CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Elsevier - Publisher Connector

Genomics 100 (2012) 352–356

Contents lists available at SciVerse ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

miRNA_Targets: A database for miRNA target predictions in coding and non-coding regions of mRNAs

Amit Kumar [a,b,*], Adam K.-L. Wong [c], Mark L. Tizard [a], Robert J. Moore [a], Christophe Lefèvre [b]

[a] CSIRO Livestock Industries, Australian Animal Health Laboratory, Geelong, Victoria 3220, Australia
[b] Institute for Frontier Materials, Deakin University, Geelong, Victoria 3216, Australia
[c] School of Information Technology, Deakin University, Geelong, Victoria 3216, Australia

ARTICLE INFO

Article history:
Received 16 April 2012
Accepted 20 August 2012
Available online 25 August 2012

Keywords:
Database
MicroRNA
Target predictions
Coding region
5 prime UTR

ABSTRACT

MicroRNAs (miRNAs) are small non-coding RNAs that play a role in post-transcriptional regulation of gene expression in most eukaryotes. They help in fine-tuning gene expression by targeting messenger RNAs (mRNA). The interactions of miRNAs and mRNAs are sequence specific and computational tools have been developed to predict miRNA target sites on mRNAs, but miRNA research has been mainly focused on target sites within 3′ untranslated regions (UTRs) of genes. There is a need for an easily accessible repository of genome wide full length mRNA — miRNA target predictions with versatile search capabilities and visualization tools. We have created a web accessible database of miRNA target predictions for human, mouse, cow, chicken, Zebra fish, fruit fly and *Caenorhabditis elegans* using two different target prediction algorithms, The database has target predictions for miRNA's on 5′ UTRs, coding region and 3′ UTRs of all mRNAs. This database can be freely accessed at http://mamsap.it.deakin.edu.au/mirna_targets/.
© 2012 Elsevier Inc. All rights reserved.

1. Introduction

The microRNAs (miRNA) are a class of small (~22 nucleotides) non-coding RNAs that post-transcriptionally regulate gene expression by interacting with mRNAs. In animals the mRNA–miRNA interaction is semi-complementary, whereas in plants miRNAs bind with near perfect complementarity on mRNA coding regions [1]. A miRNA can interact with hundreds of genes and a gene can be targeted by many miRNAs. This results in a very high number of possible interactions. Computational approaches have been used to predict mRNA–miRNA interactions (miRanda [2], RNAhybrid, TargetScan [3,4], PITA [5], PicTar [6], RNA22 [7], microT and miRtarget etc.) [8]. These algorithms use knowledge of experimentally proven mRNA–miRNA interactions to develop a scoring system (i.e. mRNA–miRNA partial complementarity, seed region, target position, sequence conservation features etc.), which is then used to predict mRNA–miRNA interactions. Each algorithm use slightly different scoring techniques, resulting in differences in prediction results.

A number of miRNA target prediction algorithms have been developed and tested for accuracy and precision using both computational and laboratory techniques. When results from miRNA knockout experiments were compared to results from computational approaches, computational algorithms were shown to produce high false negative (undetected miRNA target genes) and false positive (nonfunctional miRNA target sites) results. One possible explanation for false negative outcomes could be that most of these studies applied computational algorithms to only 3′ UTR regions of mRNAs. It is now recognized that miRNAs can also interact with mRNAs in coding regions and 5′ UTRs as well [9–11]. Secondly, it is unlikely that all possible target sites for a miRNA will always be functional in any biological condition. Gene repression also depends on a number of other factors such as the balance between quantity, half-life and location of miRNAs and target mRNAs. Current miRNA target prediction algorithms do not take into account these important factors. In general, results from target prediction algorithms should be carefully scrutinized and should be treated only as a guide to mRNA–miRNA interactions. The construction of advanced integrated miRNA target prediction resources such as ours can help guide the development of experimental approaches to target validation and database mining will enable a more detailed analysis of the complex interactions occurring across the network of miRNAs and mRNAs.

