miR2Trait: an integrated resource for investigating miRNA-disease associations

Poornima Babu and Ashok Palaniappan

Department of Bioinformatics, School of Chemical and Biotechnology, SASTRA University, Thanjavur, Tamil Nadu, India

ABSTRACT

MicroRNAs are key components of cellular regulatory networks, and breakdown in miRNA function causes cascading effects leading to pathophenotypes. A better understanding of the role of miRNAs in diseases is essential for human health. Here, we have devised a method for comprehensively mapping the associations between miRNAs and diseases by merging on a common key between two curated omics databases. The resulting bidirectional resource, miR2Trait, is more detailed than earlier catalogs, uncovers new relationships, and includes analytical utilities to interrogate and extract knowledge from these datasets. miR2Trait provides resources to compute the disease enrichment of a user-given set of miRNAs and analyze the miRNA profile of a specified diseasome. Reproducible examples demonstrating use-cases for each of these resource components are illustrated. Furthermore we used these tools to construct pairwise miRNA-miRNA and disease-disease enrichment networks, and identified 23 central miRNAs that could underlie major regulatory functions in the human genome. miR2Trait is available as an open-source command-line interface in Python3 (URL: https://github.com/miR2Trait) with a companion wiki documenting the scripts and data resources developed, under MIT license for commercial and non-commercial use. A minimal web-based implementation has been made available at https://sas.sastra.edu/pymir18. Supplementary information is available at: https://doi.org/10.6084/m9.figshare.8288825.v3.

INTRODUCTION

MicroRNAs (miRNAs) are key elements of post-transcriptional regulation in the genomic architecture of both prokaryotes and eukaryotes. They are short non-coding RNAs of about 18–25 nucleotides, first observed in *C. elegans* (*Lee, Feinbaum & Ambros, 1993*). In their canonical role in RNA silencing, the miRNA binds to the cognate mRNA and destabilizes it, thereby priming the transcript for degradation. A single miRNA is capable of silencing the expression of several genes by relaxing the specificity of hybridisation. MiRNA-based regulation plays crucial roles in health, and miRNA dysregulation is a common mechanism in the etiology of complex diseases, viz. cardio-vascular, auto-immune and neuro-degenerative diseases, and cancers (*Min et al., 2009*; *Calin et al., 2004*; *Tai, 2011*). About half of all miRNA genes are present in cancer-related genetic loci (*Calin et al., 2004*),
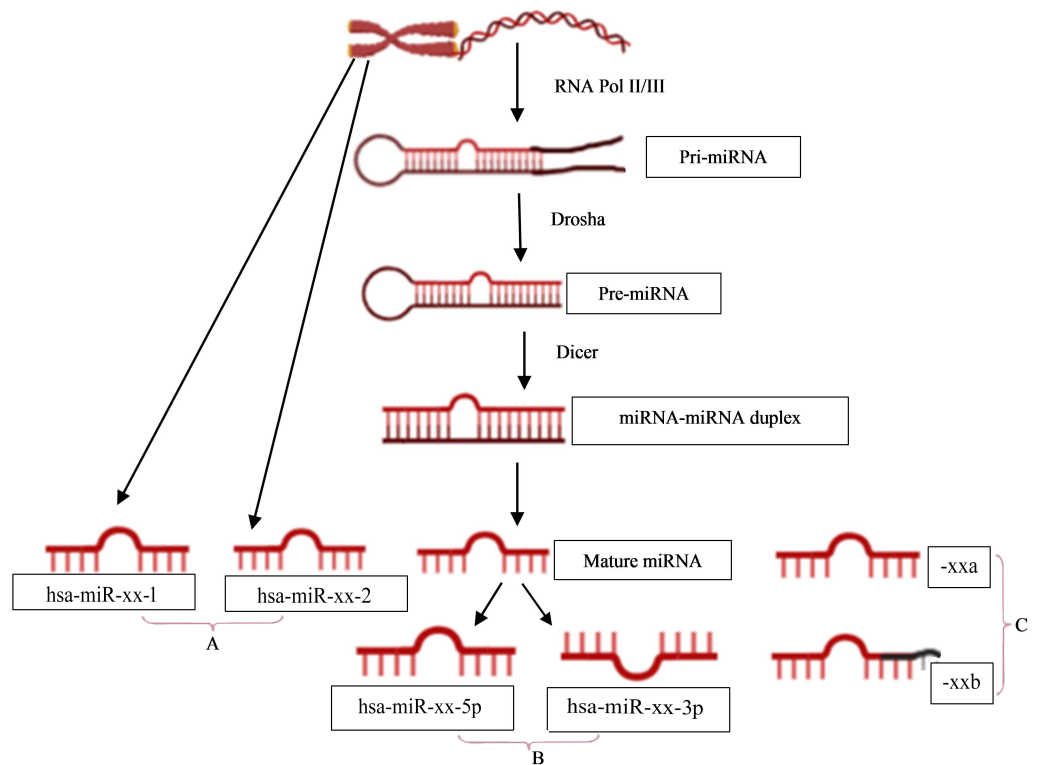
Figure 1 **miRNA nomenclature mirrors the pathway of its biogenesis.** (A) identical miRNAs encoded by different genetic loci (1,2, …); (B) source arm of the miRNA duplex (5p or 3p); and (C) nearly identical miRNAs that differ in one or two positions (a,b, …).

Full-size ☑ DOI: 10.7717/peerj.14146/fig-1

and they are known to play key roles in tumorigenesis and cancer progression (*Liu et al., 2018*; *Hayes, Peruzzi & Lawler, 2014*; *MacFarlane & Murphy, 2010*).

The naming of human miRNAs in the miRBase registry (https://www.mirbase.org) is related to the miRNA maturation pathway (Fig. 1). In brief, miRNA subtyping with numbers (−1, 2, …) indicates identical miRNAs that derive from different genomic regions; miRNA subtyping with letters (a, b, …) indicates nearly identical miRNAs that differ in only one or two positions; and miRNA designations -5p and -3p refer to the source arm of the miRNA duplex from which the mature miRNA is derived. It is the mature miRNA that binds to the mRNA and inhibits/regulates protein translation (*MacFarlane & Murphy, 2010*). It is noteworthy that the sequence responsible for miRNA-specificity of the mRNA is not translated, thus separating the protein-coding region of a gene from the miRNA-binding site in sequence space.

