TechNote TN-04 July 2018

Title: Web-based tools collection for microRNA analysis

Since the discovery of microRNAs, there has been a release of number of tools useful for both basic and advanced applications. In this technote we describe over 20 microRNA database, target prediction or sequencing tools (Figure 1) with short explanations of the content to help to find the right tool for your needs.

Figure 1: Webbased-tools

1. microRNA target prediction tools

An extremely large number of potential target sites exists for any given microRNA, and the process of validating a potential microRNA target in the laboratory is time consuming and costly. A computational approach to prediction of microRNA targets facilitates the process of narrowing down potential target sites for experimental validation. Computational approaches model how microRNAs target specific mRNAs and an increasing collection of tools is available, each with a distinct approach to microRNA target prediction.

There are four commonly used features for microRNA target prediction tools [1] :

Seed match: microRNAs regulate the gene expression by binding to the mRNA. The seed sequence is essential for the binding of the microRNA to the mRNA. The seed sequence or seed region is a conserved heptametrical sequence which is mostly situated at positions 2-7 from the microRNA 5´-end. Even though base pairing of microRNA and its target mRNA does not match perfect, the "seed sequence" has to be perfectly complementary.

Conservation: Solutions to reduce the number of false-positives in target prediction include filtering out those binding sites that do not appear to be conserved across species. Many of the initial microRNA target sites identified are indeed highly conserved across species [2].

Free energy: Free energy (or Gibbs free energy) can be used as a measure of the stability of a biological system. By predicting how the microRNA and its candidate target hybridize, regions of high and low free energy can be inferred and the overall the change in free energy during a reaction (ΔG) can be used as an indicator of how strongly bound they are [3].

Site accessibility: Site accessibility is a measure of the ease with which a microRNA can locate and hybridize with an mRNA target. Following transcription, mRNA assumes a secondary structure [4] which can interfere with a miRNA's ability to bind to a target site. To assess the likelihood that an mRNA is the target of a microRNA, the predicted amount of energy required to make a site accessible to a microRNA can be evaluated.

Below online microRNA target prediction tools are listed.

miRDB (http://mirdb.org/)

miRDB is an online database for miRNA target prediction and functional annotations. All the targets in miRDB were predicted by a bioinformatics tool, MirTarget, which was developed by analyzing thousands of miRNA-target interactions from high-throughput sequencing experiments. Common features associated with miRNA target binding have been identified and used to predict miRNA targets with machine learning methods. miRDB hosts predicted miRNA targets in five species: human, mouse, rat, dog and chicken.

TargetScan (http://www.targetscan.org/vert_72/)

TargetScan predict regulatory targets of vertebrate microRNAs (miRNAs) by identifying mRNAs with conserved complementarity to the seed (nucleotides 2-7) of the miRNA. An over-representation of conserved adenosines flanking the seed complementary sites in mRNAs indicates that primary sequence determinants can supplement base pairing to specify miRNA target recognition.

PicTar (https://pictar.mdc-berlin.de/)

PicTar is a computational method for identifying common targets of microRNAs. Statistical tests using genome-wide alignments of eight vertebrate genomes, PicTar's ability to specifically recover published microRNA targets, and experimental validation of seven predicted targets suggest that PicTar has an excellent success rate in predicting targets for single microRNAs and for combinations of microRNAs.

RNA-hybrid (https://bibiserv.cebitec.uni-bielefeld.de/rnahybrid)

RNA-hybrid predicts multiple potential binding sites of miRNAs in large target RNAs. In general, the program finds the energetically most favorable hybridization sites of a small RNA in a large RNA.

miRIAD (http://bmi.ana.med.uni-muenchen.de/miriad/)

miRIAD is a web search tool developed with the primary purpose of integrating relevant information concerning intragenic miRNAs and their host genes.

DIANA-mirPath (http://snf-515788.vm.okeanos.grnet.gr/)

DIANA-mirPath can utilize predicted miRNA targets (in CDS or 3'-UTR regions) provided by the DIANA-microT-CDS algorithm or even experimentally validated miRNA interactions derived from DIANA-TarBase. These interactions (predicted and/or validated) can be subsequently combined with sophisticated merging and meta-analysis algorithms.

miRTar.human (http://mirtar.mbc.nctu.edu.tw/human/)

miRTar is a tool that enables easily to identify the biological functions and regulatory relationships between a group of known/putative miRNAs and protein coding genes. It also provides perspective of information on the miRNA targets on alternatively spliced transcripts.

miRmap (http://mirmap.ezlab.org/)

Includes experimentally verified miRNAs and their targets, also contains RNAhybrid and TargetScan target prediction module.