Previously designed web servers focused on 3′ UTR targets only [2,12]. In the last few years many high throughput experiments have reported experimentally validated functional miRNA target sites located in 5′ UTR and coding region [13,14]. MiRNA target database miRWalk used 7-mer seed sequence matches as the main criteria to predict miRNA targets on mRNAs in promoter and flanking regions from human, mouse and rat species [15]. The miRTAar.human database used a combination of prediction algorithms (miRanda, TargetScan, RNAhybrid and pita) to scan full length mRNAs for predicted miRNA

* Corresponding author at: Institute for Frontier Materials, Deakin University, Geelong, Victoria 3216, Australia.
E-mail address: amitk@deakin.edu.au (A. Kumar).

0888-7543/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ygeno.2012.08.006

Table 1
Number of genes from Ensembl database and miRNAs from miRBase (release 18) in the webserver. miRNA target sites using miRanda with default settings and RNAhybrid at <0.05 P-value.

Species	Ensembl gene ids	Mature miRNA	miRanda target sites	RNAhybrid target sites
Human (GRCh37.p3)	54,283	1921	18,340,081	25,772,789
Mouse (NCBIM37)	37,681	1157	9,889,849	6,573,689
Chicken (WASHUC2)	17,934	544	2,099,138	984,979
Zebra fish (Zv9)	32,307	247	1,627,051	709,606
Cow (UMD4)	26,015	676	3,117,593	733,956
C. elegans (WS220)	45,435	368	1,375,889	765,604
Drosophila melanogaster (BDGP5.25)	14,867	430	1,359,496	1,327,560

targets using mainly the miRNA seed sequence (1–8 nt) and conservation filters. This approach is likely to achieve the best accuracy to date for conserved miRNA target sites but will miss non-conserved/species specific miRNA target sites [16]. Here we designed a web server for miRNA target predictions for mRNA 5′, 3′ UTRs and coding region using precompiled genome wide target predictions on human, mouse, cow, chicken, zebrafish, fruit fly and *Caenorhabditis elegans* using miRanda and RNAhybrid algorithms. Both of these algorithms apply commonly accepted miRNA target features and are not highly focused towards miRNA seed regions and highly conserved miRNA targets. This combination provides maximum sensitivity for target site predictions. We have incorporated versatile search capabilities and tools to help visualize results. This will provide a much needed resource for the biological research community.

2. Methods and results

2.1. Implementation

Full length mRNA sequences were downloaded from the Ensembl database using the BioMart tool [17]. Mature miRNA sequences were downloaded from miRBase (Release 18) [18]. MiRNA target prediction algorithms miRanda [2] and RNAhybrid [19] were downloaded from their respective web servers. These target prediction algorithms were used to predict miRNA targets on all sequence datasets of the respective species. Both types of target predictions use full miRNA sequence for searching target genes and are not highly conservation biased. This gives maximum sensitivity to the miRNA target search.

2.2. Database

The miRNA_Targets MySQL database stores annotated mRNA sequences and miRNA target prediction results. Target prediction results are available for *Homo sapiens, Mus musculus, Gallus gallus, Danio rerio, Bos Taurus, Drosophila melanogaster* and *C. elegans* (Table 1). This MySQL-PHP based pipeline can be extended to all the species present in the Ensembl database (Fig. 1). Ensembl gene IDs are used as the main reference in the database structure. Where multiple transcripts were available for a gene, the longest mRNA isoform was used with miRanda and for RNAhybrid miRNA targets with P-value<0.05 were selected.

2.3. Web server

The PHP-MySQL web interface allows the user to search for miRNA targets either by using a common name, Ensembl gene ID or miRBase mature miRNA ID. Users can search for miRNAs targeting a gene or group of gene IDs. The target gene list is sorted by best energy scores. A diagram in the results shows the position of miRNA targets on mRNA 5′, 3′ UTRs and coding region of each gene. MiRNAs predicted to target a gene by both algorithms are listed first, followed by miRNA predicted only by miRanda and then predicted only by RNAhybrid.

