Supplementary Data

Tools4miRs – one place to gather all the tools for miRNA analysis

Anna Lukasik[1], Maciej Wójcikowski[1] and Piotr Zielenkiewicz[1,2]

[1] Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawinskiego 5a, 02-106 Warsaw, Poland
[2] Department of Plant Molecular Biology, Institute of Experimental Plant Biology and Biotechnology, University of Warsaw, Miecznikowa 1, 02-096 Warsaw, Poland

Data Collection

During the collection of computational methods for the **tools4miRs** platform, all tools have been manually verified for several terms, such as, availability of reference publication (if tool was not published it was excluded from further verification), stated functionality (if some troubles occurred during sample data testing, tool was passed for later evaluation again), collected data (in case of Databases), requirement of the next-generation sequencing (NGS) data, requirement of the reference genome sequence, availability of the user support, ease of tool use/installation, cross-references and others. All gathered tools in the **tools4miRs** were categorized into four general and seven specialized categories, which can be queried via the following menus:

- Databases - wide range of miRNA-related data collections, such as miRNA sequences, experimentally validated targets, predicted targets, miRNA expression level in different cells/tissues/conditions, targets expression level, disease associations and several others.
- Sequencing Data Analysis - methods specifically dedicated to the examination of high-throughput sequencing data (e. g. sRNA-Seq, RNA-Seq, PARE and CLIP-Seq) generated from different sequencing platforms. Mentioned tools provide specific analysis concerning features like: novel and known miRNA identification, target prediction and annotation or differential expression investigation.
- All-In-One Tools – this section is similar to the Sequencing Data Analysis category, however, tools gathered here, instead of focusing the analysis on specific features (e.g. only known miRNA identification and differential expression), perform complex analysis of the high-throughput sequencing data.
- Software categories – tools collected in tools4miRs are classified in the below sub-sections depending on their functionality. These computational methods may or may not require (as an input) the high-throughput sequencing data:

  - Known miRNA identification – bioinformatics approaches designed to search for known miRNAs in the provided data sets (in most cases these are deep sequencing data sets).
  - isomiRs identification – tools useful for identification of isomiRs, that is molecules having variations in sequence length/composition with respect to the reference, known miRNAs. Described type of methods requires (usually) results from the sRNA deep sequencing.
  - Novel miRNA / Precursor Analysis – computational approaches gathered here focus on the novel miRNA molecules identification or distinguishing real miRNA precursors from the pseudo pre-miRNAs. The analyses performed by these tools are based on the divergent algorithms and consider specific features such as structure, sequence and conservation.
  - Differential Expression Analysis – computational approaches which evaluate miRNAs' or their targets expression differences between given sets. Performed analysis can be based on known differential expression algorithms, such as, edgeR and DESeq2.
  - Target Prediction – bioinformatics methods designed for the identification of potential miRNA binding sites in a target mRNA sequence. The aforementioned tools incorporate several common (seed match, conservation, free energy or site-accessibility) and less common (target-site abundance, local AU content or 3' compensatory pairing) features on which miRNA target prediction is based.
  - Target Functional Annotation – useful tools enabling the user to perform functional analysis of given miRNA targets. This may include: the GO (Gene Ontology) and DO (Disease Ontology) terms annotation/enrichment examination, metabolic KEGG pathway analysis, expression correlation between miRNA and their targets, literature mining and others.
  - miRNA-SNP analysis – methods which examine influence of single nucleotide polymorphisms (SNP) on the miRNAs and their interaction with potential targets. Certain tools from this section may require already predicted targets for a given miRNAs.
- Other tools - methods which are interesting and relevant from the miRNA-analysis point of view; however, basing on their main scope they could not be classified into any of the available general or detailed categories.

The classification of several gathered tools may be redundant– same method maybe presented in few categories. The best example showing that are computational approaches from the "Sequencing Data Analysis" section. Mentioned tools perform on the NGS data a pipeline of processes throughout which they for e.g. identify known/novel miRNAs, miRNA precursors and targets, as well as evaluate miRNAs expression differences between samples. Therefore, depending on these functional features, bioinformatic approaches from the "Sequencing Data Analysis" section are also classified into the available detailed categories, such as "Known miRNA Identification", "isomiR Identification" and "Differential Expression Analysis", however, with the underlined information that the NGS data is needed as an input.

Working Example

A simple example will show great functionality and ease of use of the **tools4miRs** service to achieve the users' goal – namely, identifying tools for the mouse miRNA targets functional analysis. To expand description of the presented task, the user does not have any NGS data available nor predicted targets for his/hers selected miRNAs. The user prefers to apply a web-based (online) tool, which must include in the analysis at least the Gene Ontology (GO) terms annotation and the KEGG metabolic pathways examination. In order to obtain desired computational methods the user should go to the "Target Functional Analysis" section where throughout the "Filer Panel" he/she can filter available tools according to the selected appropriate criteria: Organisms Specificity ("Mouse"), Analysis Features ("KEGG Pathways" and "GO Terms"), Tool Availability ("Online"), NGS Data Needed ("No") and Target Prediction Data Needed ("No"). As a result of the aforementioned searching the user will retrieve table with 3 computational methods, while the starting number of tools was 21. The final decision of choosing desired bioinformatic approach can be made based on the method's description, number of original reference citations, ease of usage, date of last software update and others. The illustration of the example described above (searching and results) is presented as Screen 1.