StarMir (http://sfold.wadsworth.org/cgi-bin/starmir.pl)

Software for statistical folding of nucleic acids and studies of regulatory RNAs based on CLIP data (sequence, thermodynamic and target structure features).

ComiRNet (http://comirnet.di.uniba.it:8080/)

ComiRNet uses 10 miRNA target prediction databases, stores approximately 5 million predicted interactions between 934 human miRNAs and 30,875 gene transcripts (mRNAs) which are exploited in the construction of the hierarchies of overlapping biclusters representing potential miRNA regulatory networks.

VIRmiRNA (http://crdd.osdd.net/servers/virmirna/)

VIRmiRNA is the first dedicated resource on experimental viral miRNA and their targets. VIRmiRNA is a resource for experimentally validated viral miRNA and its target reported in literature. It comprises of three sub-databases-VIRmiRNA, VIRmiRtar and AVIRmir. VIRmiRNA contains 1308 miRNA entries encoded by 44 viruses.

2. microRNA databases

MicroRNAs affect nearly all types of cellular pathways, from development to oncogenesis [5]. Clearly current microRNA research is not limited to their biogenesis and function. Their clinical implications are now a very topical research issue, because they have been hypothesized to be diagnostic and prognostic biomarkers and therapeutic targets for different human diseases

including cancer [6]. Given their involvement in gene regulation as well as disease processes, experiments have increased, generating an exponential flow of data scattered in thousands of articles (Figure 2). A large number of bioinformatic tools are now available to manage the mounting data flow. As listet below, most applications are accessible online.

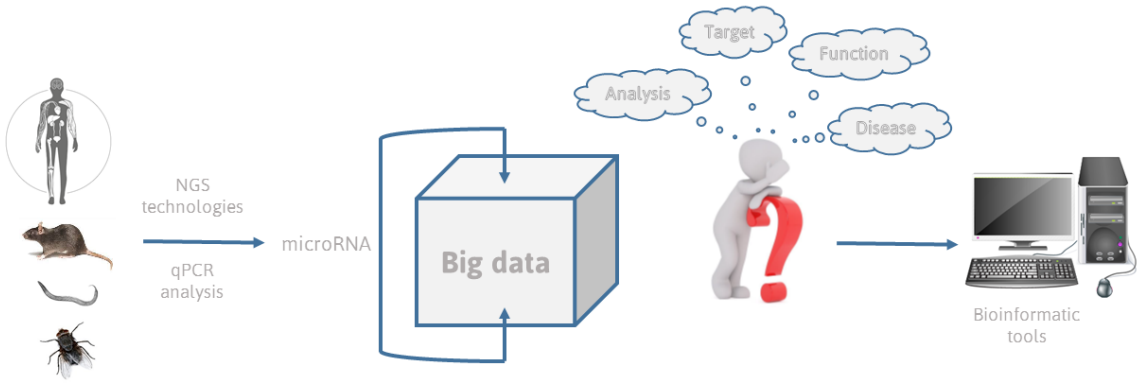

Figure 2: Illustration of the complexity of large data sets and the need for bioinformatic tools.

General purpose databases

miRBase (http://www.mirbase.org/index.shtml)

The miRBase database is a searchable database of published miRNA sequences and annotation. Each entry in the miRBase Sequence database represents a predicted hairpin portion of a miRNA transcript (termed mir in the database), with information on the location and sequence of the mature miRNA sequence (termed miR). Both hairpin and mature sequences are available for searching and browsing, and entries can also be retrieved by name, keyword, references and annotation. All sequence and annotation data are also available for download.