These prediction algorithms also use full-length mature miRNA sequences for target mRNA interactions, thus are not heavily seed biased and give different results for different members of a miRNA family. In contrast, the TargetScan algorithm considers only seed regions of miRNA families for greater accuracy. To test the sensitivity of target prediction algorithms we used High-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CHIPS)

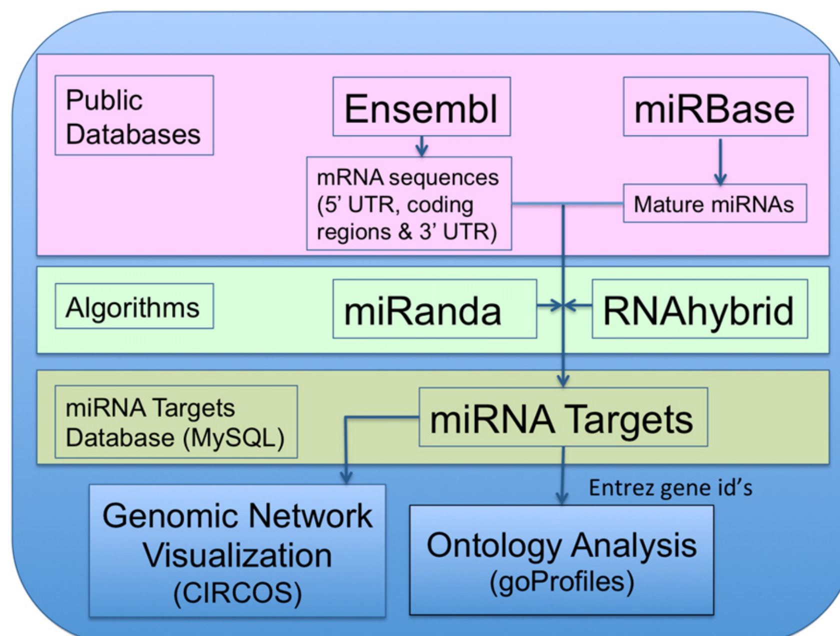


Fig. 1. Flow chart diagram of sequence datasets and algorithms used to make this database. Sequence datasets were downloaded from Ensembl and miRBase. mRNA sequences were scanned for miRNA targets using miRanda and RNAhybrid algorithms. Results were stored on MySQL database and displayed using CIRCOS and goProfiles algorithms.

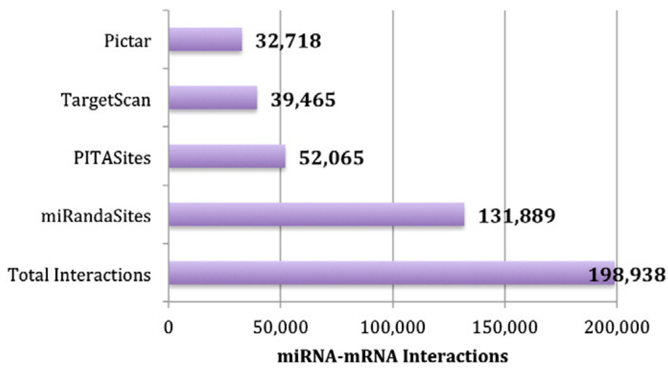


Fig. 2. Comparison of miRNA target prediction algorithms on CLIP-Seq datasets from StarBase database. Total miRNA–mRNA interactions reported in more than one CHIP-seq experiments. miRanda algorithm (applied on 3′ UTR sequences) shows the maximum coverage of miRNA target sites followed by PITA, TargetScan and Pictar.

datasets for human, mouse and *C. elegans* downloaded from the online starBase database [20]. It contains a collection of 21 CLIP-seq experiments. As shown in Fig. 2, with overall miRNA–mRNA interactions from this dataset, miRanda showed the best coverage of prediction results. Our aim was to achieve the greatest coverage of target predictions for all mature miRNAs, so that miRNA target prediction results could be used for further filtering of gene targets from microRNA overexpression and miRNA gene knockout studies.

3. Web interfaces

3.1. Genome wide miRNA target genes predictions

Single or multiple miRNA names can be used as inputs to identify target genes using miRanda or RNAhybrid prediction algorithms. MiRNA target genes can be searched by applying different energy cut-offs. The user can also restrict the miRNA target search to a list of selected genes form miRNA knockout or overexpression studies.

3.2. Using gene IDs as input

Single or multiple Ensembl gene IDs or official gene names can be used to search for miRNA targets. miRNAs are sorted first by miRNAs that have targets predicted by both algorithms then by miRanda alone, followed by those with only miRanda or RNAhybrid results.

