predicted by the TFBS applications. Similar procedures were adopted in order to analyze MicroCosm and TargetScan databases. This last one also has predictions for conserved and non-conserved sites.

All collected data was stored in a relational database and a web platform was developed in order to enable further investigations.

3 Results

After analyzing all selected data (see Materials and Methods), we found 16450 of such loops, covering 311 distinct transcription factors and 344 distinct miRNAs. Using databases concordance as reliability criteria, only 5 loops were predicted by seven of the eight miRNA targets databases used: hsa-mir-20b/STAT3, hsa-mir-200b/ZEB1, hsa-mir-200c/ZEB1, hsa-mir-373/RELA and hsa-mir-429/ZEB1. However, several TFBS tools did not contribute to these predictions. In fact, the pair hsa-mir-9/NFKB1 is the only loop predicted by all seven TFBS tools and six targets databases. However, this result is not as good as it appears to be because the average number of both predicted binding and target sites is very low. Nevertheless, there are several investigations relating NFKB1 with hsa-mir-9.

If NFKB1 is involved in the loops with highest databases concordance, MYB transcription factor is involved in the most loops with the highest target sites average (see Table 1 for details), considering TFBS and targets predicted by at least three databases. In fact, MYB is involved in eighteen of the first twenty two loops in these conditions.

Table 1. Regulation loops that have the highest target sites average.

miRNA	TF	#TFBS Apps	Avg. TFBS	# Targets Apps	Avg. Targets
hsa-mir-150	MYB	4	9.25	6	30.25
hsa-mir-182	MYB	5	21.00	4	23.00
hsa-mir-607	IKZF1	3	24.67	3	22.33
hsa-mir-155	MYB	3	4.00	5	20.00
hsa-mir-195	MYB	4	1.50	3	20.00
hsa-mir-497	MYB	4	1.50	3	20.00

At this point, it is important to say that average target sites were calculated using only six databases, because the other two used databases do not indicate the number of target sites. MirTarBase contains experimentally validated targets and mirWalk contains published targets only.

Considering mirTarBase as a reliable source of miRNA targets and selecting only loops with targets predicted by mirTarBase and whose TFBS were predicted by at least 5 tools, we have the 19 regulation loops listed in Table 2. As we can see in this table, the average number of TFBS for the hsa-mir-124/SP1 loop is much higher than

all other loops. This is because both Patch 1.0 and TFBind predicted hundreds of TFBS in this case.

Table 2. Regulation loops predicted by mirTarBase and at least five TFBS applications.

miRNA	TF	Avg. TFBS	#TFBS Apps	# Targets Apps
hsa-mir-9	NFKB1	23.86	7	6
hsa-mir-15a	NFKB1	5.86	7	5
hsa-let-7a	NFKB1	16.14	7	2
hsa-mir-146a	NFKB1	5.00	7	2
hsa-mir-23b	PLAU	2.33	6	5
hsa-mir-106a	RUNX1	78.33	6	5
hsa-mir-200b	ZEB1	8.40	5	7
hsa-mir-200c	ZEB1	3.20	5	7
hsa-mir-429	ZEB1	8.40	5	7
hsa-mir-424	MYB	15.40	5	6
hsa-mir-101	FOS	6.80	5	6
hsa-mir-16	MYB	2.40	5	6
hsa-mir-124	SP1	766.80	5	5
hsa-mir-141	ZEB1	3.20	5	4
hsa-mir-200a	ZEB1	8.40	5	4
hsa-mir-122	SRF	16.60	5	4
hsa-mir-124	AHR	30.80	5	4
hsa-mir-218	SP1	32.80	5	3
hsa-mir-27a	SP1	170.80	5	3

Since the number of predicted binding sites is a good indicator for the probability of a TF to regulate the transcription of a miRNA promoter sequence, it is important to analyze which TFs and miRNAs have the highest number of predicted TFBS. These values are shown in Table 3 and Table 4.

Table 3. Top 10 of TFs by sum of predicted TFBS.

