

6-2015

# Dietary MicroRNA Database (DMD): An Archive Database and Analytic Tool for Food-Borne microRNAs

Kevin Chiang

University of Nebraska-Lincoln, [t3-kchiang@unl.edu](mailto:t3-kchiang@unl.edu)

Jiang Shu

University of Nebraska-Lincoln, [jshu2@unl.edu](mailto:jshu2@unl.edu)

Janos Zempleni

University of Nebraska-Lincoln, [jzempleni2@unl.edu](mailto:jzempleni2@unl.edu)

Juan Cui

University of Nebraska-Lincoln, [jcui@unl.edu](mailto:jcui@unl.edu)

Follow this and additional works at: <http://digitalcommons.unl.edu/csearticles>



Part of the [Databases and Information Systems Commons](#), [Environmental Health and Protection Commons](#), and the [Food Microbiology Commons](#)

---

Chiang, Kevin; Shu, Jiang; Zempleni, Janos; and Cui, Juan, "Dietary MicroRNA Database (DMD): An Archive Database and Analytic Tool for Food-Borne microRNAs" (2015). *CSE Journal Articles*. 111.

<http://digitalcommons.unl.edu/csearticles/111>

This Article is brought to you for free and open access by the Computer Science and Engineering, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in CSE Journal Articles by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

RESEARCH ARTICLE

# Dietary MicroRNA Database (DMD): An Archive Database and Analytic Tool for Food-Borne microRNAs

Kevin Chiang<sup>1</sup>, Jiang Shu<sup>1</sup>, Janos Zemleni<sup>2</sup>, Juan Cui<sup>1\*</sup>

**1** Department of Computer Science and Engineering, University of Nebraska-Lincoln, Lincoln, NE, United States of America, **2** Department of Nutrition and Health Sciences, University of Nebraska-Lincoln, Lincoln, NE, United States of America

\* [jcui@unl.edu](mailto:jcui@unl.edu)



## OPEN ACCESS

**Citation:** Chiang K, Shu J, Zemleni J, Cui J (2015) Dietary MicroRNA Database (DMD): An Archive Database and Analytic Tool for Food-Borne microRNAs. PLoS ONE 10(6): e0128089. doi:10.1371/journal.pone.0128089

**Academic Editor:** Fengfeng Zhou, Shenzhen Institutes of Advanced Technology, CHINA

**Received:** March 6, 2015

**Accepted:** April 23, 2015

**Published:** June 1, 2015

**Copyright:** © 2015 Chiang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data within the paper are accessible at <http://sbbi.unl.edu/dmd/>.

**Funding:** DMD received funding through the NIH-supported Nebraska Center for the Prevention of Obesity Diseases through Dietary Molecules (P20GM104320). Funding was also received from the National Institute of Food and Agriculture (2015-67017-23181), the Gerber Foundation, the Egg Nutrition Center, the University of Nebraska Agricultural Research Division (Hatch Act), and the United States Department of Agriculture Multi-State Group W3002. These funders had no role in study

## Abstract

With the advent of high throughput technology, a huge amount of microRNA information has been added to the growing body of knowledge for non-coding RNAs. Here we present the Dietary MicroRNA Databases (DMD), the first repository for archiving and analyzing the published and novel microRNAs discovered in dietary resources. Currently there are fifteen types of dietary species, such as apple, grape, cow milk, and cow fat, included in the database originating from 9 plant and 5 animal species. Annotation for each entry, a mature microRNA indexed as DM0000\*, covers information of the mature sequences, genome locations, hairpin structures of parental pre-microRNAs, cross-species sequence comparison, disease relevance, and the experimentally validated gene targets. Furthermore, a few functional analyses including target prediction, pathway enrichment and gene network construction have been integrated into the system, which enable users to generate functional insights through viewing the functional pathways and building protein-protein interaction networks associated with each microRNA. Another unique feature of DMD is that it provides a feature generator where a total of 411 descriptive attributes can be calculated for any given microRNAs based on their sequences and structures. DMD would be particularly useful for research groups studying microRNA regulation from a nutrition point of view. The database can be accessed at <http://sbbi.unl.edu/dmd/>.

## Introduction

Empowered by revolutionary sequencing technology, microRNAs have been extensively discovered in various dietary resources including plants (e.g. rice and tomato) and animals (e.g. milk and meats). Given the broad implications of microRNA in health and disease [1–8], research enthusiasm for functional impacts of exogenous food microRNA in human cellular phenotypes has soared, which warrants the efforts to build related bioinformatics tools and databases. The

design, data collection and analysis, decision to publish, or preparation of the manuscript.

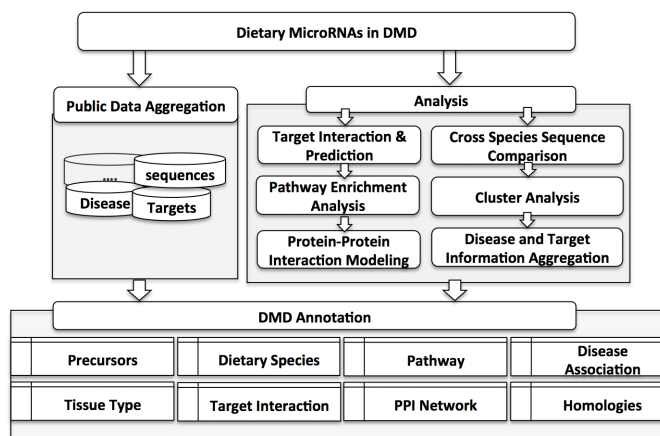
**Competing Interests:** The authors have declared that no competing interests exist.

Dietary MicroRNA Database (DMD) represents the first repository in this domain for archiving and distributing the published food-borne microRNAs in literatures and public databases.