Existing databases of miRNA–disease associations involve tedious manual curation or are limited in scope and evolution. miR2Disease was one of the earliest catalogues of miRNA-disease associations, providing relationships between 349 miRNAs and 163 diseases (*Jiang et al., 2009*). HMDD was developed by *Li et al. (2014)* to provide a more comprehensive database of miRNA-disease associations. It used text mining followed by manual annotation for identifying miRNA-disease connections. The current version of

HMDD, version 3.2, contained associations between 1206 miRNAs and 893 diseases, and included miRNA subtype information (*Huang et al, 2019*). Only one arm of the mature miRNA, indicating the -5p or -3p source, is predominant *in vivo*, a vital annotation that is missing in the HMDD databases. In this work, we have addressed the limitations of previous efforts, and designed an algorithm to uncover the spectrum of miRNA–disease associations. Our approach is based on integrating two expert-curated databases on a common key. The resulting resource, miR2Trait, is comprehensive and representative, uncovers novel relations, and provides assorted tools for investigating miRNA-disease associations. miR2Trait can be queried by both miRNA and disease, and is available as both a web server (https://sas.sastra.edu/pymir18) and command-line resource (https://github.com/miR2Trait/ or http://doi.org/10.5281/zenodo.7002878).

MATERIALS & METHODS

miRNA: disease mapping

Two expert curated databases were used as the source databases in the creation of miR2Trait: (1) miRTarbase, a database of experimentally validated miRNA–target interactions (MTIs), with >13,400 entries (*Chou et al., 2018*); and (2) DisGeNET, a widely used and standardized knowledge management platform with >1,000,000 gene–disease associations (*Piñero et al., 2015*). Both databases involved the mapping of genetic information, in one case to miRNAs, and in the other, to diseases. This provided a clue to relating miRNAs and disease via the bridge of genetic information (Fig. 2). In-house Python scripts were used to extract the gene:miRNA and the gene:disease mappings from the respective databases as dictionary data structures. The two dictionaries were merged on 'gene' to establish a dictionary of miRNA:disease mappings. Similarly, an inverse dictionary of disease:miRNA mappings was created.

Database creation

The dictionaries obtained above were used to populate relational tables in the construction of two databases: (i) miRNAs and their associated diseases, and (ii) diseases and their associated miRNAs. MySQL was used for encoding the databases.

Spectrum width calculation

The number of diseases with which an miRNA is associated is indicative of the breadth of its regulatory impact, which could flag master regulator miRNAs (or regulatory hubs). The disease spectrum width (DSW) of the ith miRNA is calculated after *Qiu, Chen & Cui (2012)*:

$$\mathrm{DSW}(i) = d_i/D_N \tag{1}$$

where d_i is the number of diseases associated with miRNA i and D_N is the total number of diseases in the database. Conversely the number of miRNAs associated with a given disease could be indicative of a complex multifactorial pathology, and the miRNA spectrum width (MSW) of the jth disease is given by:
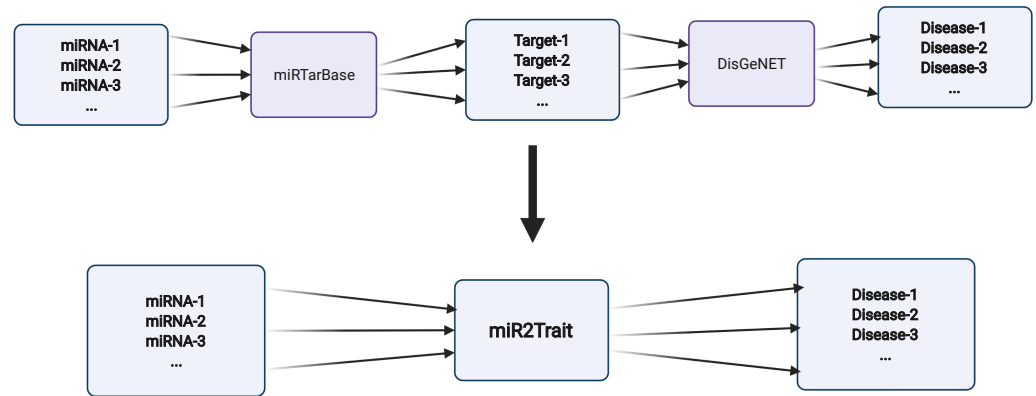
$$\mathrm{MSW}(j) = m_j/m_N \tag{2}$$

Figure 2 **Design of miR2Trait.** A back-to-back double bipartite network could be reduced to a single bipartite graph by reductive mapping on the common key. Here, the target nodes of one database serve as the source nodes of the second database. By bridging on these 'genes', a mapping between the source nodes of the first database (miRNAs) and the target nodes of the second database (diseases) could be obtained. Created with BioRender.com.

Full-size 🖼 DOI: 10.7717/peerj.14146/fig-2

where m_j is the number of miRNAs associated with the jth disease and M_N isthe total number of miRNAs in the database.

miRNA list enrichment analysis

Given a list of miRNAs that are collectively dys-regulated, it would be of interest to identify which diseases would be enriched (or more likely to occur). Statistical techniques to quantify such enrichment for genes as well as miRNAs exist (*Li et al., 2018*; *Backes et al., 2016*; *Çorapçıoğlu & Oğul, 2015*; *Rivals et al., 2007*). Here we implement an open-source tool to identify diseases in the miR2Trait database that are enriched for an input set of miRNAs. The method uses a hypergeometric test to quantify the enrichment of each disease in miR2Trait for the given set of miRNAs (in this case identical to the Fisher's exact test) (*Fisher, 1922*). The test follows the construction of the Fisher contingency table for each disease, as illustrated in Table 1.

The statistical significance (*p*-value) of enrichment is computed using Python scipy routines (http://www.scipy.org/), and diseases that pass a user-specified adjusted *p*-value threshold (with an option to choose among three methods for multiple hypothesis correction) are returned sorted by significance. In addition to significance, both the web server and standalone tool calculate the effect size (log-fold) of each enrichment.