TF	Total BS	Total loops	Mean BS
SP1	45631	197	231.63
RUNX1	26965	241	111.89
POU2F1	22013	265	83.07
CREB1	11988	242	49.54
REL	8603	62	138.76
TP53	8276	188	44.02
MYB	6756	206	32.80
NFKB1	5992	104	57.62
FOS	5676	111	51.14
PAX5	5152	119	43.29

Table 4. Top 10 of miRNAs by sum of predicted TFBS.

miRNA	Total BS	Total loops	Mean BS
hsa-miR-124	12809	107	119.71
hsa-miR-106a	6084	109	55.82
hsa-miR-607	6052	114	53.09
hsa-miR-587	5683	103	55.17
hsa-miR-425	5538	72	76.92
hsa-miR-374b	5395	123	43.86
hsa-miR-374a	5128	111	46.20
hsa-miR-122	5128	87	58.94
hsa-miR-421	4927	102	48.30
hsa-miR-92b	4829	78	61.91

As result of this work, we assembled and characterized a catalogue of such mixed transcription factor/miRNA regulation loops in human. All data is stored in a relational database and a web platform was developed in order to enable further investigations. This database has 38 tables and stores about 1.7 million records. The web interface is available at http://mirnatools.eu/TFmiRNA/loops.html and allows searching for loops using several criteria. Also presents all details of every loop such as predicted TFBS and targets, scores of each prediction, etc.

4 Discussion

We were looking for transcription factors regulating the expression of a miRNA and being simultaneously the target protein-coding gene of that same miRNA, as illustrated in Figure 2. It is known that the cell's machinery is designed in order to minimize energy consumption, so why should a gene regulate the expression of a miRNA and being simultaneously its target, usually resulting in its own translational repression?

The existence of such regulatory loops seems to reveal a complex mechanism of genes regulation. Despite the fact that we cannot yet understand the biological significance of these regulatory loops, their existence seems to be evident and should be experimentally validated. However, it is important to be aware that all loops predicted by our analysis are based on matching predictions only. Further investigations should address some more complex issues, such as:

- The fact that transcriptional and posttranscriptional regulation are very likely to occur at different time scales.
- Positive vs. negative regulatory feedbacks have important consequences in terms of network dynamics. Importantly, they are known to be prerequisites for the existence of multistability and oscillatory behavior, respectively.
- MicroRNAs regulate cellular networks as network components and it would be of key interest to assess the impact of the identified loops as part of the gene regulatory networks.

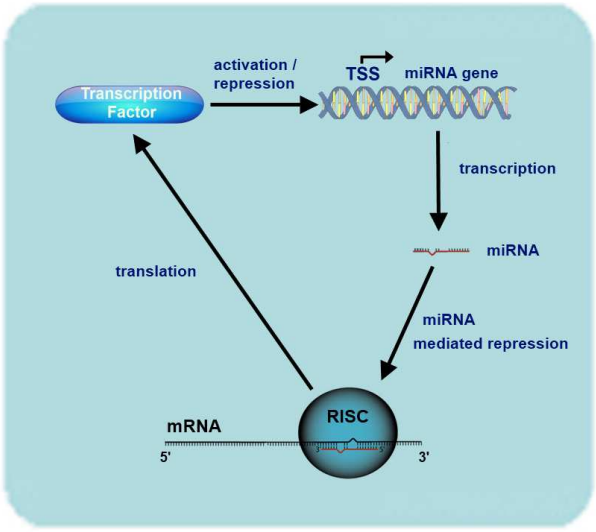

Fig. 2. Mixed transcription factor/miRNA regulation loops

Since a single miRNA can regulate multiple genes and a single gene can be regulated by multiple miRNAs, it is quite natural to think that both miRNAs and TFs may cooperate in regulating the same target genes at the transcriptional and post-transcriptional levels. In fact, the co-regulation of transcription factors and microRNAs in transcriptional regulatory networks is a subject that has been investigated by several authors [1] [2] [3] [40] [41] [42] [43] [44] [45] [46].

Clearly, miRNAs cannot independently perform a single task in cells. Instead, miRNAs regulate cellular networks as network components in many cellular functions [4]. In fact, TFs and miRNAs function together in gene regulatory networks that are not yet completely identified and understood. Consequently, all loops identified by this investigation should be seen as components of regulatory modules, instead of isolated loops. Although this is true, we can also analyze each one of these individual loops.

A similar loop was found in the developing of *Drosophila melanogaster* eye [5]. Author's investigation revealed that, in nonstimulated cells, Yan represses miR-7 transcription, whereas miR-7 RNA represses Yan protein expression in photoreceptors, by binding to sequences within its mRNA 3' UTR. This mutually inhibitory relationship helps to partition the expression of Yan into eye progenitor cells and that of miR-7 into differentiating photoreceptors, contributing to these two alternative fates. According to the authors' conclusion, this mechanism can explain how signal transduction activity can robustly generate a stable change in gene-expression patterns.