There are several public databases focused on microRNA identification and targets prediction that archive validated microRNAs with sequence, structure and interaction information. For example, miRBase (<http://www.mirbase.org>) records 64,473 microRNAs from 223 species [9] and MiRecords [10] hosts 2,705 records of interactions between 644 microRNAs and 1,901 target genes in 9 animal species. Databases such as TargetScan [11], Miranda [12] and MirTarBase [13] provide information of the validated gene targets as well as the computationally predicted targets. For example, 60% of human genes are regulated by microRNAs, participating in many major cellular processes such as cell growth, differentiation and apoptosis [14, 15]. In addition, microRNA expression data, although limited, are archived in public databases such as GEO databases [16] and TCGA [17]. However, none of the aforementioned databases cover dietary information that may represent new horizon in microRNA research. For example, miRBase has reported 808 microRNAs in bovine, whereas only 243 of them have been found in cow milk [18] and 213 in the fat of cow beef [19]. Likewise, human breast milk only contains 434 microRNAs, out of the total of 2,588 microRNAs in human [20]. We envision such diet-specific cohorts would be important for nutritionists and general biologists to investigate microRNA dietary intake and analyze subsequent regulations in human health and diseases. Expelling evidences sustaining our hypothesis include the following: it has been recently discovered that human can absorb certain exosomal microRNAs from cow's milk, e.g., miR-29b and 200c, and that endogenous microRNA synthesis does not compensate for dietary deficiency [21]; the biogenesis and function of such exogenous miRNAs are evidently health related [21–24]. However, while the evidence in support of bioavailability of milk miRNAs is unambiguous, a recent report that mammals can also absorb plant miRNAs (e.g. miR-168a) from rice [25] was met with widespread skepticism [26–29]. Based on these evidences, challenging questions may be raised regarding how humans pick up microRNAs from diet and what are the broader roles played by such exogenous microRNAs in human disease processes.

In order to facilitate more advanced research related to dietary microRNAs, DMD was developed as the first repository for archiving and analyzing the published microRNAs discovered in dietary plants and animals, such as cow milk, breast milk, grape, beef, pork, apple, banana and etc. For each reported microRNA, various types of information have been covered, including sequences, genome locations, hairpin structures of parental pre-microRNAs, disease relevance, and experimentally validated gene targets. We also integrate an analytical pipeline into this platform that includes cross-species sequence comparison, target prediction, gene enrichment analysis and microRNA-mediated gene network construction, which we will introduce in the following sections.

Compared to other microRNA-related databases, DMD also has a few unique features. For example, a feature generation tool allows users to calculate a comprehensive set of molecular discriminators based on the sequences and structures of any microRNA entry in the database or uploaded on their own. These discriminators have been considered as important features for microRNA identification and microRNA-mRNA interaction prediction and have been employed by many current tools in addition to the use of complementary seed sequences as major motifs in animal and plant species [11, 30–34]. Based on the targets, one can extract the functional pathways information and infer the functional impacts of the microRNAs through their gene regulation [35, 36]. In the later section, we will use a case study to demonstrate the usefulness of this database.



**Fig 1. DMD construction workflow and the outline of data content.**

doi:10.1371/journal.pone.0128089.g001

## Materials and Methods

### Database Construction and Access

[Fig 1](#) shows the workflow of the data collection and analysis with DMD. Through literature and database search, we compiled and reported microRNAs from 15 types of dietary species. For each entry, the basic annotation page includes ten types of information including mature sequences, genome coordinates, pre-microRNA sequence, hairpin structure, cross-species sequence comparison, disease relevance, and the experimentally validated and predicted gene targets. For entries from public databases, e.g. miRBase, we have provided links to the external annotation pages.

The DMD was created using a MySQL database, consisting of 25 tables ([S1 Fig](#)). The outline ([Fig 1](#)) shows that the database content can be categorized into three areas, namely basics, annotation and analysis. First, many external databases are integrated within DMD to allow for quick viewing and annotation of microRNAs. Second, there are the prediction tools, which allow users to quickly view homologous microRNAs via the clustering analysis by CD-HIT [37] and predict gene targets in their own species and in human. Finally, there is an intensive process to annotate microRNAs into dietary species and tissues.

The [S1 Fig](#) shows the database design, table relationships and indexing patterns among these tables. All information, including from external databases, to prediction tools and annotation, are heavily connected, as can be seen from the schema. This allows for the information within the database to be shared easily and quickly.

In addition to the MySQL databases, the graph database Neo4j (<http://neo4j.org>) was used to model protein-protein interaction. The graph database consists of two different types of nodes, a microRNA and a protein node, and two types of edges, a protein-protein interaction undirected edge, and a microRNA to protein regulation directed edge.

The information of DMD can be freely accessed from <http://sbbi.unl.edu/dmd/>. Data submission and download can be accessed through a secure user login system.

### Cross-species Sequence Comparison

In order to assess the sequence conservation of each microRNA during evolution, we conducted sequence alignment and comparison using CD-HIT [37] where the microRNA with

similar sequences can be grouped into the same clusters. In this analysis, each cluster represents a collection of microRNAs that share identical or highly similar sequences (with identity higher than 95% of the sequence length), which could originate from various species, e.g. homologous microRNAs. Within the cluster, the user will be able to view the microRNA name, sequence alignment, associated gene targets and diseases, along with the option of viewing information among diet-only or all species.

## Querying the Database (Browsing and Searching)

Users are able to browse microRNAs by species using the browse page. Accessing each species will output a whole list of microRNAs specifically discovered in that dietary species. For example, microRNAs under “cow milk” and “cow fat” are subsets of all known microRNAs from bovine organism. In addition to being able to browse by species, there are three methods of searching the datasets, by “ID” (DMD index number, e.g. DM00001), “Name” (microRNA name, e.g. bta-miR-29b or part of the name, e.g. 29b) or “Sequence” (either mature or pre-microRNA sequences, e.g. ugagguaguagguuauaguu). Again, the search can be constrained within mature microRNAs or precursors according to the user defined criteria. All outputs from the search are organized initially by their unique DMD identifiers, but the results may be re-sorted by microRNA name, or sequence.

## Experimental and Predicted Targets

Experimental targets information was extracted from miRTarBase [38], which contains 18 species and 51,460 miRNA-target interactions. A few of computational tools were included for target prediction, especially for species without validated target information, including MirTarget [39], targetScan [11] and psRNAtarget [40]. Please note prediction could be specifically designed for certain organism, e.g. human specific or plant specific. All microRNA sequences will be also subject to a target prediction against human genome no matter whether predictions on their own genomes are available or not.

## Functional Analysis Based on MicroRNA Targets

According to the predicted and experimental targets for each microRNA entry, users can choose to run a pathway enrichment analysis on selected targets. 1,955 pathways from KEGG [41] are included. Modified p-value was calculated for each relevant pathway based on Fisher’s exact test on queried targets against the whole genome. In addition to the pathway enrichment analysis, protein-protein interaction (PPI) network [42] was employed to visualize the microRNA-mRNA regulation network.