3.3. Using gene sequences

miRanda and RNAhybrid miRNA target prediction algorithms can also be used to scan a mRNA sequence for miRNA targets by selecting all miRNAs from a given list of 17 species including viruses or using

Fig. 3. Snapshot of the miRNA_Targets database web interface. Multiple interfaces for different ways of querying miRNA target predictions. (miRNA against a given list of genes, genome wide predictions for multiple miRNAs or genes and using a mRNA sequence).

any mature miRNA sequence to pin-point the exact position of a miRNA target on a mRNA sequence. It is also possible to investigate viral miRNA interactions with host mRNAs. miRNA–mRNA alignments can be viewed on the sequence web page. Users can also screen for miRNA targets with different levels of energy stringency. Fig. 3 gives a snapshot of the database interfaces. Selecting a lower binding energy cut-off results in higher specificity, greater sequence complementarity and less sensitivity (fewer targets).

Most of the results predicted by miRanda were also predicted by RNAhybrid. This showed a good overlap in the prediction results. All output gene common names are linked to the NCBI gene database and Ensembl IDs are linked to the Ensembl database. MiRNA IDs are linked to the miRBase database.

3.4. Ontology analysis

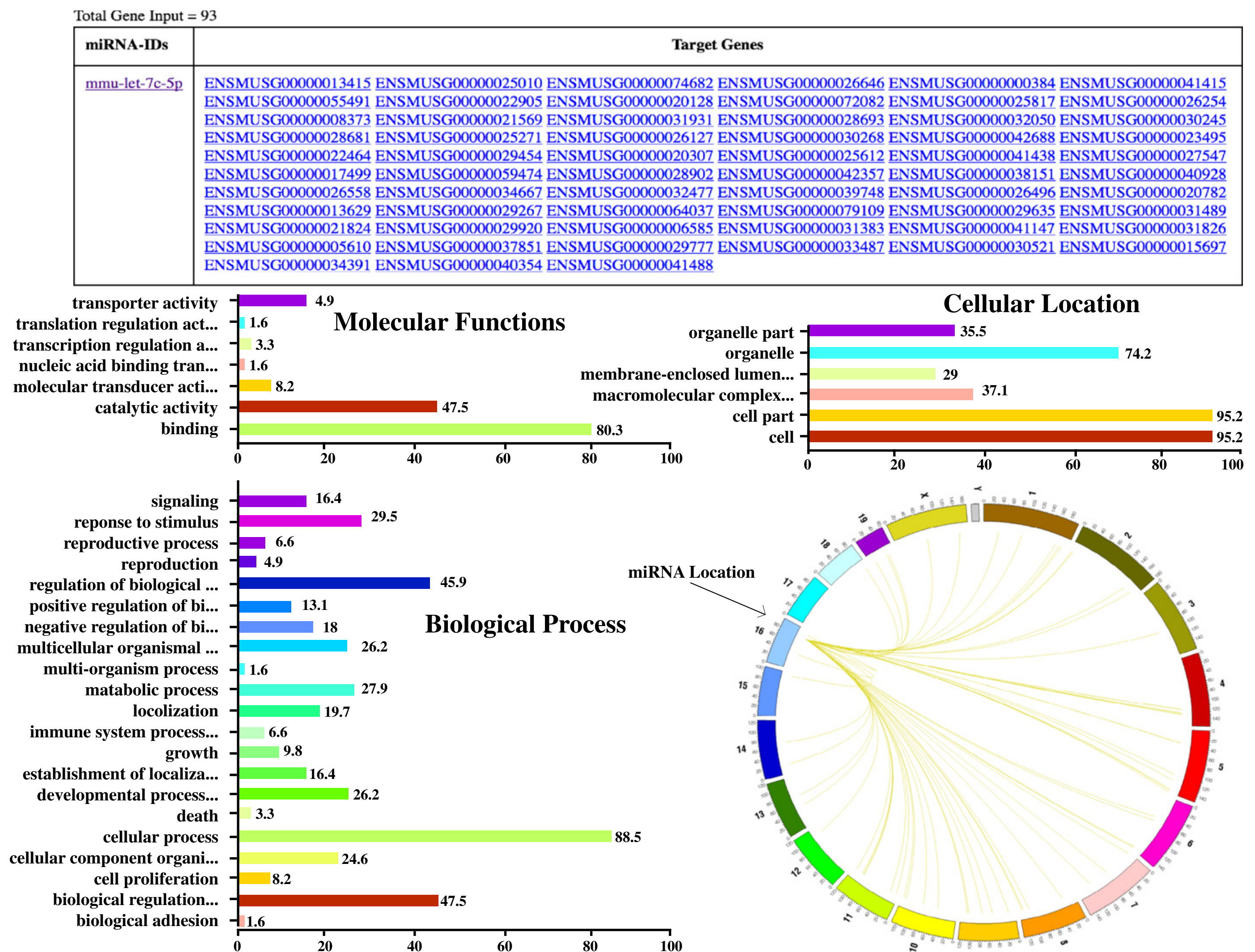
We integrated goProfiles, an R (Bioconductor) package for the functional profiling of lists of genes at the second level of Gene Ontology [21]. This package is based on the functional classification of gene ontology developed by Alex et al. [22]. Genes targeted by a miRNA or a group of miRNAs can be classified into molecular function, cellular location and biological process at the second level of GO classification. Two lists of miRNAs can also be compared against each other to get an idea of the collective ontology differences of their target genes.