As demonstrated in the Materials and Methods section, prediction of both TFBS and targets varies widely among all tools. To reduce the number of predictions and to try to raise the reliability of predicted results, the usual procedure is to consider only those results that are predicted by several algorithms, assuming this overlap as a high-

er-quality subset of predictions. However, this is not necessarily true. In fact, as indicated by Ritchie W et al. [47], this can be a trap. They suggest that searching for overlaps between miRNA target prediction algorithms should be discouraged owing to a lack of utility and rationale. For this reason and because we did not want to restrict future investigations, we decided to publish results from all used databases, despite the certainty that the vast majority of these predictions are not real loops.

Keeping in mind the fact that this is an *in silico* analysis, we should be aware that the vast majority of all detected loops have a very low probability of being real loops. All predicted loops rely on several other tools and, as postulated by GIGO (garbage in, garbage out) axiom, if invalid data is entered into a system, the resulting output will also be invalid. Therefore, further investigations should start by defining reliability criteria. Best validation would be to compare all predictions with experimentally validated targets. However, such datasets are too small to be used as benchmarks. Since each database scores each one of the predictions, these scores can be used to identify the most reliable predictions and an overall score can also be computed.

Another possible way to perform additional validations is by using principal component analysis (PCA). After applying PCA on a matrix with the number of predictions (for example) for some selected miRNAs and genes, we can visually analyze how miRNAs are related to each other concerning the TFs that control their transcription, as predicted by each one of the databases. We can cluster these results, measuring the Euclidean distance of all miRNAs (for example). However, we can also cluster all data used to perform PCA analysis and get a cluster dendrogram. Then, we can compare all resulting clusters with miRNA clusters already validated.

Our ongoing work is trying to build up more knowledge in this research area.

5 Conclusion

Since cell's machinery is designed in order to minimize energy consumption, it would be unlikely for a gene to regulate the expression of a miRNA and being simultaneously its target, usually resulting in its own translational repression at a post-transcriptional level. However, this *in silico* analysis has found 16450 potential loops, covering 311 distinct transcription factors and 344 distinct miRNAs. Some of these loops have a great probability of being experimentally confirmed. Although not being the ultimate goal of this investigation, we also computed a score for each predicted loop. With this or any other scoring system it is possible to guide experimental validations of predicted loops.

Despite the fact that we cannot yet understand the biological significance of these regulatory loops, their existence seems to be evident and this must be an important mechanism of genes regulation. However, all these data demand for further investigations and experimental validations. In order to enable further investigations, we developed a web platform through which all data can be analyzed.

6 References

1. Tsang J et al.: MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. Mol Cell. Jun 8;26(5):753-67 (2007)
2. Qiu C et al.: microRNA evolution in a human transcription factor and microRNA regulatory network. BMC Syst Biol. Jun 29;4:90 (2010)
3. Tran DH et al.: Computational discovery of miR-TF regulatory modules in human genome. Bioinformation. Feb 28;4(8):371-7 (2010)
4. Yichen Li et al.: MicroRNA-mediated positive feedback loop and optimized bistable switch in a cancer network Involving miR-17-92. Plos One, 6(10): e26302 (2011)
5. Li X, Carthew RW: A microRNA mediates EGF receptor signaling and promotes photoreceptor differentiation in the Drosophila eye. Cell Dec 29;123(7):1267-77 (2005)
6. Marson et al.: Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell, 134, 521-533 (2008)
7. Lee RC, Feinbaum RL, Ambros V: The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell, 75, 843-854 (1993)
8. Ambros et al.: Identification of microRNAs and other tiny noncoding RNAs by cDNA cloning. Methods Mol Biol, 265, 131–158 (2004)
9. K Chen and N Rajewsky: The evolution of gene regulation by transcription factors and microRNAs. Nature Reviews, 8:93 (2007)
10. Cai X, Hagedorn CH, Cullen BR: Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. RNA 10 (12): 1957–66 (2004)
11. Murchison E, Hannon G: miRNAs on the move: miRNA biogenesis and the RNAi machinery. Current opinion in cell biology 16 (3): 223–229 (2004)
12. Schwarz DS et al.: Asymmetry in the assembly of the RNAi enzyme complex. Cell. Oct 17;115(2):199-208 (2003)
13. Okamura K et al: The long and short of inverted repeat genes in animals: microRNAs, mirtrons and hairpin RNAs. Cell Cycle. Sep 15;7(18):2840-5 (2008)
14. Yutaka Akiyama, Kyoto University, http://www.cbrc.jp/research/db/TFSEARCH.html (1995)
15. Marinescu VD, Kohane IS, Riva A: MAPPER: a search engine for the computational identification of putative transcription factor binding sites in multiple genomes. BMC Bioinformatics, 6:79 (2005)
16. Kel AE et al.: MATCH: A tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Res., 31(13):3576-9 (2003)
17. Chekmenev DS, Haid C, Kel AE: P-Match: transcription factor binding site search by combining patterns and weight matrices. Nucleic Acids Res., 33(Web Server issue):W432-7 (2005)
18. Farré D et al.: Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. Nucleic Acids Res., 31(13):3651-3 (2003)
19. Tsunoda T, Takagi T: Estimating transcription factor bindability on DNA. Bioinformatics, 15(7-8):622-30 (1999)
20. Berg OG, von Hippel PH: Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters. J Mol Biol, 193(4):723-50 (1987)
21. Fields DS et al: Quantitative specificity of the Mnt repressor. J Mol Biol., 271(2):178-94 (1997)
22. Heinemeyer T et al.: Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. Nucleic Acids Res., 26(1):362-7 (1998)