DIANA tools (http://carolina.imis.athena-innovation.gr/diana_tools/web/index.php?r=site%2Ftools)

DIANA-TarBase v8.0 is a reference database devoted to the indexing of experimentally supported microRNA (miRNA) targets. Its eighth version is the first database indexing >1 million entries, corresponding to ~670 000 unique miRNA-target pairs. The interactions are supported by >33 experimental methodologies, applied to ~600 cell types/tissues under ~451 experimental conditions. It integrates information on cell-type specific miRNA-gene regulation, while hundreds of thousands of miRNA-binding locations are reported.

Specialized databases

miRBase Tracker (http://www.mirbasetracker.org/)

An easy-to-use online database that keeps track of all historical and current miRNA annotation present in the miRBase database.

FirePlex Discovery Engine (https://www.fireflybio.com/portal/search)

The FirePlex Discovery Engine assembles a list of the most important microRNAs and associated genes from the scientific literature for any keyword or topic, putting years of research at your fingertips in seconds.

miROrtho (http://cegg.unige.ch/mirortho)

miROrtho contains predictions of precursor miRNA genes covering several animal genomes combining orthology and a Support Vector Machine. It provides homology extended alignments of already known miRBase families and putative miRNA families.

miRCancer : microRNA Cancer Association Database (http://mircancer.ecu.edu/)

miRCancer provides comprehensive collection of microRNA (miRNA) expression profiles in various human cancers which are automatically extracted from published literatures in PubMed. Search by cancer names or the on-site sequence analysis tools is possible.

miRWalk (http://mirwalk.umm.uni-heidelberg.de/)

The new version of miRWalk stores predicted data obtained with a maschine learning algorithm including experimentally verified miRNA-target interactions. The focus lies on accuracy, simplicity, user-friendly design and mostly up to date informations.

omiRas http://tools.genxpro.net/omiras/)

omiRas provides an interface to generate a miRNA-mRNA interaction network for user selected microRNAs. Interactions between genes and miRNAs are determined by the interesection of 7 databases (TargetScan, miRanda, PICTAR, PITA, miRDB, TarBase, miRConnect).

---

3. microRNA deep sequencing tools

Profiling of microRNAs by deep sequencing measures absolute abundance and allows for the discovery of novel microRNAs that have eluded previous cloning and standard sequencing efforts. Public databases provide in silico predictions of microRNA gene targets by various algorithms [7].

miRDeep2 (https://www.mdc-berlin.de/n-rajewsky#t-data,software&resources)

miRDeep2 is a completely overhauled tool which discovers microRNA genes by analyzing sequenced RNAs.

miRNAkey (http://ibis.tau.ac.il/miRNAkey/)

miRNAkey is a software package designed to be used as a base-station for the analysis of miRNA deep sequencing data.

---

4. microRNA data analyses tools

A series of microarray experiments produces observations of differential expression for thousands of genes across multiple conditions. It is often not clear whether a set of experiments are measuring fundamentally different gene expression states or are measuring similar states created through different mechanisms. It is useful, therefore, to define a core set of independent features for the expression states that allow them to be compared directly [8].

Principal components analysis (PCA): is a statistical technique for determining the key variables in a multidimensional data set that explain the differences in the observations, and can be used to simplify the analysis and visualization of multidimensional data sets.

microRNA heat maps**:** in heat maps the data is displayed in a grid where each row represents a microRNA and each column represents a sample. The colour and intensity of the boxes is used to represent changes (not absolute values) of microRNA expression.

Venn diagrams**:** are graphical ways of representing interactions among sets to display information that can be read easily.

ClustVis (https://biit.cs.ut.ee/clustvis/)

This web tool allows users to upload their own data and easily create Principal Component Analysis (PCA) plots and heatmaps.

Venny (http://bioinfogp.cnb.csic.es/tools/venny/)

An interactive tool for comparing lists using Venn Diagrams with support for up to 4 sets.

---

5. Others web-based tools

Tools4miRs (https://tools4mirs.org/)

Tools4miRs is a first, manually curated platform gathering at the present over 170 methods for the broadly-defined miRNA analysis. All tools in Tools4miRs are classified in the four general and seven more detailed categories. In each of the aforementioned sections user can additionally filter available methods according to his research needs, capabilities and preferences.

TissueAtlas (https://ccb-web.cs.uni-saarland.de/tissueatlas/)

TissueAtlas allows identifying the tissues of origin of miRNAs and provides insights into specificity and heterogeneity of miRNAs with respect to tissues.