3.5. MiRNA and target circular diagrams

MiRNAs are mapped to the target genes on respective chromosomes in all given species using the Circos algorithm [23]. This network visualization presents insight into any preferential targeting of certain chromosomes by particular miRNAs.

4. Experimental example

To demonstrate the value of the miRNA_Targets database we used the data from a study published by Melton et al. in 2010 [24]. In this paper it was shown that the inhibition of the let-7 miRNA family promotes de-differentiation of somatic cells to induced pluripotent stem cells. We used this microarray dataset to show the presence of miRNA target sites in 5′ UTRs and coding regions in addition to 3′ UTRs of the target genes. Expression of let-7c miRNA down regulated 694 genes. The Ensembl biomart tool was used to match the gene names to unique Ensembl gene IDs (559). By using our miRNA_Targets database, we found 488 of the 559 genes have predicted let-7c miRNA target sites. Melton et al. [24] reported only 294 genes as having miRNA target sites when they only analyzed the 3′ UTRs. From the same paper,


Total Gene Input = 93

miRNA-IDs	Target Genes
mmu-let-7c-5p	ENSMUSG00000013415 ENSMUSG00000025010 ENSMUSG00000074682 ENSMUSG00000026646 ENSMUSG00000000384 ENSMUSG00000041415 ENSMUSG00000055491 ENSMUSG00000022905 ENSMUSG00000020128 ENSMUSG00000072082 ENSMUSG00000025817 ENSMUSG00000026254 ENSMUSG00000008373 ENSMUSG00000021569 ENSMUSG00000031931 ENSMUSG00000028693 ENSMUSG00000032050 ENSMUSG00000030245 ENSMUSG00000028681 ENSMUSG00000025271 ENSMUSG00000026127 ENSMUSG00000030268 ENSMUSG00000042688 ENSMUSG00000023495 ENSMUSG00000022464 ENSMUSG00000029454 ENSMUSG00000020307 ENSMUSG00000025612 ENSMUSG00000041438 ENSMUSG00000027547 ENSMUSG00000017499 ENSMUSG00000059474 ENSMUSG00000028902 ENSMUSG00000042357 ENSMUSG00000038151 ENSMUSG00000040928 ENSMUSG00000026558 ENSMUSG00000034667 ENSMUSG00000032477 ENSMUSG00000039748 ENSMUSG00000026496 ENSMUSG00000020782 ENSMUSG00000013629 ENSMUSG00000029267 ENSMUSG00000064037 ENSMUSG00000079109 ENSMUSG00000029635 ENSMUSG00000031489 ENSMUSG00000021824 ENSMUSG00000029920 ENSMUSG00000006585 ENSMUSG00000031383 ENSMUSG00000041147 ENSMUSG00000031826 ENSMUSG00000005610 ENSMUSG00000037851 ENSMUSG00000029777 ENSMUSG00000033487 ENSMUSG00000030521 ENSMUSG00000015697 ENSMUSG00000034391 ENSMUSG00000040354 ENSMUSG00000041488