Implementation

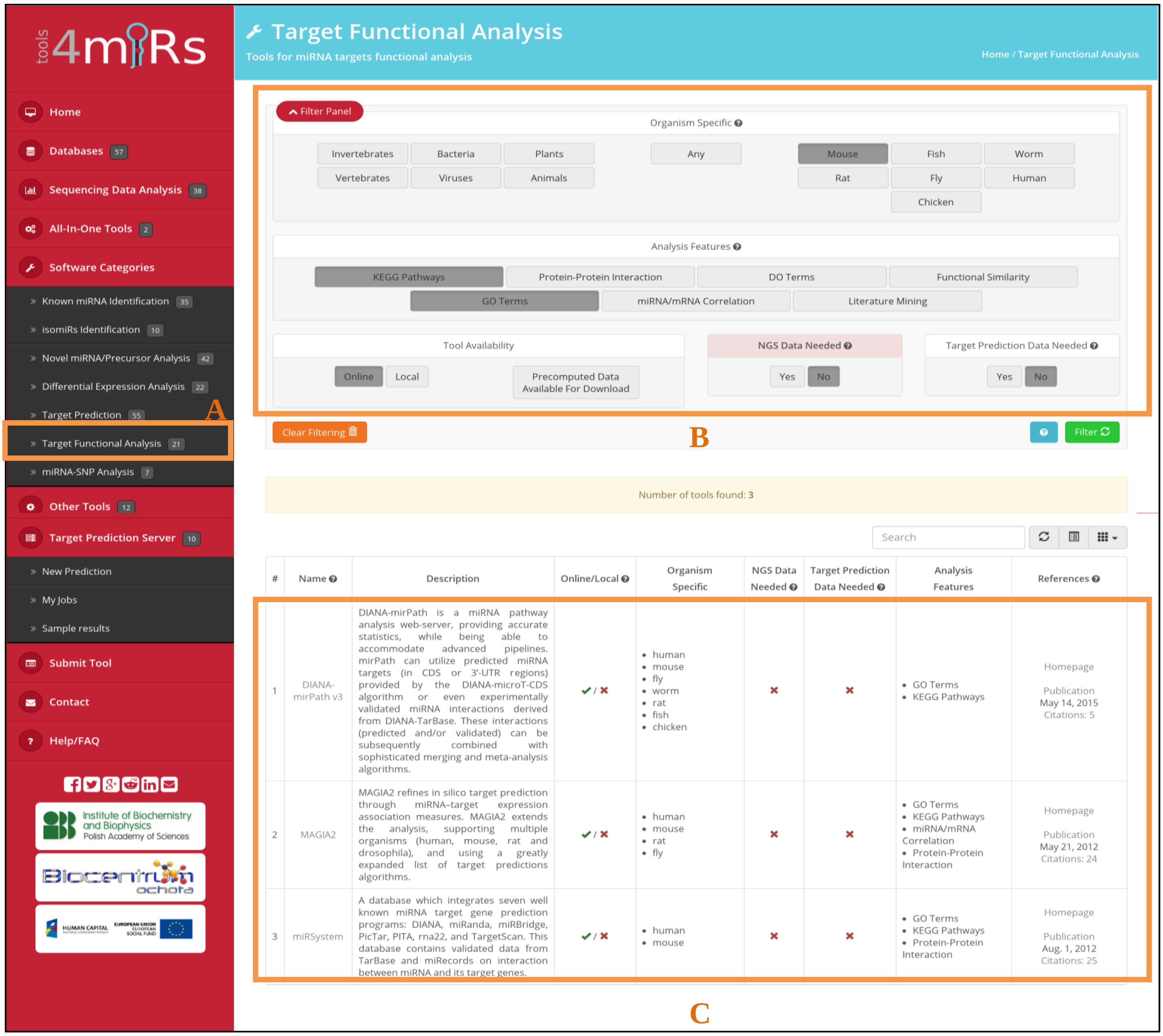
Tools4miRs was coded in Python using the Django framework. The HTML frontend is mainly powered by Bootstrap (http://getbootstrap.com/), Bootstrap Tables (http://bootstrap-table.wenzhixin.net.cn/) and jQuery (https://jquery.com/). The computing subsection (target prediction meta-server) of the **tools4miRs** is driven by Python and employs the Slurm cluster management software. BioPython and additional filters are used for the input data validation and pre-processing to ensure data quality for further steps. Uniform interaction with underlying target prediction software is possible through the additional Python class wrappers, which united all of the programs under common API. The miRNA/target sequences provided by the user serve as an input for the selected target prediction bioinformatic approaches. The computations are done separately for each method and afterwards generated outputs are parsed for hits. Since the predicted target binding site ranges obtained from various tools (for a single unique miRNA:target pair) may differ slightly, the additional step o merging these binding site ranges is introduced. The larger site include the smaller ones that it covers, but only if they are fully within the larger range. Otherwise, mentioned target binding sites (of the same unique miRNA:target pair) are treated as separate ones. The Pandas library handles data collection and post-processing, which includes the on-demand post-processing available on the "Job Status Page". This enables the user to tailor computed data to his/her needs (custom logic, methods union/intersection, sorting, cut-offs, etc.). The Pandas is additionally used for the result exporting to CSV, XLS and XLSX formats. The unprocessed outputs from all used target prediction software are also available for download.

User-Interactive Features

To make **tools4miRs** more user-interactive, an interface has been developed that allows the addition of comments and ratings for a given tool. Mentioned "Comments & Ratings" section is available on the bottom of every page that provides detailed, individual description of a specific tool. The users willing to add to the **tools4miRs** their already published computational method are able to do it through the "Submit Tool" form or by contacting the authors directly. After data validation, the submitted method will be incorporated into the service and the meta-server, if possible.

Screens

Screen 1. A simple example describing functionality of the **tools4miRs** service – searching a desired tool for mouse miRNA target functional analysis. Pipeline: (A) Choosing appropriate category – "Target Functional Analysis" and (B) Choosing appropriate criteria throughout the "Filter Panel". Results: (C) Table with the bioinformatic method meeting the users' requirements, preferences and needs. Screen on the next page.