GeneCards (https://www.genecards.org/)

GeneCards is a searchable, integrative database that provides comprehensive, user-friendly information on all annotated and predicted human genes. It automatically integrates gene-centric data from ~125 web sources, including genomic, transcriptomic, proteomic, genetic, clinical and functional information.

The GeneCards Suite of Databases

MalaCards: A database of human maladies and their annotations
PathCards: The integrated database of human pathways and their annotations
GeneALaCart: Generates a file of GeneCards annotations for your list of genes
GeneLoc: Presents an integrated map for each human chromosome, based on data integrated by the GeneLoc algorithm

BloodcellAtlas (http://134.245.63.235/ikmb-tools/bloodmiRs/)

A comprehensive reference dataset of detailed miRNA expression profiles from seven types of human peripheral blood cells, serum, exosomes and whole blood. The overall dataset was generated from 450 small RNA libraries using high-throughput sequencing.

6. Overview of the web-based tools

Table 1 provides an overview of the tool collection and their features.

tool name	miRNA database	target prediction algorithm	text mining	tissue expression atlas	miRNA data/sequence analysis	organism
miRDB	•	•				human, rat, mouse, chicken and dog
TargetScan	•	•				human, rat, mouse, and many more
PicTar		•				vertebrate and fly
RNA-hybrid		•				human, fly, worm
miRIAD		•				human, rat, mouse, and many more
DIANA-mirPath		•				human, rat, mouse, and many more
miRTar.human		•				human
miRmap		•				human, rat, mouse, and many more
StarMir	•					human, mouse, worm
ComiRNet		•				human
VIRmiRNA	•	•				virus
miRBase	•					human, viruses, plants and many more
miRBase Tracker	•					human, viruses, plants and many more
FirePlex Discovery Engine	•		•			human, rat, mouse, and many more
miROrtho						human, mouse,fly
DIANA tools	•	•				human, rat, mouse, and many more
miRCancer			•			human
miRWalk	•	•				human, rat, mouse
omiRas	•					human, mouse and many more
miRDeep2					•	human, mouse and five more
miRNAkey					•	human and many more
ClustVis					•	
Venny					•	
Tools4miRs						tool collection
TissueAtlas				•		human - 61 tissue biopsie
GeneCards	•			•		
BloodcellAtlas				•		human

Table 1: Overview of the features of the online tools

References:

[1] S.M. Peterson, J.A. Thompson, M.L. Ufkin, P. Sathyanarayana, L. Liaw, C.B. Congdon, Common features of microRNA target prediction tools, Front. Genet. 5 (2014) 1–10. doi:10.3389/fgene.2014.00023.

[2] 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 (2010) 15–20. doi:10.1016/j.cell.2004.12.035.

[3] D. Yue, H. Liu, Y. Huang, Survey of Computational Algorithms for MicroRNA Target Prediction, Curr. Genomics. 10 (2009) 478–492. doi:10.2174/138920209789208219.

[4] E.M. Mahen, P.Y. Watson, J.W. Cottrell, M.J. Fedor, mRNA secondary structures fold sequentially but exchange rapidly in vivo, PLoS Biol. 8 (2010). doi:10.1371/journal.pbio.1000307.

[5] B. Zhang, X. Pan, Q. Wang, G.P. Cobb, T.A. Anderson, Computational identification of microRNAs and their targets, Comput. Biol. Chem. 30 (2006) 395–407. doi:10.1016/j.compbiolchem.2006.08.006.

[6] W.C.S. Cho, MicroRNAs: Potential biomarkers for cancer diagnosis, prognosis and targets for therapy, Int. J. Biochem. Cell Biol. 42 (2010) 1273–1281. doi:10.1016/j.biocel.2009.12.014.

[7] C.J. Creighton, J.G. Reid, P.H. Gunaratne, Expression profiling of microRNAs by deep sequencing, Brief. Bioinform. 10 (2009) 490–497. doi:10.1093/bib/bbp019.

[8] S. Raychaudhuri, J.M. Stuart, R.B. Altman, 2E- Principal components analysis to summarize microarray experiments: application to sporulation time series., Pac. Symp. Biocomput. (2000) 455–66.