23. Sandelin A et al: JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res., 32(Database issue):D91-4 (2004)
24. Lai EC: Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. Nature Genetics 30, 363 – 364 (2002)
25. Rajewsky N: microRNA target predictions in animals. Nature Genetics 38, Suppl:S8-13 (2006)
26. Thomas M, Lieberman J, Lal A.: Desperately seeking microRNA targets. Nat Struct Mol Biol. Oct;17(10):1169-74 (2010)
27. Barbato C et al.: Computational challenges in miRNA target predictions: to be or not to be a true target. J Biomed Biotechnol., 2009:803069 (2009)
28. Saito T, Sætrom P: A two-step site and mRNA-level model for predicting microRNA targets. BMC Bioinformatics, 11:612 (2010)
29. Maragkakis M et al.: DIANA-microT web server: elucidating microRNA functions through target prediction. Nucleic Acids Res., 37(Web Server issue):W273-6 (2009)
30. John B et al.: Human MicroRNA targets. PLoS Biol., 2(11):e363 (2004)
31. Krek A et al.: Combinatorial microRNA target predictions. Nature Genetics, 37(5):495-500 (2005)
32. Kertesz M et al.: The role of site accessibility in microRNA target recognition. Nature Genetics, 39(10):1278-84 (2007)
33. Miranda KC et al.: A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. Cell, 22, 126(6):1203-17 (2006)
34. Lewis BP, Burge CB, Bartel DP: Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell, 120(1):15-20 (2005)
35. Griffiths-Jones S. et al.: miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. Jan 1;34(Database issue):D140-4, (2006)
36. Sethupathy P: A guide through present computational approaches for the identification of mammalian microRNA targets. Nature Methods, 3(11):881-6. (2006)
37. Yoon S, De Micheli G.: Prediction of regulatory modules comprising microRNAs and target genes. Bioinformatics. Sep 1;21 Suppl 2:ii93-100 (2005)
38. Dweep H et al.: miRWalk--database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. Journal of Biomedical Informatics, (2011)
39. Hsu SD et al.: miRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic Acids Res., 39(Database issue):D163-9 (2011)
40. Chen CY et al.: Coregulation of transcription factors and microRNAs in human transcriptional regulatory network. BMC Bioinformatics, Feb 15;12 Suppl 1:S41 (2011)
41. Flynt AS, Lai EC: Biological principles of microRNA-mediated regulation: shared themes amid diversity. Nat Rev Genet. Nov;9(11):831-42 (2008)
42. El Baroudi M et al.: A curated database of miRNA mediated feed-forward loops involving MYC as master regulator. PLoS One 3;6(3):e14742 (2011)
43. Shalgi R et al.: Global and local architecture of the mammalian microRNA-transcription factor regulatory network. PLoS Comput Biol. Jul;3(7):e131 (2007)
44. Su N et al.: Combinatorial regulation of transcription factors and microRNAs. BMC Syst Biol. Nov 8;4:150 (2010)
45. Lee D, Shin C.: MicroRNA-target interactions: new insights from genome-wide approaches. Ann N Y Acad Sci. Oct;1271:118-28 (2012)
46. Marco A et al: MicroRNAs from the same precursor have different targeting properties. Silence. Sep 27;3(1):8 (2012)
47. Ritchie W et al.: Predicting microRNA targets and functions: traps for the unwary. Nature Methods, 6(6):397-8 (2009)