Diseasome analysis

Given an input set of diseases related in some way (called a diseasome), it would be of interest to find shared dysregulated miRNAs. We approached this problem with an abundance analysis of miRNAs in the diseasome, and identifying miRNAs that passed a user-specified count. Such a set of miRNAs could constitute master regulators acting on common pathways in the specified set of diseases. If we are interested in the miRNA spectra of umbrella disease terms such as 'carcinoma', instead of a specific set of diseases, this is also allowed. In this case, diseases containing the general keyword are first identified,

Table 1 Fisher's contingency table for miRNA list enrichment analysis for one disease in the database. Each cell value represents the count for the intersecting row and column.

miR2Trait	miRNAs associated with disease	miRNAs not associated with disease	(Sum)
miRNAs in input	a	b	a+b
miRNAs not in input	c	d	c+d
(Sum)	a+c	b+d	n

and then the miRNA abundance for this diseasome is computed in a manner similar to above. A CLI script to compute the miRNA profile of a specified diseasome is provided with miR2Trait (https://github.com/miR2Trait).

miR2Trait web server and standalone tools

A web interface using PHP to connect the user HTML front-end with the MySQL backend was created to offer access to all the services developed, including the two databases and the assorted tools. MiRNA list enrichment analysis (miLEA) was implemented using the python DBI module, and receives the user's miRNA list from the HTML form via PHP intermediate. The command-line version of the enrichment analysis provides adjusted *p*-value calculation using Bonferroni, Holm-Bonferroni or Benjamini–Hochberg (default) correction. Other CLI scripts developed include:

(i) Query standalone miR2Trait by miRNA or disease

(ii) Retrieve miRNA-specific information from miRBase by user-given miRNA name(s) or sequence(s).

(iii) Network creation: creating an miRNA-miRNA/disease-disease adjacency network from miR2Trait dictionaries

(iv) Postprocessing constructed networks for consensus central nodes (miRNAs or diseases)

The user is referred to the wiki (https://github.com/miR2Trait/miR2Trait/wiki) for reproducible examples and use-cases of all the scripts developed.

RESULTS

The mapping sizes extracted from the respective databases is given in Table 2. In summary, 2,599 miRNAs were mapped to 15,062 genes using miRTarbase, and 8,819 genes were mapped to 13,075 diseases using DisGeNET. Merging the two relations yielded a mapping between 2,595 miRNAs and 11,689 diseases (Table 2). This mapping laid the foundation for the miRNA:disease and disease:miRNA databases, which could be freely downloaded as csv files from aforementioned URLs. The density function of the mappings is shown in Fig. 3.

Using the miRNA:disease db, miRNAs with the top ten DSWs were identified (Table 3). Such miRNAs are potential master (or hub) regulators at the intersection of multiple pathways, indicatively contributing to one-fifth of all diseases. Using the disease:miRNA db, diseases with the top ten MSWs were identified (Table 4). Complex neurological

Table 2 Establishing the miR2Trait mappings. Only four miRNAs from miRTarBase did not map to any disease in the final analysis.

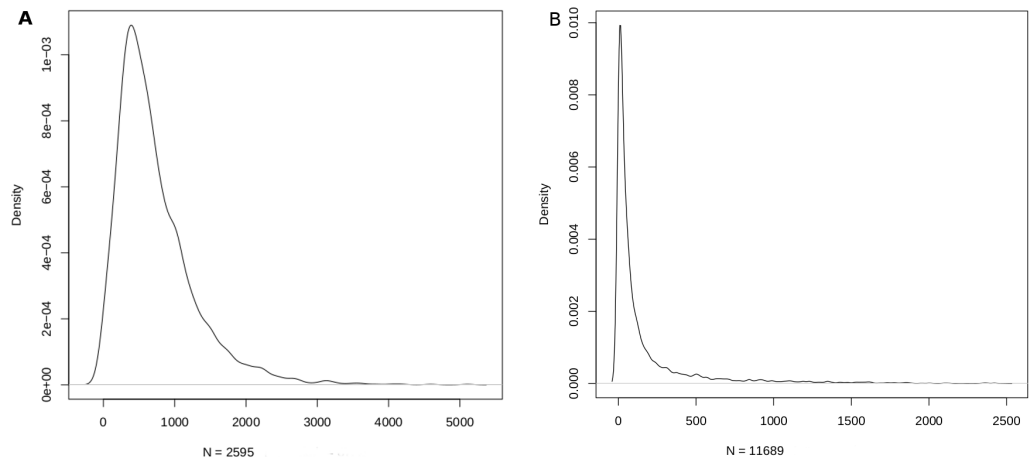| Database/resource	Mapping relation	Mapping size
miRTarBase	miRNAs → Target genes	2,599 → 15,062
DisGeNET	Genes → Diseases	8,819 → 13,075
miR2Trait	miRNAs → Diseases	2,595 → 11,689

Figure 3 Density function plots of (A) miRNA-disease; and (B) disease-miRNA mappings. The mode of the number of diseases/miRNA is about 389 and the mode of the number of miRNAs/disease is about 9.

Table 3 DSW of miRNAs. miRNA corresponding to the top ten DSWs are shown.

S.No	miRNA	DSW	No. of associated diseases
1	hsa-miR-335-5p	0.22	2,530
2	hsa-miR-26b-5p	0.20	2,321
3	hsa-miR-124-3p	0.19	2,175
4	hsa-miR-16-5p	0.18	2,110
5	hsa-miR-92a-3p	0.17	1,974
6	hsa-miR-1-3p	0.16	1,913
7	hsa-miR-17-5p	0.16	1,912
8	hsa-let-7b-5p	0.16	1,864
9	hsa-miR-155-5p	0.16	1,856
10	hsa-miR-93-5p	0.16	1,766

conditions sweep >90% of all miRNAs, reflecting multifactorial, far from understood aetiologies.

The miRNA Enrichment Analyzer computes both the odds-ratio size and significance *p*-value for a given set of miRNAs. The input could be variable in the number of miRNAs provided. We used "hsa-miR-346 hsa-miR-26a-5p hsa-miR-7-5p hsa-miR-34a-5p" as

Table 4 MSW of diseases. Diseases corresponding to the top ten MSWs are shown.