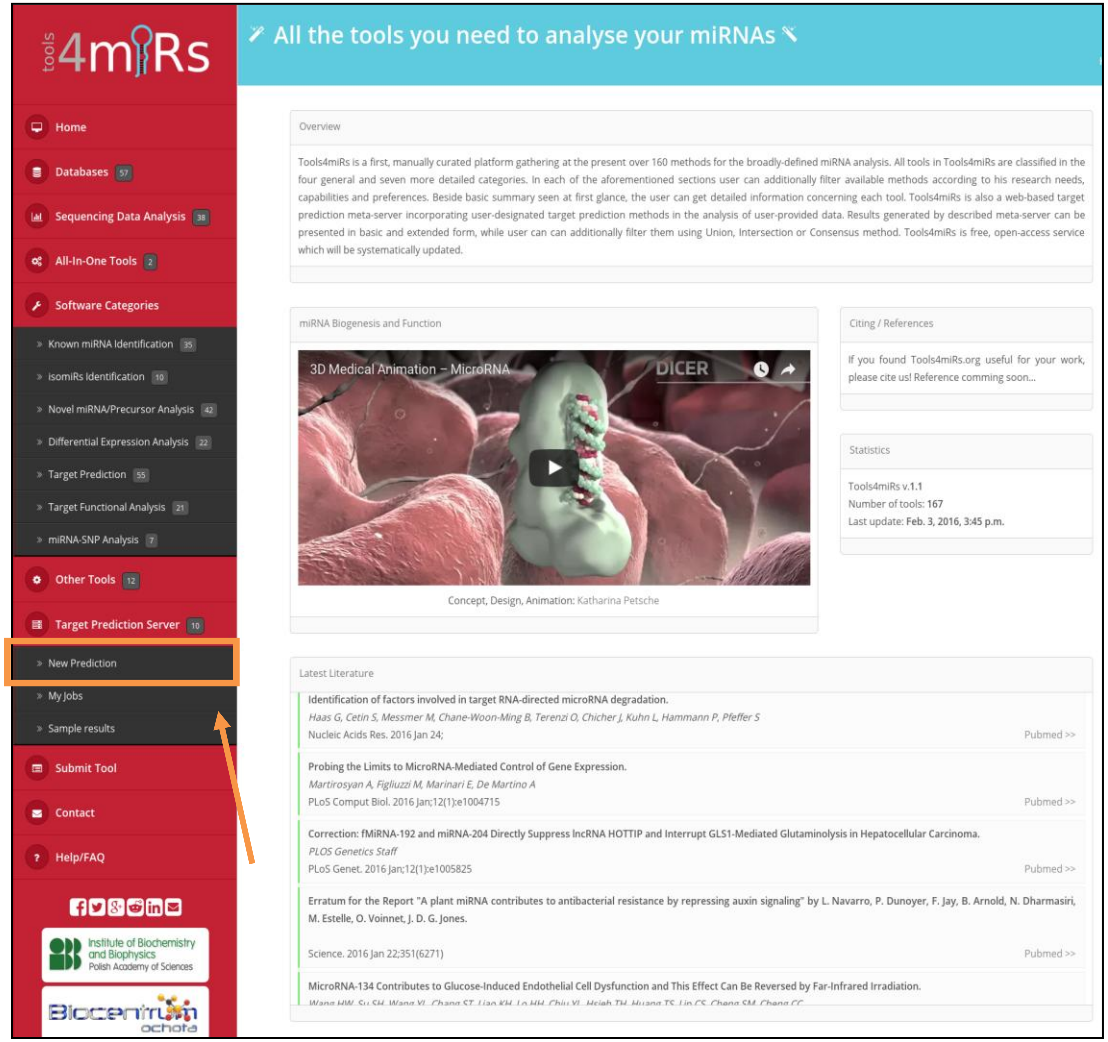


Number of tools found: 3

#	Name	Description	Online/Local	Organism Specific	NGS Data Needed	Target Prediction Data Needed	Analysis Features	References
1	DIANA-mirPath v3	DIANA-mirPath is a miRNA pathway analysis web-server, providing accurate statistics, while being able to accommodate advanced pipelines. 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.	✓ / ✗	human mouse fly worm rat fish chicken	✗	✗	GO Terms KEGG Pathways	Homepage Publication May 14, 2015 Citations: 5
2	MAGIA2	MAGIA2 refines in silico target prediction through miRNA–target expression association measures. MAGIA2 extends the analysis, supporting multiple organisms (human, mouse, rat and drosophila), and using a greatly expanded list of target predictions algorithms.	✓ / ✗	human mouse rat fly	✗	✗	GO Terms KEGG Pathways miRNA/mRNA Correlation Protein-Protein Interaction	Homepage Publication May 21, 2012 Citations: 24
3	miRSystem	A database which integrates seven well known miRNA target gene prediction programs: DIANA, miRanda, miRBridge, PicTar, PITA, rna22, and TargetScan. This database contains validated data from TarBase and miRecords on interaction between miRNA and its target genes.	✓ / ✗	human mouse	✗	✗	GO Terms KEGG Pathways Protein-Protein Interaction	Homepage Publication Aug. 1, 2012 Citations: 25

Screen 2. Workflow of the miRNA target prediction using the **tools4miRs** meta-server: (A) Choosing "New Prediction" from the main menu, (B) Selecting desired tools from available ones (or filtering them with the "Filter Panel" beforehand), (C) Providing miRNAs and mRNA sequences (in FASTA format) for the analysis. Optionally, providing job name, description and e-mail address. Setting parameters for the selected targets prediction methods or leaving them as default, (D) Waiting for calculations to finish or for e-mail with notification that submitted job has completed, (E) Examination of the obtained results from the Basic and (F) Extended analysis mode. The number of results generated from the Basic and Extended analysis modes may differ even significantly. The explanation for this is that the Basic mode denotes unique pairs of miRNA and mRNAs, while the Extended mode treats triplet (miRNA, mRNA and position in given mRNA) unique. Screen on the next three pages.

A



All the tools you need to analyse your miRNAs

Overview

Tools4miRs is a first, manually curated platform gathering at the present over 160 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. Beside basic summary seen at first glance, the user can get detailed information concerning each tool. Tools4miRs is also a web-based target prediction meta-server incorporating user-designated target prediction methods in the analysis of user-provided data. Results generated by described meta-server can be presented in basic and extended form, while user can can additionally filter them using Union, Intersection or Consensus method. Tools4miRs is free, open-access service which will be systematically updated.

miRNA Biogenesis and Function

Concept, Design, Animation: Katharina Petsche

Citing / References

If you found Tools4miRs.org useful for your work, please cite us! Reference comming soon...