## Feature Construction

As molecular properties of microRNA sequences and structures are key for target identification, we have developed a *feature* page to allow users to calculate for any given microRNA sequences a list of features categorized into two classes: sequence-based features and secondary structure features. Particularly, for each mature miRNA, features were generated on both mature sequences and the corresponding pre-miRNA sequences, such as existence of palindromic sequences, sequence length and the composition of monomers and dimers. Such features have been shown to be discriminants when used for machine learning [18, 31, 32, 43–46].

Secondary structure features were calculated based on the pre-miRNA sequences. For example, RNAfold [47] was employed to predict secondary structure and calculate Minimum Free Energy (MFE) [48]. Based on the predicted structure of pre-miRNA, 32 triplet features and 11

base-paired features were calculated, such as A (((the frequency of 3 paired nucleotides leading with A) and %pairGC (percentage of the paired G-C bases). Additionally, RNAshape was used to map secondary structures to tree-like domain of shapes, retaining adjacency and nesting of structural features, but disregarding helix length [49]. STOAT, packaged in the NOBAI web server, was utilized to compute Shannon Entropy (Q) and Frobenius Norm (F) [50]. See Table 1 for a complete list of features.

## Results and Discussion

As a dietary microRNA database, DMD acts as the first repository archiving microRNA sequence and annotation that are related to any dietary species. Currently there are 15 dietary species have been curated in DMD, including five animal species (human, chicken, cow, pig and salmon) and nine plant species (soybean, tomato, corn, apple, orange, banana, grape, rice, and wheat). Please note that dietary species might originate from the same biological organism.

**Table 1. List of features available for generation.**

Category	Feature Details	Feature Dimensions	Reference
Primary Sequence	Single Nucleotide Frequency	$4 \times 3^1$	[31, 46]
	Pairwise Nucleotide Frequency	$16 \times 3^1$	[31, 32, 43, 46]
	Triplet Nucleotide Frequency	$64 \times 3^1$	[31, 43, 46]
	Quadruplet Nucleotide Frequency	$256 \times 3^1$	[46]
	A + U Frequency	$1 \times 3^1$	[43, 44]
	G + C Frequency	$1 \times 3^1$	[31, 32, 43, 44, 46]
	G + U Frequency	$1 \times 3^1$	[43, 44, 46]
	Number of Palindromes in Sequence	$1 \times 3^1$	[45]
	Length	$1 \times 3^1$	[31]
	Pairs of A-U in Premature microRNA	1	[43]
	Pairs of G-C in Premature microRNA	1	[43, 44]
	Pairs of G-U in Premature microRNA	1	[43]
Secondary Structure	Nucleotide to RNAfold <sup>2</sup> triplet match. (A(((, C(., G(. . . etc. . .)	32	[44, 46]
	Minimum Free Energy, Normalized Minimum Free Energy, Frequency of Minimum Free Energy Structures	3	[31, 32, 43, 44]
	Ensemble Free Energy, Normalized Ensemble Free Energy	2	[32, 44]
	Stem Statistics (Stems, Average Stem Length, Maximum Stem Length, Stem containing AU, Stem containing GC, Stem containing GU)	6	[32, 43, 44]
	Minimum Free Energy Statistics (mfe/G+C frequency, mfe/stems, mfe/unpaired nucleotides, mfe/paired nucleotides, difference in mfe and efe, and ensemble diversity).	6	[32, 43, 44]
	Percentage of sequence composing of pairs.	1	[46]
	Frequency of Nucleotides that occur outside of UA, GU, GC pairs.	4	[46]
	Predicted shape type probability base on RNAshapes <sup>3</sup> .	5	[51]
	STOAT <sup>4</sup> statistics (Shannon Entropy, Frobenius Norm, Base-pairing propensity, and mean stem length)	4	[32]

<sup>1</sup>These features may be calculated for the premature sequence, mature sequence, and seed region sequence.

<sup>2</sup>RNAfold is an external tool that is run with the—p option to generate the partition function and base pairing probability.

<sup>3</sup>RNAshapes is an external tool that is run with the—t option to specify 5 different shape types.

<sup>4</sup>STOAT is an external tool that is run with the—x 31 option to signify 31 character states and the—v option to display a verbose option that is easier to parse.

Table 2. Statistics of microRNAs and species in DMD.

Types	Species	Mature [9] #. of dietary miRNAs/#. of known miRNAs in the organism	Precursor [9]	Exp. Target [38]	Pred. Target [38, 40]
<b>Animal</b>	Human Breastmilk	434 / 2588	402 / 1881	9995	17,416
	Cow Milk	243 / 793	245 / 808	3	16,451
	Cow Fat	205 / 793	229 / 808	5	16,269
	Atlantic Salmon	498 / 498	371 / 371	-	17,319
	Chicken	994 / 994	740 / 740	19	18,075
	Pig	411 / 411	382 / 382	-	17,406
<b>Plant</b>	Apple	203 / 207	202 / 206	-	13,098
	Banana	360 / 360	180 / 180	-	16,537
	Corn	309 / 321	166 / 172	-	16,456
	Grape	108 / 186	157 / 163	-	12,236
	Orange	047 / 064	057 / 060	-	12,655
	Rice	634 / 713	526 / 592	-	21,355
	Soybean	620 / 639	554 / 573	-	20,261
	Tomato	040 / 110	068 / 77	-	10,827
	Wheat	111 / 119	108 / 116	-	15,327

doi:10.1371/journal.pone.0128089.t002

For example, the bovine microRNAs are organized under different dietary groups, including cow milk and cow fat. Table 2 shows the statistics of different types of information archived for each species.

## Database Content

The information stored in DMD is categorized into *Sequences* and *Annotation*.