S.No	Disease	MSW	No. of associated miRNAs
1	Autosomal recessive predisposition	0.97	2,491
2	Schizophrenia	0.96	2,466
3	Intellectual disability	0.95	2,452
4	Low intelligence	0.94	2,421
5	Dull intelligence	0.94	2,421
6	Poor school performance	0.94	2,421
7	Mental deficiency	0.94	2,421
8	Mental retardation	0.94	2,421
9	Mental and motor retardation	0.92	2,381
10	Global developmental delay	0.92	2,381

Table 5 miRNA list enrichment analysis. Results for the query: "hsa-miR-2355-3p hsa-miR-214-3p hsa-miR-4801", and containing the word 'neoplasm' are shown, sorted by adjusted significance (*q*-value). An odds ratio (effect size) of 'Inf' indicates that the entire set of miRNAs is contained in the disease definition.

Disease	q-value	Odds-ratio
Colorectal Neoplasms	0.014	Inf
Neoplasm of the anterior pituitary	0.020	74.50
Stomach Neoplasms	0.022	Inf
Prostatic Neoplasms	0.027	Inf
Malignant neoplasm of ovary	0.034	43.32
Biliary Tract Neoplasm	0.035	42.83

a query with an adj. *p*-value (*i.e.*, *q*-value) cutoff of 0.05. This returned a set of 4,121 significantly enriched diseases. As another example, we used a set of three random miRNAs: "hsa-miR-2355-3p hsa-miR-214-3p hsa-miR-4801" as a query. This returned 830 significantly enriched diseases (Table 5). Dysregulation of key cancer-associated miRNAs (*Chan, Prado & Weidhaas, 2011*) could contribute to the emergence of cancer hallmarks (*Hanahan & Weinberg, 2011*) and could be tested likewise.

Analysis of diseasomes for over-represented miRNAs could yield valuable hypotheses for further research, in terms of both pathophysiology and therapeutic options. In this line, we pursued the investigation of diseasomes related to cancer. Since the evidence for miRNA involvement in all stages of cancer including tumorigenesis, progression and metastasis is well-established, we probed miR2Trait for the keyword 'Neoplasm' diseasome, and performed an miRNA occurrence analysis. We also performed occurrence analysis for diseasomes identified using the keywords 'Carcinoma', 'Sarcoma', 'Lymphoma', and 'Leukemia'. Table 6 shows the results of these analyses for 25 miRNAs ranked by the 'Neoplasm' occurrence analysis. MiRNAs documented in the literature as playing key roles in various cancers appeared in the top miRNAs for each keyword (*Sandoval-Bórquez et al., 2017*; *Miyamoto et al., 2016*; *Zhang et al., 2015*; *Wang et al., 2018*; *Cloonan et al., 2008*; *Xu et al., 2014*; *Robertson et al., 2014*; *Fabbri et al., 2016*; *Garzon et al., 2008*; *Marcucci et al.,

Table 6 Diseasome analysis of miRNA over-abundance. Statistics for only the top 25 miRNAs from 'Neoplasm' occurrence analysis are shown with respect to cancer-related keyword searches.

'Neoplasm' rank	miRNA	Carcinoma	Leukemia	Lymphoma	Sarcoma
1	hsa-miR-34a-5p	43	21	17	16
2	hsa-miR-16-5p	47	24	15	18
3	hsa-miR-124-3p	42	16	18	21
4	hsa-miR-335-5p	41	25	17	20
5	hsa-miR-21-5p	45	12	17	15
6	hsa-miR-155-5p	48	17	17	18
7	hsa-miR-19a-3p	38	16	15	13
8	hsa-miR-17-5p	47	24	16	16
9	hsa-miR-125b-5p	36	16	15	21
10	hsa-let-7b-5p	35	15	13	20
11	hsa-miR-221-3p	35	15	–	17
12	hsa-miR-15a-5p	40	16	13	16
13	hsa-miR-130b-3p	33	11	11	11
14	hsa-miR-26b-5p	39	15	14	18
15	hsa-miR-181a-5p	37	17	14	16
16	hsa-miR-19b-3p	36	13	14	13
17	hsa-miR-106a-5p	42	13	17	15
18	hsa-miR-192-5p	41	20	11	14
19	hsa-miR-106b-5p	43	15	16	11
20	hsa-miR-218-5p	37	26	15	16
21	hsa-miR-24-3p	36	14	14	12
22	hsa-miR-27a-3p	41	21	12	10
23	hsa-miR-20a-5p	43	20	16	16
24	hsa-miR-454-3p	32	10	10	10
25	hsa-miR-30a-5p	47	16	13	14

2008; *Calin et al., 2005*; *Stefano et al., 2010*; *Jima et al., 2010*). Databases devoted to miRNA associations in cancer have been developed (*Yang et al., 2017*; *Sarver et al., 2018*; *Ahmed et al., 2018*), and our work here would augment efforts in this direction. Expression of specific microRNAs has been recognized as a vital diagnostic/prognostic biomarker and therapeutic target in a variety of cancers (*Blenkiron et al., 2007*; *Sempere et al., 2007*; *Michael et al., 2006*; *Iorio et al., 2005*; *Lowery et al., 2009*; *Lebanony et al., 2009*; *Ueda, 2010*; *Fridman et al., 2010*; *Gutiérrez et al., 2010*).