Fig. 4. let-7c miRNA targets from the down regulated set of genes with ontology analysis and chromosomal location map of the target genes on mouse genome. Percentage of genes classified in each GO category is given in front of respective GO ontology.

expression of miR-294, down regulated 1899 genes. Again, only 638 genes had predicted miRNA target sites in 3′ UTR regions compared to 1371 on full length mRNA (5′ UTR, coding region and 3′ UTRs combined). Fig. 4 shows a snapshot of predicted miRNA targets for the set of down regulated genes. Ontology classification and chromosomal locations of genes are shown on the mouse genome. This diagram shows the connecting network of miRNA-gene interactions on different chromosomes. Multiple genes on different chromosomes can be controlled simultaneously, in a sequence specific manner, by miRNA interactions. Chromosomes 4, 5 and 11 have higher densities of closely located genes interacting with let-7c miRNA. This approach can also give insights into chromosomal biases in miRNA–mRNA interactions and can highlight over-represented gene ontologies in the list of potential target genes.

MiRNA target prediction algorithms give false predictions, but if used on differentially expressed genes, we can map possible miRNA interaction sites on a given list of down regulated genes.

5. Discussion

MiRNA target prediction algorithms and publically available databases are continuously evolving. As more information about miRNA–mRNA interactions is becoming available, new publically available database tools are being developed to incorporate the new data. Here we applied the two well-known target prediction algorithms miRanda and RNAhybrid to full length mRNA sets. Variations in prediction results from different algorithms are due to different weightages of miRNA–mRNA interaction properties. The seed region is the best-known indicator of possible interaction, but this does not cover all interactions [24]. Complementarity at the 3′ end of miRNA is also known to affect miRNA target interactions [9]. Multiple experimental studies have reported that a large number of miRNA targets are present in coding regions and 5′ UTR regions of mRNA. miRNA_Targets database fills the gap in publically available databases by providing full length miRNA target site prediction for multiple species.

This database is particularly helpful for screening the differentially regulated genes from experimental studies related to miRNAs. In such studies a proportion of the down-regulated genes will be directly modulated by miRNA interactions whereas others are not subjected to miRNA interactions but rather are regulated by indirect effects of a regulatory cascade. New algorithms are required to computationally screen through a large number of genes linked in pathway, which are not direct targets of miRNAs. Currently this step can only be performed manually for individual pathways. As more and more sequencing datasets are becoming available, expression of miRNA and mRNA transcripts at multiple time points will provide further quantitative evidence of the degree of repression caused by each miRNAs. Algorithms and publically available databases have to keep up with each other for the smooth translation of bioinformatics studies to laboratory experiments.

6. Conclusions

miRNA_Targets database will provide researchers a query platform to investigate miRNA interactions in non-coding and coding regions of RNA and should promote research activity on these otherwise neglected 5′ and coding regions of mRNA. The user can query the database for precompiled miRNA targets from 7 species and miRNA target predictions can be performed for multiple genes using mature miRNAs from 17 species and viruses using miRanda and RNAhybrid algorithms. This is a more complete platform than previously available databases for the analysis of miRNA targeting biology. In the future, we will continue to update and maintain this database with addition of new miRNAs, gene annotations and incorporate more advanced open source pathway and ontology analysis algorithms as they become available.