**Sequences:** Currently, there are 11,569 microRNA entries in DMD, including 5,217 unique mature sequences and 5,865 unique pre-microRNA sequences. Duplicates contribute to the total microRNA, which is due to a microRNA being present in multiple dietary species. DMD follows the same naming standard for each entry, e.g. microRNA gene names have the form hsa-miR-200, consistent with other databases, e.g. miRBase. The prefix signifies the organism, in this case Homo Sapiens. Each entry in the database, indexed as DM0000\*, represents the mature sequence, with the information on the genomic location and hairpin sequence of the parental pre-microRNA, indexed as DP0000\*, which will have the corresponding miRBase index if entries from both database are the same. Homologous microRNA loci in different species are assigned the same number. Paralogous microRNAs are assigned names with lettered and numbered suffixes, depending on whether the derived mature microRNA is identical in sequence, or contains sequence differences. The derived mature microRNAs were previously assigned names of the form dme-miR-100 and dme-miR-100\*, for the guide and passenger strand, respectively while hsa-miR-100-5p and hsa-miR-100-3p were assigned for sequences derived from the 5' and 3' arms of the hsa-miR-100 hairpin precursor.

**Annotation:** In addition to the general information such as sequence and structure, DMD has also generated features for each of the microRNAs. The *feature* page provides users with 411 molecular attributes that can be calculated based on the microRNA sequences and



structures. [Table 1](#) lists all the features included in this study, and also presents a non-exhaustive list of studies that have reported use of a particular feature. Other annotation information covers targets, sequence comparison against other species, and pathway and interaction network analysis based on the either experimentally validated targets or computational predicted targets. Specifically, the pathway information was compiled from KEGG. PPI information has also been used to visualize the interaction network based on the given targets. Furthermore, microRNA disease information was extracted from mirCancer and obesity, the Human microRNA Disease Database [52], PhenomiR [53], and PubMed literature search.

The *browse* page allows users to access microRNAs under each dietary species while the database metadata can be downloaded as plain text. For each microRNA sequence entry, there are links to other databases providing the primary references that describe its discovery, links to the evidence supporting the microRNA annotation, genomic coordinates and links to databases of predicted and validated microRNA target sites. Entries can be searched either by sequence or keyword. Note the name cannot be used as a substitute for rigorous sequence analysis.

## Unique Analytical Workflow

DMD offers not only the basic access to sequence and annotation data stored in the database but also a few analytical tools integrated as the web services. For examples, the feature page allows users to calculate for any given microRNA a list of 411 features based on the sequences and structures; the cross-species sequence comparison was integrated on each entry page. Other analyses are focused on the functional inference through microRNA target identification.

**Target prediction.** On each annotation page for a given microRNA, we have provided targets from both experimental studies and computational prediction. A total of 91,032 experimental targets have been included in this database while computational predictions are mainly through MirTarget and psRNAtarget. Since the target prediction could be organism specific, we separated the cases for plant and animal and provide for each microRNA the potential targets in human.

**Functional inference and Network Analysis.** Pathway enrichment analysis and gene network construction have been implemented for users who are interested in functional analysis. Protein-protein interaction visualization network can be seen from a microRNA entry's predicted and experimental targets. First, selecting from the list of targets will display a new list, which shows the interactions between two targets with up to three intermediates. Opting to show the interaction will open a new window, which visualizes the interaction between the two targets. Included in the network are the two selected targets, all its shared gene intermediates, and all the microRNAs that regulate the two targets. We will illustrate in the following section of these analyses.

In addition to the analysis, the password-protected *upload* page allows registered users to deposit newly identified evidences of microRNA occurrences in food. Novel microRNAs can be submitted after an article describing their identification is accepted for publication in a peer-reviewed journal. Since this field needs fast gain of knowledge, we also allow users to upload their raw data prior to the publication, for instance, a single microRNA detected based on PCR or a raw microRNA array or sequencing data that might under review or in-house use. We will take these data in and analyze internally using the broadly used bioinformatics standards.



The Dietary microRNA Database is a searchable database of published and novel microRNA sequences discovered in human food. Each entry in the database, indexed as DMDM0000\*, represents the mature sequence of a food-born microRNA and may be retrieved along with its information about the genomic location and hairpin structure of the pre-microRNA, indexed as DMDP0000\*. The annotation covers the hairpin structure of the pre-microRNAs, cross-species sequence comparison on each mature microRNA, reported disease relevance, and the experimentally validated gene targets, of which a gene regulation network was built for each entry. Novel microRNA sequences are also subjected to computational target prediction and gene network inference.

Users are able to upload their novel findings of any food microRNAs to the database prior to publication results. Visit the [help](#) page for more information about the data sharing.

DMD is managed by the Systems Biology and Biomedical Informatics Lab at the Department of Computer Science and Engineering, University of Nebraska at Lincoln with funding through NIH supported by the Center for the Prevention of Obesity Diseases through Dietary Molecules.

**Search DMD**

Search by: ☒ miRNA name ☐ Sequence  
Show: ☒ Dietary Species Only ☐ All Species  
Show: ☒ Mature miRNA Only ☐ Pre-miRNA Only

Search Entry: **bta-mir-29b**

ID	Name	Length	Sequence
DM00812	bta-mir-29b	23	UAGCACCAUUUGAAAUCAGUGUU

**DMD Annotation**

Precursors Dietary Species Pathway Disease Association  
Tissue Type Target Interaction PPI Network Homologies

2015 is an exciting year for nutrition meeting topics. The topics under discussion in all of our sci...[more](#)

**Metabolism, Diet, and Disease**  
Internationally recognized speakers held talks on topics including metabolic links to cancer, obe...[more](#)

**NPOD**  
The NPOD was launched in August of 2014 as a Center of Biomedical Research Excellence (COBRE) with s...[more](#)

Contact Us  
Systems Biology and Biomedical Informatics (SBI) Laboratory  
University of Nebraska-Lincoln  
1228/122C Avery Hall, 1144 T St  
Lincoln, NE 68588-0115  
Phone: (402) 472-5023  
For any questions or concerns regarding the use of DMD, please email us at [bioinfo@cse.unl.edu](mailto:bioinfo@cse.unl.edu).