DISCUSSION

Benchmarking

Dedicated miRNA omics databases have been developed to aid researchers unravel the role of miRNAs in biological processes (for *e.g.*, see *Ruepp et al., 2010*; *Keller et al., 2011*). Computational methods have also been advanced for inferring miRNA-disease connections (*Chen, Liu & Yan, 2012*; *Xuan et al., 2013*; *Qu et al., 2018*). HMDDv3.2 is a
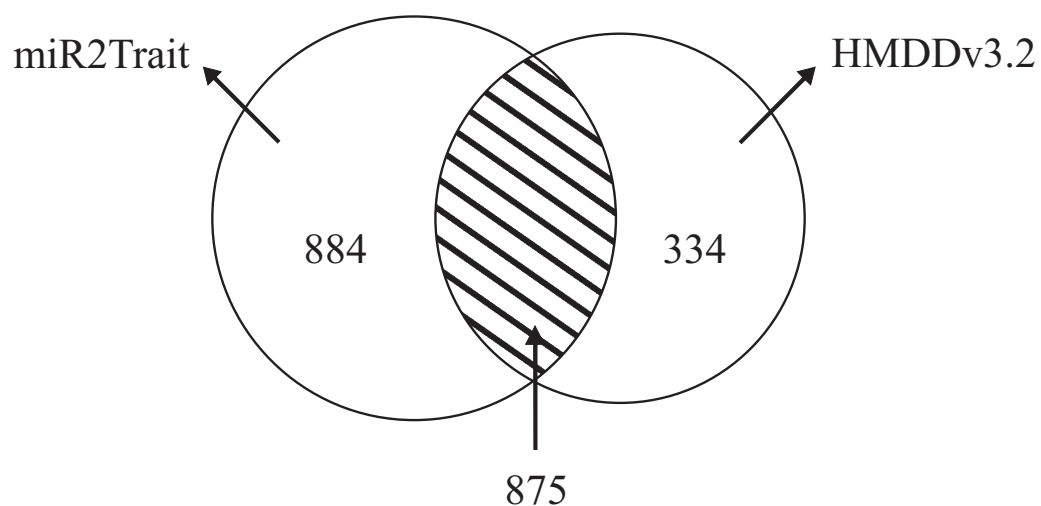
manually curated database of miRNA-disease associations based on text-mining. While miR2Trait contains the full mature miRNA information, HMDD does not provide subtype information such as -5p or -3p; 1a or 1b. The miRNA arm designation (5p or 3p) is a major ingredient of disease etiology, and key to naming miRNAs (for *e.g.,* the ones with the top DSWs in Table 3). To compare the databases, it is therefore necessary to 'normalize' the miRNA nomenclature between the databases (by suppressing the subtype information in miR2Trait). This contracted the unique miRNAs in miR2Trait to 1,759 in number. Following this, the databases were compared, and the results are shown in Fig. 4. Even after contraction, miR2Trait appears more comprehensive in the catalog of miRNA-disease associations. miR2Trait-unique miRNAs are 2.5-fold more in number than HMDDv3.2 - unique miRNAs (884 vs 334). The unique miRNA complements suggest directions for research. Some of the diseases missed by miR2Trait could be those that involve deleterious mutations in the miRNA-binding region of the gene itself. Finally, the significant size of the overlap between the two databases increases confidence in the methods advanced here, which are also easily extensible with updates to the underlying primary databases. These observations further suggest the use of miR2Trait to complement HMDD.

Network analysis
The availability of two-way mappings between miRNAs and diseases with various analysis tools paves the way to varied investigations. One direction would be to identify pairs of miRNAs that lead to clusters of similar disease phenotypes. We addressed this question by using the miRNA enrichment analyzer discussed herein to compute the disease enrichment scores (2-tuples of BH-adjusted *p*-value and odds-ratio) of each miRNA-miRNA pair

using Fisher's exact test. This process is iterated over the entire set of miRNA-miRNA pairs (2595*2595) to obtain an adjacency matrix of ~3 million interactions. This matrix was used to construct the corresponding network made of significant edges weighted by the odds-ratio. The interactions were sorted by -log$_{10}$ p-value and the top 10,000 interactions were used to construct the miRNA-miRNA adjacency network, yielding 2,442 nodes (and 10,000 edges). The network was read into Cytoscape (*Shannon et al., 2003*), and centrality analysis was performed using CytoNCA (*Tang et al., 2015*). The weighted versions of the following seven centrality measures were used: betweenness, closeness, eigenvector, degree, network, information, and local average centralities. The top 100 miRNAs from each centrality measure were used to obtain the consensus (five or more measures) top miRNAs (Table 7; algorithm details and Python script given here: https://github.com/miR2Trait/miR2Trait/wiki/Top-Nodes-from-Adjacency-Networks). Edge-percolation centrality of the network was also computed using cytoHubba to rank the miRNAs (*Chin et al., 2014*). The top 23 miRNAs contained 14 consensus miRNAs and the top 75 contained all but one of the consensus miRNAs. Such a consensus list could identify the critical regulatory miRNAs in human for further investigation. A disease-disease adjacency network was similarly constructed, and both the above adjacency networks and the scripts to perform the network construction and analysis are available from miR2Trait (https://github.com/miR2Trait).

Deployment

The miR2Trait web server and standalone CLI could be queried by either disease or miRNA. The standard use-case would be to identify miRNAs associated with some disease. For instance, a search by disease for 'Myocardial infarction' returned 1,445 associated miRNAs. The top hits included hsa-miR-302a-3p, hsa-miR-302b-3p, hsa-miR-302c-3p, hsa-miR-302d-3p, hsa-miR-367-3p and hsa-miR-367-5p, all of which are well-documented in the literature (*Sun et al., 2017*). The top hits for a search by disease of 'Alzheimer's disease' included hsa-miR-9, hsa-miR-29a, hsa-miR34a, hsa-miR-106b, hsa-miR-125b, hsa-miR-146a, and hsa-miR-155, again all of which are well-documented in the literature (*Leidinger et al., 2013*). As a last example, the top hits for a search by disease of 'Diabetes mellitus' included hsa-miR-577, hsa-miR-37a, hsa-miR-375, hsa-miR-181a, hsa-miR-17, and hsa-miR-24, all of which are reported in the literature (*Kim & Zhang, 2019*; *Guay et al., 2011*). The search term is case-insensitive, matches anywhere, and interpreted according to regex grammar (https://dev.mysql.com/doc/refman/8.0/en/regexp.html#operator_regexp). Such searches might yield miRNAs lacking documented disease connections, and offer hypotheses for future investigations. In this regard, miR2Trait provides a script to retrieve known information about miRNAs from miRBase upon querying with the miRNA name or sequence (https://github.com/miR2Trait/miR2Trait/wiki/Querying-miRNA-information-from-miRBase).

In summary, we have developed miR2Trait, a comprehensive resource for investigating miRNA-disease associations, based on a simple yet surprisingly effective technique for fusing two curated primary databases. One outcome of the approach has been the enhanced

Table 7 Consensus identification of key miRNAs in disease development. Seven centrality measures were used to identify the top 100 nodes in the miRNA-miRNA adjacency network and their consensus is determined based on the number of measures that agree upon an miRNA (agreement ≥ 5). The consensus miRNAs appear central even by the Edge-percolation centrality (EPC) measure.