Statistics

Tools4miRs v.1.1
Number of tools: 167
Last update: Feb. 3, 2016, 3:45 p.m.

Latest Literature

Identification of factors involved in target RNA-directed microRNA degradation.
Haas G, Cetin S, Messmer M, Chane-Woon-Ming B, Terenzi O, Chicher J, Kuhn L, Hammann P, Pfeffer S
Nucleic Acids Res. 2016 Jan 24; Pubmed >>

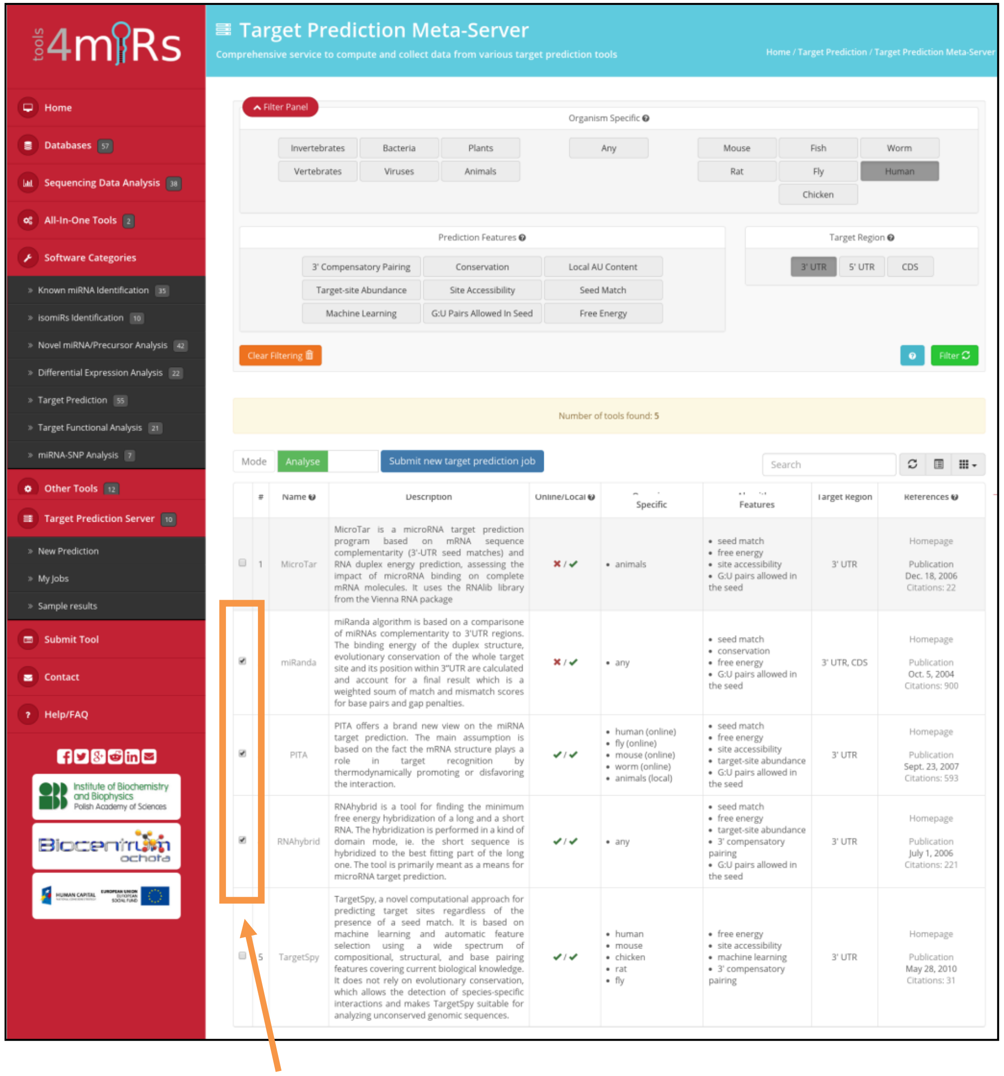
Probing the Limits to MicroRNA-Mediated Control of Gene Expression.
Martirosyan A, Figliuzzi M, Marinari E, De Martino A
PLoS Comput Biol. 2016 Jan;12(1):e1004715 Pubmed >>

Correction: fMiRNA-192 and miRNA-204 Directly Suppress lncRNA HOTTIP and Interrupt GLS1-Mediated Glutaminolysis in Hepatocellular Carcinoma.
PLOS Genetics Staff
PLoS Genet. 2016 Jan;12(1):e1005825 Pubmed >>

Erratum for the Report "A plant miRNA contributes to antibacterial resistance by repressing auxin signaling" by L. Navarro, P. Dunoyer, F. Jay, B. Arnold, N. Dharmasiri, M. Estelle, O. Voinnet, J. D. G. Jones.
Science. 2016 Jan 22;351(6271) Pubmed >>

MicroRNA-134 Contributes to Glucose-Induced Endothelial Cell Dysfunction and This Effect Can Be Reversed by Far-Infrared Irradiation.

