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)

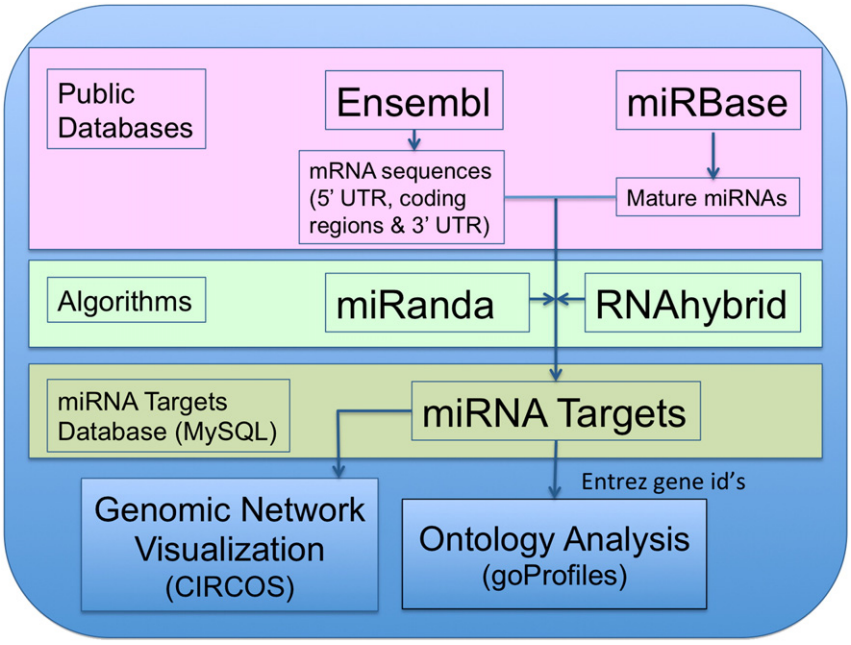

Fig. 1. Flow chart diagram of sequence datasets and algorithms used to make this database. Sequence datasets were downloaded from Ensembl and miRBase. mRNA sequences were scanned for miRNA targets using miRanda and RNAhybrid algorithms. Results were stored on MySQL database and displayed using CIRCOS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

Download English Version:

https://daneshyari.com/en/article/5907951

Download Persian Version:

https://daneshyari.com/article/5907951

Daneshyari.com