SNo	miRNA	Agreement	EPC rank, score
1	hsa-miR-1253	7	36, 208
2	hsa-miR-6755-5p	7	13, 247
3	hsa-miR-3911	6	62, 188
4	hsa-miR-6804-5p	6	16, 232
5	hsa-miR-4689	6	12, 247
6	hsa-miR-873-5p	6	11, 247
7	hsa-miR-548aq-3p	5	8, 254
8	hsa-miR-548x-3p	5	6, 257
9	hsa-miR-548ah-3p	5	3, 260
10	hsa-miR-6885-3p	5	28, 222
11	hsa-miR-548aj-3p	5	7, 255
12	hsa-miR-548j-3p	5	5, 258
13	hsa-miR-548ae-3p	5	2, 261
14	hsa-miR-543	5	41, 203
15	hsa-miR-194-5p	5	20, 229
16	hsa-miR-3140-3p	5	73, 181
17	hsa-miR-3156-3p	5	43, 202
18	hsa-miR-548am-3p	5	4, 260
19	hsa-miR-892b	5	10, 253
20	hsa-miR-4280	5	n/a
21	hsa-miR-802	5	25, 224
22	hsa-miR-877-5p	5	1, 266
23	hsa-miR-425-3p	5	26, 223

interpretability and transparency of the resource constituents, aspects that are lost in more complex methods to arrive at reliable miRNA-disease associations.

CONCLUSIONS

MicroRNAs play key roles in health and the development of disease by virtue of their crucial regulatory activities. To document miRNA-disease associations, we used a novel method integrating two curated and validated sources of miRNA-gene and gene-disease relations, and created vastly expanded databases of miRNA-disease and disease-miRNA associations. A set of tools to interrogate these databases and uncover novel findings has been developed. Taken together, miR2Trait provides a starting point for the investigation of the miRNA-ome in the context of disease. A web-server interface to all the functionalities has been developed and is available at https://sas.sastra.edu/pymir18/. The service allows for flexible querying using regular expression-based pattern matching. It includes a tool for miRNA-enrichment analysis that would enable the identification of statistically over-represented diseases in a user-given list of miRNAs. This would help formulate hypotheses of miRNA-mediated

common mechanisms underlying multiple disease dysregulation pathologies. A number of examples have been discussed to illustrate the use-cases for the resource. The resource has also been benchmarked with the best alternative miRNA-disease database (HMDDv3.2). Further, the source code for the entire project is made freely available, and includes a plethora of tools for working with the resource, along with an extensive wiki. All the data developed in the project are freely available as supplementary information, including the miRNA-miRNA and disease-disease adjacency networks.

ACKNOWLEDGEMENTS

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Poornima Babu performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Ashok Palaniappan conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:
    The code is available at GitHub: https://github.com/miR2Trait.
    The data is available at FigShare: Babu, Poornima; Palaniappan, Ashok (2021): miR2Trait: an integrated resource for mining miRNA - disease associations. figshare. Online resource. https://doi.org/10.6084/m9.figshare.8288825.v3.

REFERENCES

Ahmed M, Nguyen H, Lai T, Kim DR. 2018. miRCancerdb: a database for correlation analysis between microRNA and gene expression in cancer. *BMC Research Notes* **11(1)**:103 DOI 10.1186/s13104-018-3160-9.

Backes C, Khaleeq QT, Meese E, Keller A. 2016. miEAA: microRNA enrichment analysis and annotation. *Nucleic Acids Research* **44(W1,8)**:W110–W116 DOI 10.1093/nar/gkw345.

Blenkiron C, Goldstein LD, Thorne NP, Spiteri I, Chin S-F, Dunning MJ, Barbosa-Morais NL, Teschendorff AE, Green AR, Ellis IO, Tavaré S, Caldas C, Miska Eric A. 2007. MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. *Genome Biology* **8(10)**:R214 DOI 10.1186/gb2007-8-10-r214.

Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, Shimizu M, Rattan S, Bullrich F, Negrini M, Croce CM. 2004. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proceedings of the National Academy of Sciences of the United States of America* **101(9)**:2999–3004 DOI 10.1073/pnas.0307323101.

Calin GA, Ferracin M, Cimmino A, Di Leva G, Shimizu M, Wojcik SE, Iorio VM, Visone R, Sever NI, Fabbri M, Iuliano R, Palumbo T, Pichiorri F, Roldo C, Garzon R, Sevignani C, Rassenti L, Alder H, Volinia S, Liu C-G, Kipps TJ, Negrini M, Croce CM. 2005. A microRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *New England Journal of Medicine* **353(17)**:1793–1801 DOI 10.1056/NEJMoa050995.

Çorapçıoğlu ME, Oğul H. 2015. miSEA: microRNA set enrichment analysis. *Biosystems* **134**:37–42 DOI 10.1016/j.biosystems.2015.05.004.

Chan E, Prado DE, Weidhaas JB. 2011. Cancer microRNAs: from subtype profiling topredictors of response to therapy. *Trends in Molecular Medicine* **17(5)**:235–243 DOI 10.1016/j.molmed.2011.01.008.

Chen X, Liu MX, Yan GY. 2012. RWRMDA: predicting novel human microRNA-disease associations. *Molecular BioSystems* **8(10)**:2792–2798 DOI 10.1039/c2mb25180a.

Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. 2014. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology* **8(Suppl 4)**:S11 DOI 10.1186/1752-0509-8-S4-S11.

Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, Huang WC, Sun TH, Tu SJ, Lee WH, Chiew MY, Tai CS, Wei TY, Tsai TR, Huang HT, Wang CY, Wu HY, Ho SY, Chen PR, Chuang CH, Hsieh PJ, Wu YS, Chen WL, Li MJ, Wu YC, Huang XY, Ng FL, Buddhakosai W, Huang PC, Lan KC, Huang CY, Weng SL, Cheng YN, Liang C, Hsu WL, Huang HD. 2018. miRTarBase update: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Research* **46(D1)**:D296–D302 DOI 10.1093/nar/gkx1067.

Cloonan N, Brown MK, Steptoe AL, Wani S, Chan WL, Forrest ARR, Gabriel K, Gabrielli B, Grimmond SM. 2008. The miR-17-5p microRNA is a key regulator of the G1/S phase cell cycle transition. *Genome Biology* **9**:R127 DOI 10.1186/gb-2008-9-8-r127.

Fabbri E, Montagner G, Bianchi N, Finotti A, Borgatti M, Lampronti I, Cabrini G, Gambari R. 2016. MicroRNA miR-93-5p regulates expression of IL-8 and VEGF in neuroblastoma SK-N-AS cells. *Oncology Reports* **35**:2866–2872 DOI 10.3892/or.2016.4676.