B



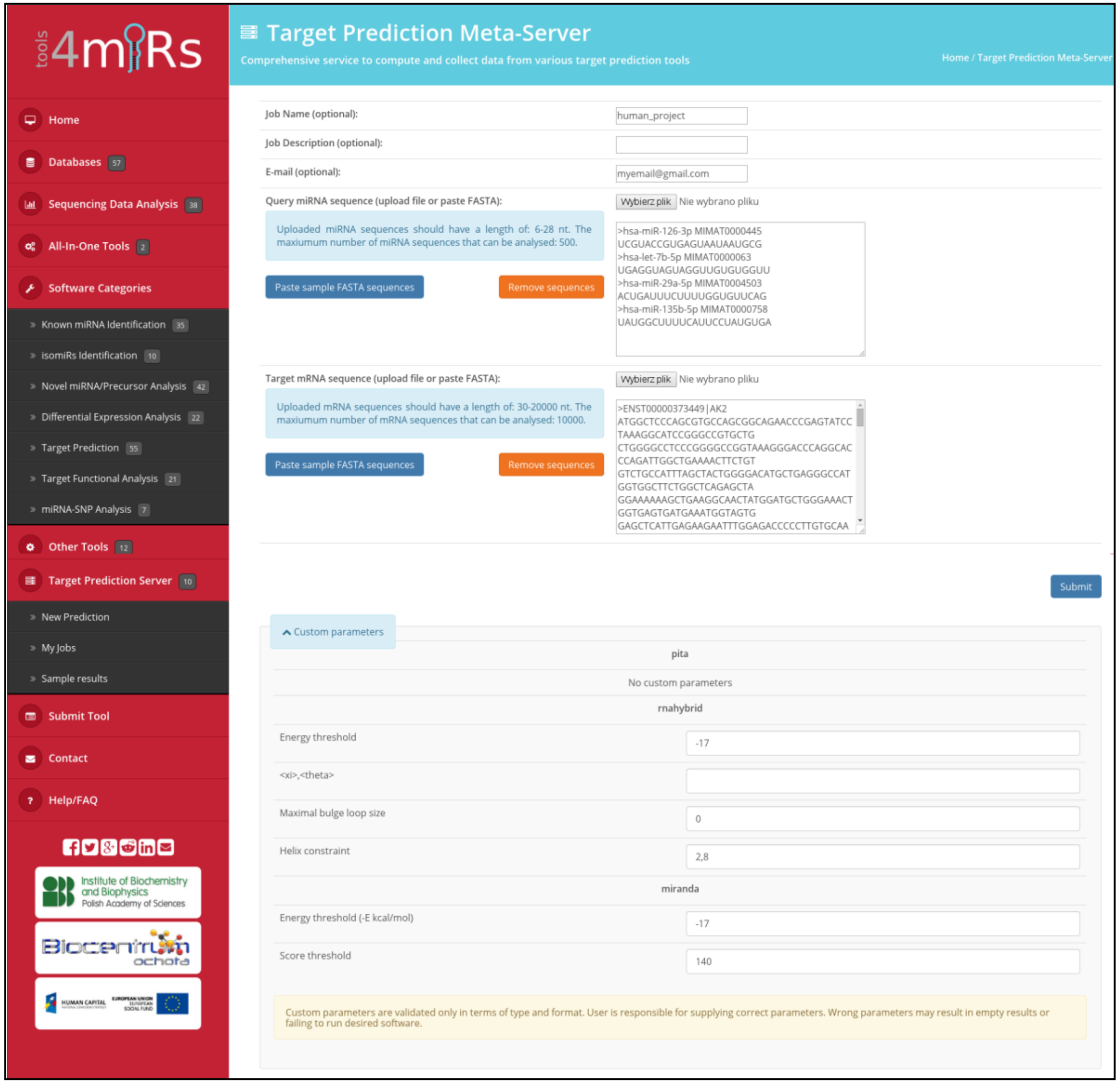
Target Prediction Meta-Server

Comprehensive service to compute and collect data from various target prediction tools

Number of tools found: 5

	#	Name	Description	Online/Local	Specific	Features	Target Region	References
☐	1	MicroTar	MicroTar is a microRNA target prediction program based on mRNA sequence complementarity (3'-UTR seed matches) and RNA duplex energy prediction, assessing the impact of microRNA binding on complete mRNA molecules. It uses the RNAlib library from the Vienna RNA package	✗ / ✓	• animals	• seed match • free energy • site accessibility • G:U pairs allowed in the seed	3' UTR	Homepage Publication Dec. 18, 2006 Citations: 22
☑		miRanda	miRanda algorithm is based on a comparisone of miRNAs complementarity to 3'UTR regions. The binding energy of the duplex structure, evolutionary conservation of the whole target site and its position within 3"UTR are calculated and account for a final result which is a weighted soum of match and mismatch scores for base pairs and gap penalties.	✗ / ✓	• any	• seed match • conservation • free energy • G:U pairs allowed in the seed	3' UTR, CDS	Homepage Publication Oct. 5, 2004 Citations: 900
☑		PITA	PITA offers a brand new view on the miRNA target prediction. The main assumption is based on the fact the mRNA structure plays a role in target recognition by thermodynamically promoting or disfavoring the interaction.	✓ / ✓	• human (online) • fly (online) • mouse (online) • worm (online) • animals (local)	• seed match • free energy • site accessibility • target-site abundance • G:U pairs allowed in the seed	3' UTR	Homepage Publication Sept. 23, 2007 Citations: 593
☑		RNAhybrid	RNAhybrid is a tool for finding the minimum free energy hybridization of a long and a short RNA. The hybridization is performed in a kind of domain mode, ie. the short sequence is hybridized to the best fitting part of the long one. The tool is primarily meant as a means for microRNA target prediction.	✓ / ✓	• any	• seed match • free energy • target-site abundance • 3' compensatory pairing • G:U pairs allowed in the seed	3' UTR	Homepage Publication July 1, 2006 Citations: 221
☐	5	TargetSpy	TargetSpy, a novel computational approach for predicting target sites regardless of the presence of a seed match. It is based on machine learning and automatic feature selection using a wide spectrum of compositional, structural, and base pairing features covering current biological knowledge. It does not rely on evolutionary conservation, which allows the detection of species-specific interactions and makes TargetSpy suitable for analyzing unconserved genomic sequences.	✓ / ✓	• human • mouse • chicken • rat • fly	• free energy • site accessibility • machine learning • 3' compensatory pairing	3' UTR	Homepage Publication May 28, 2010 Citations: 31

C



Target Prediction Meta-Server

Comprehensive service to compute and collect data from various target prediction tools

Job Name (optional): human_project

Job Description (optional):