Site Info  
Visits: 483

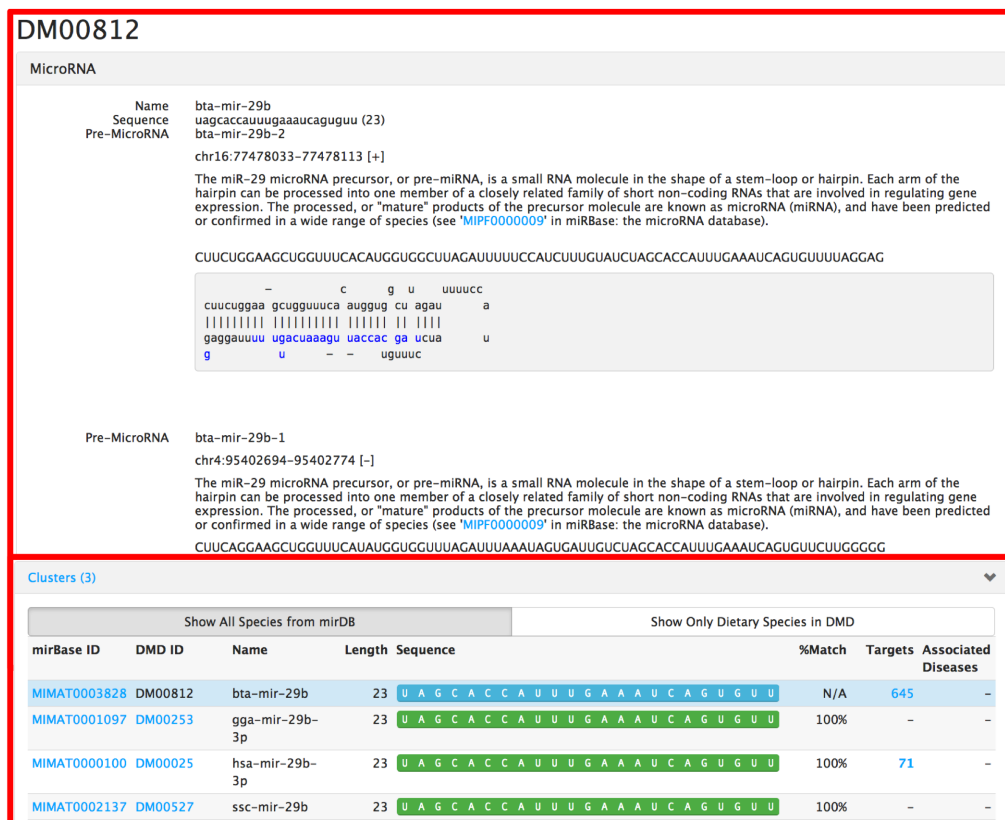
**Fig 2. Illustration of searching "bta-mir-29b" and the search results.**

doi:10.1371/journal.pone.0128089.g002

## An Application Example to Reconstruct microRNA-29b Regulation Network

Bta-miR-29b, a cow-milk microRNA that has been discovered in individuals involved in a milk feeding study [21], is used for illustration of searching and analyzing a specific microRNA in DMD. Fig 2 shows the returned entry when user searching by microRNA name on the search page.

Selecting the ID DM00812 leads to the annotation page shown in Fig 3 and Fig 4. In the top panel of Fig 3, it gives the basic sequence information along with its precursor sequence, structure, and coordinates. The panel labeled "Clusters" shows a list of homologs that have at least 95% sequence similarity in other species. In this case it shows the mir-29b homologs for chicken, human, and pig. This panel also shows the number of targets for each homolog as well as associated diseases, if available. The top panel labeled "Targets" in Fig 4 shows experimentally validated and computationally predicted gene targets. Users have the option to run gene set enrichment analysis on the selected targets as well as visualize the protein-protein network.



**Fig 3. The DMD entry for 'bta-mir-29b,' which contains sequence information, precursor annotation, and homologous sequences and their associated targets and diseases.**

doi:10.1371/journal.pone.0128089.g003

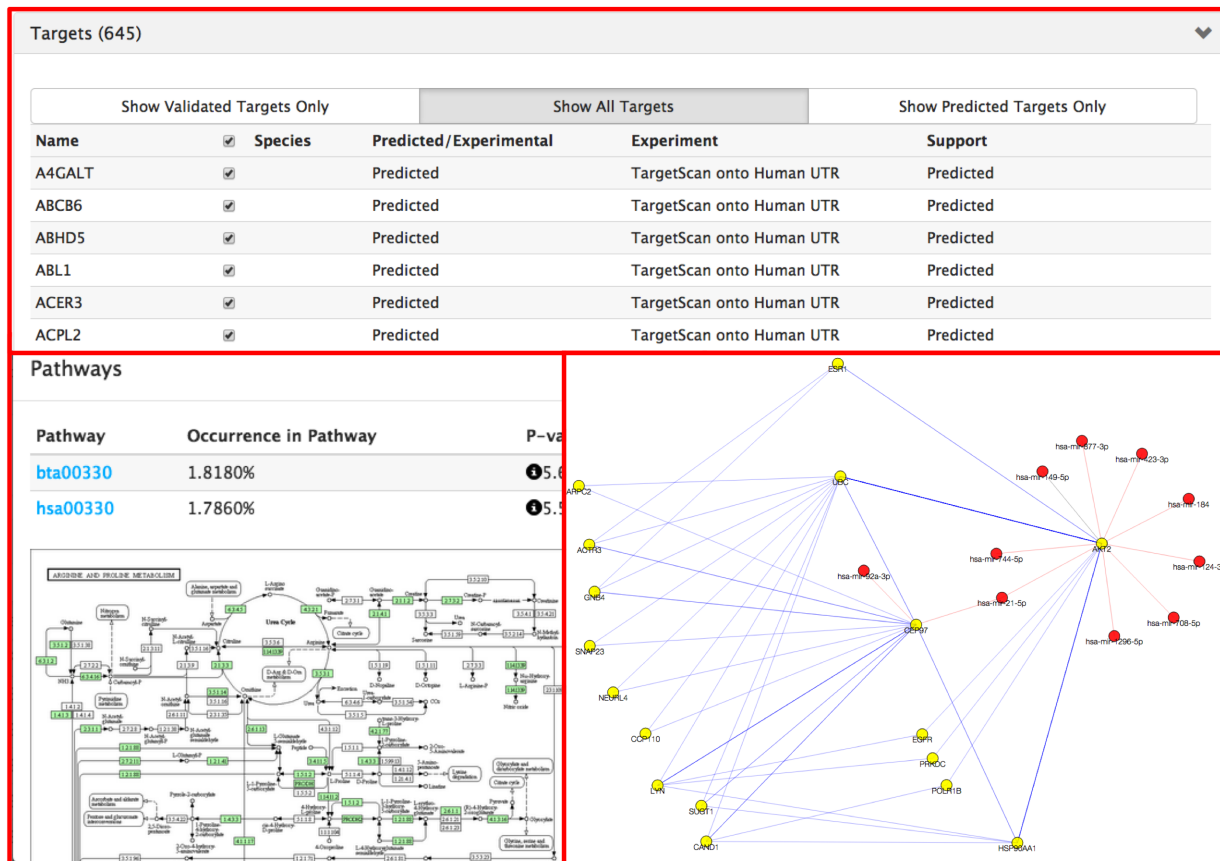
To use bta-mir-29b in a data mining capacity, users may then copy the microRNA sequence and navigate to the feature page and calculate up to 411 features (Fig 5). Users may also submit a fasta file to generate features. Clicking on "Generate Features" will begin the process of computationally generating the features listed in Table 2. Once complete, a page containing tab-separated values is presented to the user to copy and paste. The first line contains a header file describing the features generated.