Fisher RA. 1922. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A* **222**:309–368 DOI 10.1098/rsta.1922.0009.

Fridman E, Dotan Z, Barshack I, David BM, Dov A, Tabak S, Zion O, Benjamin S, Benjamin H, Kuker H, Avivi C, Rosenblatt K, Polak-Charcon S, Ramon J, Rosenfeld N, Spector Y. 2010. Accurate molecular classification of renal tumors using MicroRNA expression. *The Journal of Molecular Diagnostics* **12(5)**:687–696 DOI 10.2353/jmoldx.2010.090187.

Garzon R, Garofalo M, Martelli PM, Briesewitz R, Wang L, Fernandez-Cymering C, Volinia S, Liu C-G, Schnittger S, Haferlach T, Liso A, Diverio D, Mancini M, Meloni G, Foa R, Martelli FM, Mecucci C, Croce MC, Falini B. 2008. Distinctive microRNA signature of acute myeloid leukemia bearing cytoplasmic mutated nucleophosmin. *Proceedings of the National Academy of Sciences of the United States of America* **105(10)**:3945–3950 DOI 10.1073/pnas.0800135105.

Guay C, Roggli E, Nesca V, Jacovetti C, Regazzi R. 2011. Diabetes mellitus, a microRNA-related disease? *Translational Research* **157(4)**:253–264 DOI 10.1016/j.trsl.2011.01.009.

Gutiérrez NC, Sarasquete ME, Misiewicz-Krzeminska I, Delgado M, De Las Rivas J, Ticona FV, Fermiñán E, Martín-Jiménez P, Chillón C, Risueño A, Hernández JM, García-Sanz R, González M, San Miguel JF. 2010. Deregulation of microRNA expression in the different genetic subtypes of multiple myeloma and correlation with gene expression profiling. *Leukemia* **24(3)**:629–637 DOI 10.1038/leu.2009.274.

Hanahan D, Weinberg RA. 2011. Hallmarks of cancer: the next generation. *Cell* **144**:646–674 DOI 10.1016/j.cell.2011.02.013.

Hayes J, Peruzzi PP, Lawler S. 2014. MicroRNAs in cancer: biomarkers, functions and therapy. *Trends in Molecular Medicine* **20(8)**:460–469 DOI 10.1016/j.molmed.2014.06.005.

Huang Z, Shi J, Gao Y, Cui C, Zhang S, Li J, Zhou Y, Cui Q. 2019. HMDD v3.0: a database for experimentally supported human microRNA-disease associations. *Nucleic Acids Research* **47(D1)**:D1013–D1017 DOI 10.1093/nar/gky1010.

Iorio MV, Ferracin M, Liu C-G, Veronese A, Spizzo R, Sabbioni S, Magri E, Pedriali M, Fabbri M, Campiglio M, Menard S, Palazzo PJ, Rosenberg A, Musiani P, Volinia S, Nenci I, Calin AG, Querzoli P, Negrini M, Croce MC. 2005. MicroRNA gene expression deregulation in human breast cancer. *Cancer Research* **65(16)**:7065–7070 DOI 10.1158/0008-5472.CAN-05-1783.

Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, Li M, Wang G, Liu Y. 2009. miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Research* **37(Database issue)**:D98–D104 DOI 10.1093/nar/gkn714.

Jima DD, Zhang J, Jacobs C, Richards KL, Dunphy CH. 2010. Deep sequencing of the small RNA transcriptome of normal and malignant human B-cells identifies hundreds of novel microRNAs. *Blood* **116(23)**:e118–e127 DOI 10.1182/blood-2010-05-285403.

Keller A, Leidinger P, Bauer A, Elsharawy A. 2011. Toward the blood-borne miRNome of human diseases. *Nature Methods* **8**(10):841–843 DOI 10.1038/nmeth.1682.

Kim M, Zhang X. 2019. The profiling and role of miRNAs in diabetes mellitus. *Journal of Diabetes and Clinical Research* **1**(1):5–23 DOI 10.33696/diabetes.1.003.

Lebanony D, Benjamin H, Gilad S, Ezagouri M, Dov A, Ashkenazi K, Gefen N, Izraeli S, Rechavi G, Pass H, Nonaka D, Li J, Spector Y, Rosenfeld N, Chajut A, Cohen D, Aharonov R, Mansukhani M. 2009. Diagnostic assay based on hsa-miR-205 expression distinguishes squamous from nonsquamous non-small-cell lung carcinoma. *Journal of Clinical Oncology* **27**(12):2030–2037 DOI 10.1200/JCO.2008.19.4134.

Lee RC, Feinbaum RL, Ambros V. 1993. The *C*, elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**(5):843–854 DOI 10.1016/0092-8674(93)90529-Y.

Leidinger P, Backes C, Deutscher S, Schmitt K, Mueller SC, Frese K, Haas J, Ruprecht K, Paul F, Stähler C, Lang CJ, Meder B, Bartfai T, Meese E, Keller A. 2013. A blood based 12-miRNA signature of Alzheimer disease patients. *Genome Biology* **14**(7):R78 DOI 10.1186/gb-2013-14-7-r78.

Li Y, Qiu CX, Tu J, Geng B, Yang J, Jiang T, Cui Q. 2014. HMDD v, 2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Research* **42**(Database issue):D1070–D1074 DOI 10.1093/nar/gkt1023.

Li J, Han X, Wan Y, Zhang S, Zhao Y, Fan R, Cui Q, Zhou YTAM. 2.0: tool for MicroRNA set analysis. *Nucleic Acids Research* W180–W185 DOI 10.1093/nar/gky509.

Liu B, Shyr Y, Cai J, Liu Q. 2018. Interplay between miRNAs and host genes and their role in cancer. *Briefings in Functional Genomics* **18**(4):255–266 DOI 10.1093/bfgp/elz002.

Lowery AJ, Miller N, Devaney A, McNeill RE, Davoren PA, Lemetre C, Benes V, Schmidt S, Blake J, Ball G, Kerin MJ. 2009. MicroRNA signatures predict oestrogen receptor, progesterone receptor and HER2/neu receptor status in breast cancer. *Breast Cancer Research* **11**(3):R27 DOI 10.1186/bcr2257.