References

[1] E. Bonnet, Y. He, K. Billiau, Y. Van de Peer, TAPIR, a web server for the prediction of plant microRNA targets, including target mimics, Bioinformatics 26 (2010) 1566–1568.
[2] B. John, A.J. Enright, A. Aravin, T. Tuschl, C. Sander, D.S. Marks, Human MicroRNA targets, PLoS Biol. 2 (2004) 1862–1879.
[3] B.P. Lewis, I.H. Shih, M.W. Jones-Rhoades, D.P. Bartel, C.B. Burge, Prediction of mammalian microRNA targets, Cell 115 (2003) 787–798.
[4] B.P. Lewis, C.B. Burge, D.P. Bartel, Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets, Cell 120 (2005) 15–20.
[5] M. Kertesz, N. Iovino, U. Unnerstall, U. Gaul, E. Segal, The role of site accessibility in microRNA target recognition, Nat. Genet. 39 (2007) 1278–1284.
[6] A. Krek, D. Grun, M.N. Poy, R. Wolf, L. Rosenberg, E.J. Epstein, P. MacMenamin, I. da Piedade, K.C. Gunsalus, M. Stoffel, N. Rajewsky, Combinatorial microRNA target predictions, Nat. Genet. 37 (2005) 495–500.
[7] K.C. Miranda, T. Huynh, Y. Tay, Y.-S. Ang, W.-L. Tam, A.M. Thomson, B. Lim, I. Rigoutsos, A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes, Cell 126 (2006) 1203–1217.
[8] M. Thomas, J. Lieberman, A. Lal, Desperately seeking microRNA targets, Nat. Struct. Mol. Biol. 17 (2010) 1169–1174.
[9] I. Lee, S.S. Ajay, J.I. Yook, H.S. Kim, S.H. Hong, N.H. Kim, S.M. Dhanasekaran, A.M. Chinnaiyan, B.D. Athey, New class of microRNA targets containing simultaneous 5 ′-UTR and 3 ′-UTR interaction sites, Genome Res. 19 (2009) 1175–1183.
[10] Y. Tay, J.Q. Zhang, A.M. Thomson, B. Lim, I. Rigoutsos, MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation, Nature 445 (2008) 1124–1128.
[11] T.A.Y.J. Robln Lytle, Joan A. Steltz, Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5′ UTR as in the 3′ UTR, PNAS 104 (2007) 9667–9672.
[12] D. Betel, M. Wilson, A. Gabow, D.S. Marks, C. Sander, The microRNA.org resource: targets and expression, Nucleic Acids Res. 36 (2008) D149–D153.
[13] A.L.-M. Joshua, J. Forman, Hilary A. Coller, A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence, PNAS 105 (2008) 14879–14884.
[14] X. Zhou, X.C. Duan, J.J. Qian, F. Li, Abundant conserved microRNA target sites in the 5′-untranslated region and coding sequence, Genetica 137 (2009) 159–164.
[15] H. Dweep, C. Sticht, P. Pandey, N. Gretz, MiRWalk — database: prediction of possible miRNA binding sites by “walking” the genes of three genomes, J. Biomed. Inform. 44 (2011) 839–847.
[16] J.B. Hsu, C.M. Chiu, S.D. Hsu, W.Y. Huang, C.H. Chien, T.Y. Lee, H.D. Huang, miRTar: an integrated system for identifying miRNA-target interactions in human, BMC Bioinforma. 12 (2011) 300.
[17] S. Haider, B. Ballester, D. Smedley, J. Zhang, P. Rice, A. Kasprzyk, BioMart central portal — unified access to biological data, Nucleic Acids Res. 37 (2009) W23–W27.
[18] S. Griffiths-Jones, miRBase: microRNA sequences and annotation, Curr. Protoc. Bioinformatics (2010) (Chapter 12 Unit 12 19 11–10).
[19] J. Kruger, M. Rehmsmeier, RNAhybrid: microRNA target prediction easy, fast and flexible, Nucleic Acids Res. 34 (2006) W451–W454.
[20] J.H. Yang, J.H. Li, P. Shao, H. Zhou, Y.Q. Chen, L.H. Qu, starBase: a database for exploring microRNA–mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data, Nucleic Acids Res. 39 (2011) D202–D209.
[21] Alex Sanchez, Jordi Ocana, M. Salicru, goProfiles: An R Package for the Statistical Analysis of Functional Profiles, R Package Version 1.14.0, 2010.
[22] A. Sánchez-Pla, M. Salicrú, J. Ocaña, Statistical methods for the analysis of high-throughput data based on functional profiles derived from the gene ontology, J. Stat. Plan. Infer. 137 (2007) 3975–3989.
[23] M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones, M.A. Marra, Circos: an information aesthetic for comparative genomics, Genome Res. 19 (2009) 1639–1645.
[24] C. Melton, R.L. Judson, R. Blelloch, Opposing microRNA families regulate self-renewal in mouse embryonic stem cells, Nature 436 (2010) 621–626.