## Concluding Remarks

Here, we introduce the Dietary MicroRNA Database (DMD) that distributes all published microRNA sequence from dietary resources, for browsing and searching by sequence and keywords, through a web interface (<http://sbbi.unl.edu/dmd/>). In addition, we have provided sequence comparison, target prediction and feature generation functionality on the websites.

## Future Development

We will continue the literature search for newly published data in other dietary species to update the database. We also plan to integrate the real-time analytical pipeline so users can submit their sequences and evaluate the results from their side. Another plan for this database is to let user upload paired microRNA and gene expression profiles where both mRNA expression



**Fig 4. Clockwise from top: "bta-mir-29b" shows 645 targets; the protein-protein network visualization; and the gene enrichment analysis and pathway information.**

doi:10.1371/journal.pone.0128089.g004

## Feature Generation

Available here is tool that is used to generate a list of features for use in machine learning applications. It is split up into two major categories: Primary, and Secondary. The Primary features are all sequence based features while the Secondary features are based on structural attributes. It is recommended to get the structural features for the premature microRNA sequence, the mature sequence, and the seed region, and get the Secondary features for only the premature sequence.

Incl.	Primary Features	#	Description	Reference
✓	Single Nucleotide Frequency	4	The frequency of a single nucleotide (A, U, G, C) with respect to the length of the sequence.	<a href="#">Link</a>
✓	Dinucleotide Frequency	16	The frequency of nucleotide pairs (AA, AU, AG, AC, UA...) with respect to the length of the sequence.	<a href="#">Link</a> <a href="#">Link</a>
✓	RNAshape Shapes	5	Probability of folding into 5 different shapes as provided by RNAshapes. (Max length of sequence = 220)	
✓	Stoat Statistics	4	Shannon Entropy, Frobenius Norm, Mean Stem Length, and Base Pairing Propensity	

File

--OR--

Input Fasta

Name	freq000A	freq000C	freq000G	freq000U	freq00AA	freq00AC	freq00AG	freq00AU
bta-mir-29b	0.30435	0.17391	0.17391	0.34783	0.09091	0.04545	0.09091	0.09091

**Fig 5. The feature generation page showing the entry of "bta-mir-29b" in fasta format and its output in a tab separated values format.**

doi:10.1371/journal.pone.0128089.g005

changes and microRNA expression alterations under the same conditions can be captured, so that one can employ dynamic modeling [54] to infer the conditional-dependent dynamic regulation network.

## Supporting Information

**S1 Fig. DMD Database Schema.**  
(TIFF)

## Acknowledgments

DMD is managed by the Systems Biology and Biomedical Informatics Lab at the Department of Computer Science and Engineering, University of Nebraska at Lincoln.

## Author Contributions

Conceived and designed the experiments: KC JC. Performed the experiments: KC JS. Analyzed the data: KC JS. Contributed reagents/materials/analysis tools: KC JC JS. Wrote the paper: KC JC JS JZ.

## References

1. Calin GA, Ferracin M, Cimmino A, Di Leva G, Shimizu M, Wojcik SE, et al. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med*. 2005; 353(17):1793–801. PMID: [16251535](#)
2. Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell*. 2006; 9(3):189–98. PMID: [16530703](#)
3. Bloomston M, Frankel WL, Petrocca F, Volinia S, Alder H, Hagan JP, et al. MicroRNA expression patterns to differentiate pancreatic adenocarcinoma from normal pancreas and chronic pancreatitis. *JAMA*. 2007; 297(17):1901–8. PMID: [17473300](#)
4. Schetter AJ, Leung SY, Sohn JJ, Zanetti KA, Bowman ED, Yanaihara N, et al. MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. *JAMA*. 2008; 299(4):425–36. doi: [10.1001/jama.299.4.425](#) PMID: [18230780](#)
5. Garzon R, Volinia S, Liu CG, Fernandez-Cymering C, Palumbo T, Pichiorri F, et al. MicroRNA signatures associated with cytogenetics and prognosis in acute myeloid leukemia. *Blood*. 2008; 111(6):3183–9. doi: [10.1182/blood-2007-07-098749](#) PMID: [18187662](#)
6. Wyman SK, Parkin RK, Mitchell PS, Fritz BR, O'Brian K, Godwin AK, et al. Repertoire of microRNAs in epithelial ovarian cancer as determined by next generation sequencing of small RNA cDNA libraries. *PLoS One*. 2009; 4(4):e5311. doi: [10.1371/journal.pone.0005311](#) PMID: [19390579](#)
7. Ambs S, Prueitt RL, Yi M, Hudson RS, Howe TM, Petrocca F, et al. Genomic profiling of microRNA and messenger RNA reveals deregulated microRNA expression in prostate cancer. *Cancer Res*. 2008; 68(15):6162–70. doi: [10.1158/0008-5472.CAN-08-0144](#) PMID: [18676839](#)
8. Croce CM. Causes and consequences of microRNA dysregulation in cancer. *Nat Rev Genet*. 2009; 10(10):704–14. doi: [10.1038/nrg2634](#) PMID: [19763153](#)
9. Griffiths-Jones S. miRBase: microRNA sequences and annotation. *Current protocols in bioinformatics / editorial board, Andreas D Baxevanis [et al]*. 2010;Chapter 12:Unit 12 9 1–0.
10. Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. miRecords: an integrated resource for microRNA-target interactions. *Nucleic acids research*. 2009; 37(Database issue):D105–10. doi: [10.1093/nar/gkn851](#) PMID: [18996891](#)
11. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. 2005; 120(1):15–20. PMID: [15652477](#)
12. Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, et al. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell*. 2006; 126(6):1203–17. PMID: [16990141](#)
13. Hsu SD, Tseng YT, Shrestha S, Lin YL, Khaleel A, Chou CH, et al. miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic acids research*. 2014; 42(Database issue):D78–85. doi: [10.1093/nar/gkt1266](#) PMID: [24304892](#)

14. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004; 116(2):281–97. PMID: [14744438](#)
15. Ambros V. The functions of animal microRNAs. *Nature*. 2004; 431(7006):350–5. PMID: [15372042](#)
16. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic acids research*. 2002; 30(1):207–10. PMID: [11752295](#)
17. Tomczak K, Czerwinska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemporary oncology*. 2015; 19(1A):A68–77. doi: [10.5114/wo.2014.47136](#) PMID: [25691825](#)
18. Chen X, Gao C, Li H, Huang L, Sun Q, Dong Y, et al. Identification and characterization of microRNAs in raw milk during different periods of lactation, commercial fluid, and powdered milk products. *Cell Res*. 2010; 20(10):1128–37. doi: [10.1038/cr.2010.80](#) PMID: [20548333](#)
19. Wang H, Zheng Y, Wang G, Li H. Identification of microRNA and bioinformatics target gene analysis in beef cattle intramuscular fat and subcutaneous fat. *Molecular bioSystems*. 2013; 9(8):2154–62. doi: [10.1039/c3mb70084d](#) PMID: [23728155](#)
20. Weber JA, Baxter DH, Zhang S, Huang DY, Huang KH, Lee MJ, et al. The microRNA spectrum in 12 body fluids. *Clinical chemistry*. 2010; 56(11):1733–41. doi: [10.1373/clinchem.2010.147405](#) PMID: [20847327](#)
21. Baier SR, Nguyen C, Xie F, Wood JR, Zemleni J. MicroRNAs are absorbed in biologically meaningful amounts from nutritionally relevant doses of cow milk and affect gene expression in peripheral blood mononuclear cells, HEK-293 kidney cell cultures, and mouse livers. *J Nutr*. 2014; 144(10):1495–500. doi: [10.3945/jn.114.196436](#) PMID: [25122645](#)
22. Izumi H, Kosaka N, Shimizu T, Sekine K, Ochiya T, Takase M. Bovine milk contains microRNA and messenger RNA that are stable under degradative conditions. *J Dairy Sci*. 2012; 95(9):4831–41. doi: [10.3168/jds.2012-5489](#) PMID: [22916887](#)
23. Arnold CN, Pirie E, Dosenovic P, McInerney GM, Xia Y, Wang N, et al. A forward genetic screen reveals roles for Nfkbid, Zeb1, and Ruvbl2 in humoral immunity. *Proc Natl Acad Sci U S A*. 2012; 109(31):12286–93. doi: [10.1073/pnas.1209134109](#) PMID: [22761313](#)
24. Liu R, Ma X, Xu L, Wang D, Jiang X, Zhu W, et al. Differential microRNA expression in peripheral blood mononuclear cells from Graves' disease patients. *J Clin Endocrinol Metab*. 2012; 97(6):E968–72. doi: [10.1210/jc.2011-2982](#) PMID: [22456620](#)
25. Zhang L, Hou D, Chen X, Li D, Zhu L, Zhang Y, et al. Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA. *Cell Res*. 2012; 22(1):107–26. doi: [10.1038/cr.2011.158](#) PMID: [21931358](#)
26. Snow JW, Hale AE, Isaacs SK, Baggish AL, Chan SY. Ineffective delivery of diet-derived microRNAs to recipient animal organisms. *RNA biology*. 2013; 10(7):1107–16. Epub 2013/05/15. doi: [10.4161/rna.24909](#) PMID: [23669076](#)
27. Dickinson B, Zhang Y, Petrick JS, Heck G, Ivashuta S, Marshall WS. Lack of detectable oral bioavailability of plant microRNAs after feeding in mice. *Nature biotechnology*. 2013; 31(11):965–7. doi: [10.1038/nbt.2737](#) PMID: [24213763](#)
28. Chen X, Zen K, Zhang CY. Reply to Lack of detectable oral bioavailability of plant microRNAs after feeding in mice. *Nature biotechnology*. 2013; 31(11):967–9. Epub 2013/11/12. doi: [10.1038/nbt.2741](#) PMID: [24213764](#)
29. Wang K, Li H, Yuan Y, Etheridge A, Zhou Y, Huang D, et al. The complex exogenous RNA spectra in human plasma: an interface with human gut biota? *PLoS ONE*. 2012; 7(12):e51009. doi: [10.1371/journal.pone.0051009](#) PMID: [23251414](#)
30. Gomes CP, Cho JH, Hood L, Franco OL, Pereira RW, Wang K. A Review of Computational Tools in microRNA Discovery. *Frontiers in genetics*. 2013; 4:81. doi: [10.3389/fgene.2013.00081](#) PMID: [23720668](#)
31. Leclercq M, Diallo AB, Blanchette M. Computational prediction of the localization of microRNAs within their pre-miRNA. *Nucleic acids research*. 2013; 41(15):7200–11. doi: [10.1093/nar/gkt466](#) PMID: [23748953](#)
32. Batuwita R, Palade V. microPred: effective classification of pre-miRNAs for human miRNA gene prediction. *Bioinformatics*. 2009; 25(8):989–95. doi: [10.1093/bioinformatics/btp107](#) PMID: [19233894](#)
33. Wang H, Xiao S, Wang M, Kim NH, Li H, Wang G. In silico identification of conserved microRNAs and their targets in bovine fat tissue. *Gene*. 2015; 559(2):119–28. doi: [10.1016/j.gene.2015.01.021](#) PMID: [25626178](#)
34. Xuan P, Han K, Guo Y, Li J, Li X, Zhong Y, et al. Prediction of potential disease-associated microRNAs based on random walk. *Bioinformatics*. 2015.