E-mail (optional): myemail@gmail.com

Query miRNA sequence (upload file or paste FASTA): Nie wybrano pliku

Uploaded miRNA sequences should have a length of: 6-28 nt. The maxiumum number of miRNA sequences that can be analysed: 500.

```
>hsa-miR-126-3p MIMAT0000445
UCGUACCGUGAGUAAUAAUGCG
>hsa-let-7b-5p MIMAT0000063
UGAGGUAGUAGGUUGUGUGGUU
>hsa-miR-29a-5p MIMAT0004503
ACUGAUUUCUUUUGGUGUUCAG
>hsa-miR-135b-5p MIMAT0000758
UAUGGCUUUUCAUUCCUAUGUGA
```

Target mRNA sequence (upload file or paste FASTA): Nie wybrano pliku

Uploaded mRNA sequences should have a length of: 30-20000 nt. The maximum number of mRNA sequences that can be analysed: 10000.

```
>ENST00000373449|AK2
ATGGCTCCCAGCGTGCCAGCGGCAGAACCCGAGTATCC
TAAAGGCATCCGGGCCGTGCTG
CTGGGGCCTCCCGGGGCCGGTAAAGGGACCCAGGCAC
CCAGATTGGCTGAAAACTTCTGT
GTCTGCCATTTAGCTACTGGGGACATGCTGAGGGCCAT
GGTGGCTTCTGGCTCAGAGCTA
GGAAAAAAGCTGAAGGCAACTATGGATGCTGGGAAACT
GGTGAGTGATGAAATGGTAGTG
GAGCTCATTGAGAAGAATTTGGAGACCCCCTTGTGCAA
```

pita

No custom parameters

rnahybrid

Energy threshold: -17

<xi>,<theta>:

Maximal bulge loop size: 0

Helix constraint: 2,8

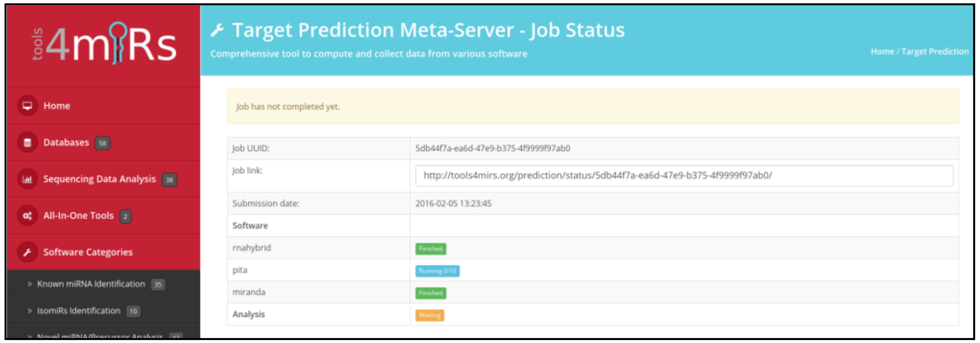
miranda

Energy threshold (-E kcal/mol): -17

Score threshold: 140

Custom parameters are validated only in terms of type and format. User is responsible for supplying correct parameters. Wrong parameters may result in empty results or failing to run desired software.

D



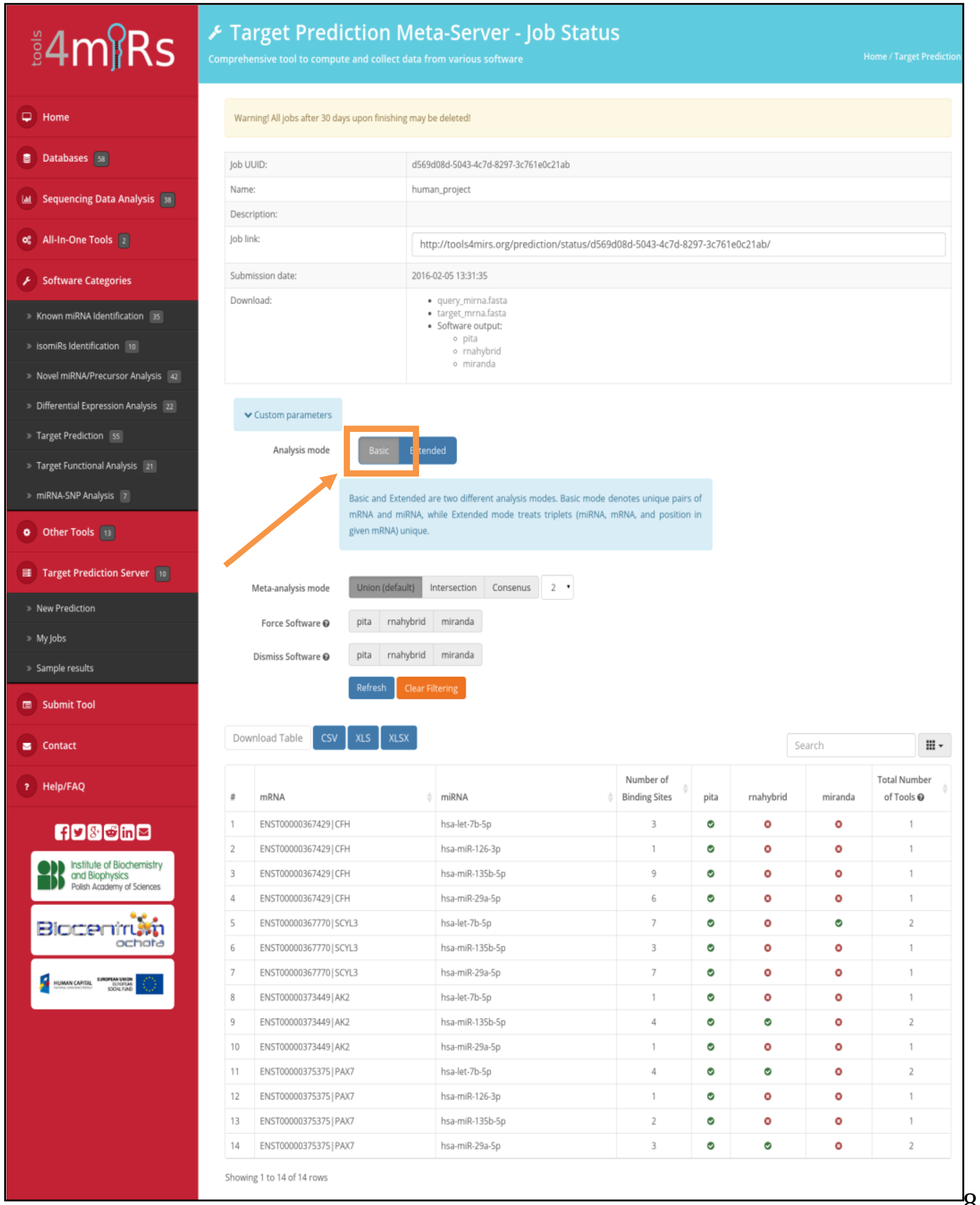
Target Prediction Meta-Server - Job Status