35. Blevins R, Bruno L, Carroll T, Elliott J, Marcais A, Loh C, et al. microRNAs Regulate Cell-to-Cell Variability of Endogenous Target Gene Expression in Developing Mouse Thymocytes. *PLoS genetics*. 2015; 11(2):e1005020. doi: [10.1371/journal.pgen.1005020](https://doi.org/10.1371/journal.pgen.1005020) PMID: [25714103](https://pubmed.ncbi.nlm.nih.gov/25714103/)
36. Cui J, Eldredge JB, Xu Y, Puett D. MicroRNA expression and regulation in human ovarian carcinoma cells by luteinizing hormone. *PLoS One*. 2011; 6(7):e21730. doi: [10.1371/journal.pone.0021730](https://doi.org/10.1371/journal.pone.0021730) PMID: [21765906](https://pubmed.ncbi.nlm.nih.gov/21765906/)
37. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012; 28(23):3150–2. doi: [10.1093/bioinformatics/bts565](https://doi.org/10.1093/bioinformatics/bts565) PMID: [23060610](https://pubmed.ncbi.nlm.nih.gov/23060610/)
38. Hsu SD, Lin FM, Wu WY, Liang C, Huang WC, Chan WL, et al. miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic acids research*. 2011; 39(Database issue): D163–9. doi: [10.1093/nar/gkq1107](https://doi.org/10.1093/nar/gkq1107) PMID: [21071411](https://pubmed.ncbi.nlm.nih.gov/21071411/)
39. Wang X, El Naqa IM. Prediction of both conserved and nonconserved microRNA targets in animals. *Bioinformatics*. 2008; 24(3):325–32. PMID: [18048393](https://pubmed.ncbi.nlm.nih.gov/18048393/)
40. Dai X, Zhao PX. psRNATarget: a plant small RNA target analysis server. *Nucleic acids research*. 2011; 39(Web Server issue):W155–9. doi: [10.1093/nar/gkr319](https://doi.org/10.1093/nar/gkr319) PMID: [21622958](https://pubmed.ncbi.nlm.nih.gov/21622958/)
41. Aoki KF, Kanehisa M. Using the KEGG database resource. *Current protocols in bioinformatics / editorial board, Andreas D Baxevanis [et al]*. 2005;Chapter 1:Unit 1 12.
42. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. *Nucleic acids research*. 2006; 34(Database issue):D535–9. PMID: [16381927](https://pubmed.ncbi.nlm.nih.gov/16381927/)
43. Lopes Ide O, Schliep A, de Carvalho AC. The discriminant power of RNA features for pre-miRNA recognition. *BMC Bioinformatics*. 2014; 15:124. doi: [10.1186/1471-2105-15-124](https://doi.org/10.1186/1471-2105-15-124) PMID: [24884650](https://pubmed.ncbi.nlm.nih.gov/24884650/)
44. Xuan P, Guo M, Liu X, Huang Y, Li W, Huang Y. PlantMiRNAPred: efficient classification of real and pseudo plant pre-miRNAs. *Bioinformatics*. 2011; 27(10):1368–76. doi: [10.1093/bioinformatics/btr153](https://doi.org/10.1093/bioinformatics/btr153) PMID: [21441575](https://pubmed.ncbi.nlm.nih.gov/21441575/)
45. Mathelier A, Carbone A. Large scale chromosomal mapping of human microRNA structural clusters. *Nucleic acids research*. 2013; 41(8):4392–408. doi: [10.1093/nar/gkt112](https://doi.org/10.1093/nar/gkt112) PMID: [23444140](https://pubmed.ncbi.nlm.nih.gov/23444140/)
46. Lertampiporn S, Thammarongtham C, Nukoolkit C, Kaewkamnerdpong B, Ruengjitchachawalya M. Identification of non-coding RNAs with a new composite feature in the Hybrid Random Forest Ensemble algorithm. *Nucleic acids research*. 2014; 42(11):e93. doi: [10.1093/nar/gku325](https://doi.org/10.1093/nar/gku325) PMID: [24771344](https://pubmed.ncbi.nlm.nih.gov/24771344/)
47. Hofacker IL. Vienna RNA secondary structure server. *Nucleic Acids Res*. 2003; 31(13):3429–31. PMID: [12824340](https://pubmed.ncbi.nlm.nih.gov/12824340/)
48. Denman RB. Using RNAFOLD to predict the activity of small catalytic RNAs. *BioTechniques*. 1993; 15(6):1090–5. PMID: [8292343](https://pubmed.ncbi.nlm.nih.gov/8292343/)
49. Janssen S, Giegerich R. The RNA shapes studio. *Bioinformatics*. 2015; 31(3):423–5. doi: [10.1093/bioinformatics/btu649](https://doi.org/10.1093/bioinformatics/btu649) PMID: [25273103](https://pubmed.ncbi.nlm.nih.gov/25273103/)
50. Knudsen V, Caetano-Anolles G. NOBAL: a web server for character coding of geometrical and statistical features in RNA structure. *Nucleic acids research*. 2008; 36(Web Server issue):W85–90. doi: [10.1093/nar/gkn220](https://doi.org/10.1093/nar/gkn220) PMID: [18448469](https://pubmed.ncbi.nlm.nih.gov/18448469/)
51. Steffen P, Voss B, Rehmsmeier M, Reeder J, Giegerich R. RNASHAPES: an integrated RNA analysis package based on abstract shapes. *Bioinformatics*. 2006; 22(4):500–3. PMID: [16357029](https://pubmed.ncbi.nlm.nih.gov/16357029/)
52. Li Y, Qiu C, Tu J, Geng B, Yang J, Jiang T, et al. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic acids research*. 2014; 42(Database issue): D1070–4. doi: [10.1093/nar/gkt1023](https://doi.org/10.1093/nar/gkt1023) PMID: [24194601](https://pubmed.ncbi.nlm.nih.gov/24194601/)
53. Ruepp A, Kowarsch A, Theis F. PhenomiR: microRNAs in human diseases and biological processes. *Methods in molecular biology*. 2012; 822:249–60. doi: [10.1007/978-1-61779-427-8\\_17](https://doi.org/10.1007/978-1-61779-427-8_17) PMID: [22144204](https://pubmed.ncbi.nlm.nih.gov/22144204/)
54. Schulz MH, Pandit KV, Lino Cardenas CL, Ambalavanan N, Kaminski N, Bar-Joseph Z. Reconstructing dynamic microRNA-regulated interaction networks. *Proc Natl Acad Sci U S A*. 2013; 110(39):15686–91. doi: [10.1073/pnas.1303236110](https://doi.org/10.1073/pnas.1303236110) PMID: [23986498](https://pubmed.ncbi.nlm.nih.gov/23986498/)