Comprehensive tool to compute and collect data from various software

Job has not completed yet.

Job UUID:	5db44f7a-ea6d-47e9-b375-4f9999f97ab0
Job link:	http://tools4mirs.org/prediction/status/5db44f7a-ea6d-47e9-b375-4f9999f97ab0/
Submission date:	2016-02-05 13:23:45
Software	
rnahybrid	Finished
pita	Running 0:10
miranda	Finished
Analysis	Waiting

E



Target Prediction Meta-Server - Job Status

Comprehensive tool to compute and collect data from various software

Warning! All jobs after 30 days upon finishing may be deleted!

Job UUID:	d569d08d-5043-4c7d-8297-3c761e0c21ab
Name:	human_project
Description:	
Job link:	http://tools4mirs.org/prediction/status/d569d08d-5043-4c7d-8297-3c761e0c21ab/
Submission date:	2016-02-05 13:31:35
Download:	• query_mirna.fasta • target_mrna.fasta • Software output: ○ pita ○ rnahybrid ○ miranda

Custom parameters

Analysis mode: Basic | Extended

Basic and Extended are two different analysis modes. Basic mode denotes unique pairs of mRNA and miRNA, while Extended mode treats triplets (miRNA, mRNA, and position in given mRNA) unique.

Meta-analysis mode: Union (default) | Intersection | Consensus | 2

Force Software: pita | rnahybrid | miranda

Dismiss Software: pita | rnahybrid | miranda

Refresh | Clear Filtering

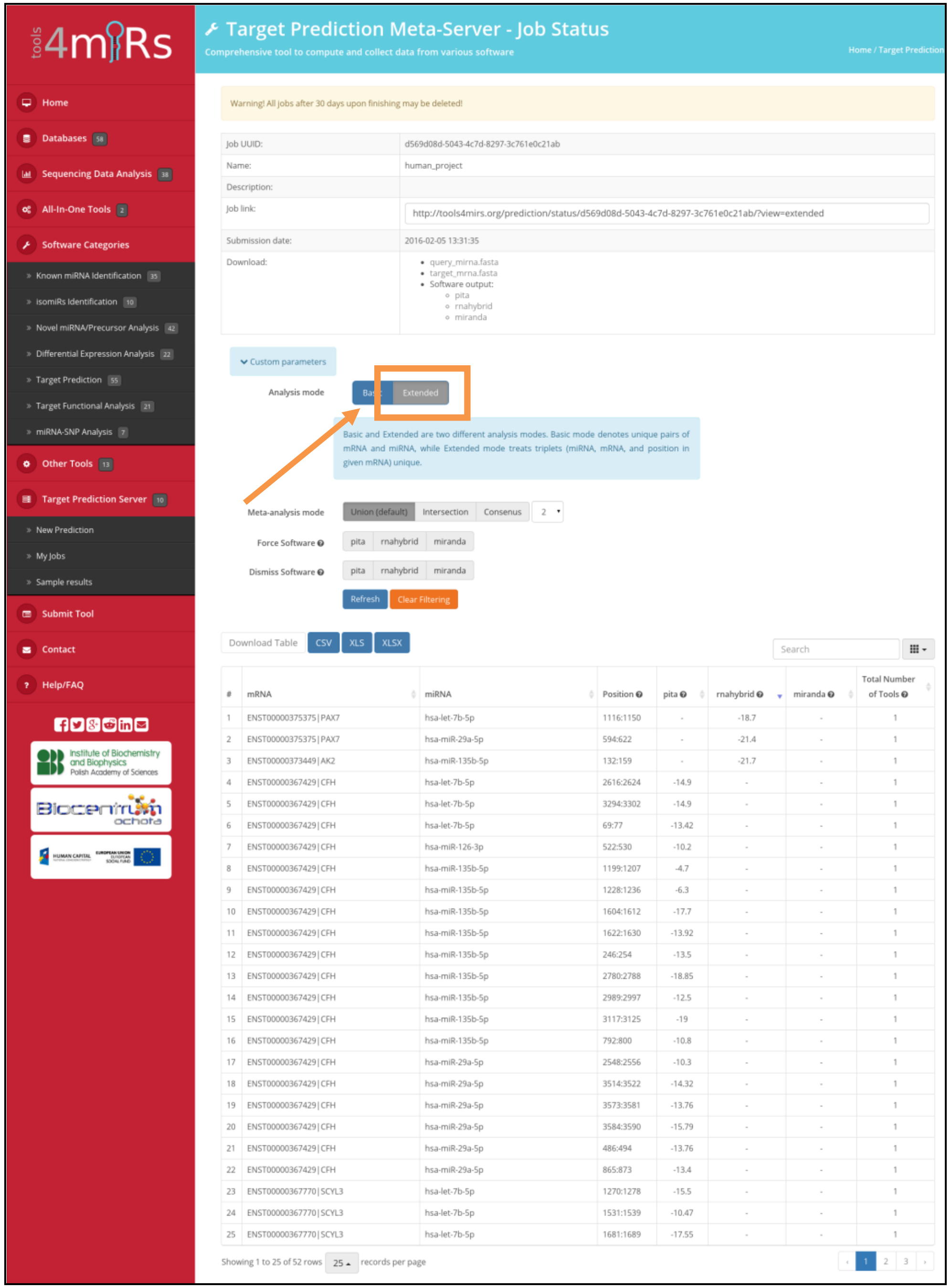
Download Table | CSV | XLS | XLSX

Search

#	mRNA	miRNA	Number of Binding Sites	pita	rnahybrid	miranda	Total Number of Tools
1	ENST00000367429|CFH	hsa-let-7b-5p	3	✓	✗	✗	1
2	ENST00000367429|CFH	hsa-miR-126-3p	1	✓	✗	✗	1
3	ENST00000367429|CFH	hsa-miR-135b-5p	9	✓	✗	✗	1
4	ENST00000367429|CFH	hsa-miR-29a-5p	6	✓	✗	✗	1
5	ENST00000367770|SCYL3	hsa-let-7b-5p	7	✓	✗	✓	2
6	ENST00000367770|SCYL3	hsa-miR-135b-5p	3	✓	✗	✗	1
7	ENST00000367770|SCYL3	hsa-miR-29a-5p	7	✓	✗	✗	1
8	ENST00000373449|AK2	hsa-let-7b-5p	1	✓	✗	✗	1
9	ENST00000373449|AK2	hsa-miR-135b-5p	4	✓	✓	✗	2
10	ENST00000373449|AK2	hsa-miR-29a-5p	1	✓	✗	✗	1
11	ENST00000375375|PAX7	hsa-let-7b-5p	4	✓	✓	✗	2
12	ENST00000375375|PAX7	hsa-miR-126-3p	1	✓	✗	✗	1
13	ENST00000375375|PAX7	hsa-miR-135b-5p	2	✓	✗	✗	1
14	ENST00000375375|PAX7	hsa-miR-29a-5p	3	✓	✓	✗	2

Showing 1 to 14 of 14 rows

F



Target Prediction Meta-Server - Job Status

Comprehensive tool to compute and collect data from various software

Warning! All jobs after 30 days upon finishing may be deleted!

Job UUID:	d569d08d-5043-4c7d-8297-3c761e0c21ab
Name:	human_project
Description:	
Job link:	http://tools4mirs.org/prediction/status/d569d08d-5043-4c7d-8297-3c761e0c21ab/?view=extended
Submission date:	2016-02-05 13:31:35
Download:	• query_mirna.fasta • target_mrna.fasta • Software output: ○ pita ○ rnahybrid ○ miranda

Custom parameters

Analysis mode: Basic | Extended

Basic and Extended are two different analysis modes. Basic mode denotes unique pairs of mRNA and miRNA, while Extended mode treats triplets (miRNA, mRNA, and position in given mRNA) unique.

Meta-analysis mode: Union (default) | Intersection | Consensus | 2

Force Software: pita | rnahybrid | miranda

Dismiss Software: pita | rnahybrid | miranda

Refresh | Clear Filtering

Download Table | CSV | XLS | XLSX

Search

#	mRNA	miRNA	Position	pita	rnahybrid	miranda	Total Number of Tools
1	ENST00000375375|PAX7	hsa-let-7b-5p	1116:1150	-	-18.7	-	1
2	ENST00000375375|PAX7	hsa-miR-29a-5p	594:622	-	-21.4	-	1
3	ENST00000373449|AK2	hsa-miR-135b-5p	132:159	-	-21.7	-	1
4	ENST00000367429|CFH	hsa-let-7b-5p	2616:2624	-14.9	-	-	1
5	ENST00000367429|CFH	hsa-let-7b-5p	3294:3302	-14.9	-	-	1
6	ENST00000367429|CFH	hsa-let-7b-5p	69:77	-13.42	-	-	1
7	ENST00000367429|CFH	hsa-miR-126-3p	522:530	-10.2	-	-	1
8	ENST00000367429|CFH	hsa-miR-135b-5p	1199:1207	-4.7	-	-	1
9	ENST00000367429|CFH	hsa-miR-135b-5p	1228:1236	-6.3	-	-	1
10	ENST00000367429|CFH	hsa-miR-135b-5p	1604:1612	-17.7	-	-	1
11	ENST00000367429|CFH	hsa-miR-135b-5p	1622:1630	-13.92	-	-	1
12	ENST00000367429|CFH	hsa-miR-135b-5p	246:254	-13.5	-	-	1
13	ENST00000367429|CFH	hsa-miR-135b-5p	2780:2788	-18.85	-	-	1
14	ENST00000367429|CFH	hsa-miR-135b-5p	2989:2997	-12.5	-	-	1
15	ENST00000367429|CFH	hsa-miR-135b-5p	3117:3125	-19	-	-	1
16	ENST00000367429|CFH	hsa-miR-135b-5p	792:800	-10.8	-	-	1
17	ENST00000367429|CFH	hsa-miR-29a-5p	2548:2556	-10.3	-	-	1
18	ENST00000367429|CFH	hsa-miR-29a-5p	3514:3522	-14.32	-	-	1
19	ENST00000367429|CFH	hsa-miR-29a-5p	3573:3581	-13.76	-	-	1
20	ENST00000367429|CFH	hsa-miR-29a-5p	3584:3590	-15.79	-	-	1
21	ENST00000367429|CFH	hsa-miR-29a-5p	486:494	-13.76	-	-	1
22	ENST00000367429|CFH	hsa-miR-29a-5p	865:873	-13.4	-	-	1
23	ENST00000367770|SCYL3	hsa-let-7b-5p	1270:1278	-15.5	-	-	1
24	ENST00000367770|SCYL3	hsa-let-7b-5p	1531:1539	-10.47	-	-	1
25	ENST00000367770|SCYL3	hsa-let-7b-5p	1681:1689	-17.55	-	-	1

Showing 1 to 25 of 52 rows 25 records per page

‹ 1 2 3 ›