Marcucci G, Radmacher MD, Maharry K, Mrózek K, Ruppert AS, Paschka P, Vukosavljevic T, Whitman SP, Baldus CD, Langer C, Chang-Gong C, Carroll AJ, Powell BL, Garzon R, Croce CM, Kolitz JE, Caligiuri MA, Larson RA, Bloomfield CD. 2008. MicroRNA expression in cytogenetically normal acute myeloid leukemia. *New England Journal of Medicine* **358**(18):1919–1928 DOI 10.1056/NEJMoa074256.

Min L, Christian M-M, Bharadwaj U, Chow K-H, Yao Q, Chen C. 2009. MicroRNAs: control and loss of control in human physiology and disease. *World Journal of Surgery* **33**(4):667–684 DOI 10.1007/s00268-008-9836-x.

MacFarlane L-A, Murphy PR. 2010. MicroRNA: biogenesis, function and role in cancer. *Current Genomics* **11**(7):537–561 DOI 10.2174/138920210793175895.

Michael DM, Benz CC, Bowers J, Sensinger K, Wong L, Scott GK, Fedele V, Ginzinger D, Getts R, Haqq C. 2006. Optimized high-throughput microRNA expression profiling provides novel biomarker assessment of clinical prostate and breast cancer biopsies. *Molecular Cancer* **5**(1):24 DOI 10.1186/1476-4598-5-24.

Miyamoto K, Seki N, Matsushita R, Yonemori M, Yoshino H, Nakagawa M, Enokida H. 2016. Tumour-suppressive miRNA-26a-5p and miR-26b-5p inhibit cell aggressiveness by regulating PLOD2 in bladder cancer. *British Journal of Cancer* **115(3)**:354–363 DOI 10.1038/bjc.2016.179.

Piñero J, Queralt-Rosinach N, Bravo À, Deu-Pons J, Bauer-Mehren A, Baron M, Sanz F, Furlong L. 2015. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database* **2015**:bav028 DOI 10.1093/database/bav028.

Qiu C, Chen G, Cui Q. 2012. Towards the understanding of microRNA and environmental factor interactions and their relationships to human diseases. *Scientific Reports* **2**:318 DOI 10.1038/srep00318.

Qu Y, Zhang H, Lyu C, Liang C. 2018. LLCMDA: a novel method for predicting miRNA gene and disease relationship based on locality-constrained linear coding. *Frontiers in Genetics* **9**:576 DOI 10.3389/fgene.2018.00576.

Rivals I, Personnaz L, Taing L, Potier MC. 2007. Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics* **23(4)**:401–407 DOI 10.1093/bioinformatics/btl633.

Robertson DE, Wasylyk C, Ye T, Jung AC, Wasylyk B. 2014. The oncogenic MicroRNA Hsa-miR-155-5p targets the transcription factor ELK3 and links it to the hypoxia response. *PLOS ONE* **9(11)**:e113050 DOI 10.1371/journal.pone.0113050.

Ruepp A, Kowarsch A, Schmidl D, Buggenthin F, Brauner B, Dunger I, Fobo G, Frishman G, Montrone C, Theis FJ. 2010. PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. *Genome Biology* **11(1)**:R6 DOI 10.1186/gb-2010-11-1-r6.

Sandoval-Bórquez A, Polakovicova I, Carrasco-Véliz N, Lobos-González L, Riquelme I, Carrasco-Avino G, Bizama C, Norero E, Owen GI, Roa JC, Corvalán AH. 2017. MicroRNA-335-5p is a potential suppressor of metastasis and invasion in gastric cancer. *Clinical Epigenetics* **9**:114 DOI 10.1186/s13148-017-0413-8.

Sarver AL, Sarver AE, Yuan C, Subramanian S. 2018. OMCD: OncomiR cancer database. *BMC Cancer* **18(1)**:1223 DOI 10.1186/s12885-018-5085-z.

Sempere LF, Christensen M, Silahtaroglu A, Bak M, Heath CV, Schwartz G, Wells W, Kauppinen S, Cole CN. 2007. Altered MicroRNA expression confined to specific epithelial cell subpopulations in breast cancer. *Cancer Research* **67(24)**:11612–11620 DOI 10.1158/0008-5472.CAN-07-5019.

Shannon P, Markiel A, Ozier O. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13(11)**:2498–2504 DOI 10.1101/gr.1239303.

Stefano C, Egyházi S, Rodolfo M, Witten D, Hansson J, Larsson C, Lui Wg-O. 2010. MicroRNA expression profiles associated with mutational status and survival in malignant melanoma. *Journal of Investigative Dermatology* **130(8)**:2062–2070 DOI 10.1038/jid.2010.63.

Sun T, Dong Y-H, Du W, Shi C-Y, Wang K, Tariq M-A, Wang J-X, Li P-F. 2017. The role of microRNAs in myocardial infarction: from molecular mechanism

to clinical application. *International Journal of Molecular Sciences* **18**(4):745 DOI 10.3390/ijms18040745.

Tai Y-H. 2011. MicroRNAs in human diseases: from cancer to cardiovascular disease. *Immune Network* **11**(3):135–154 DOI 10.4110/in.2011.11.3.135.

Tang Y, Li M, Wang J, Pan Y, Wu FX. 2015. CytoNCA: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks. *Biosystems* **127**:67–72 DOI 10.1016/j.biosystems.2014.11.005.

Ueda T, Volinia S, Okumura H, Shimizu M, Taccioli C, Rossi S, Alder H, Liu C-G, Oue N, Yasui W, Yoshida K, Sasaki H, Nomura S, Seto Y, Kaminishi M, Calin GA, Croce CM. 2010. Relation between microRNA expression and progression and prognosis of gastric cancer:a microRNA expression analysis. *The Lancet Oncology* **11**(2):136–146 DOI 10.1016/S1470-2045(09)70343-2.

Wang J-Y, Huang J-C, Chen G, Wei D-M. 2018. Expression level and potential target pathways of miR-1-3p in colorectal carcinoma based on 645 cases from 9 microarray datasets. *Molecular Medicine Reports* **17**:5013–5020 DOI 10.3892/mmr.2018.8532.

Xu H, Liu C, Zhang Y, Guo X, Liu Z, Luo Z, Chang Y, Liu S, Sun Z, Wang X. 2014. Let-7b-5p regulates proliferation and apoptosis in multiple myeloma by targeting IGF1R. *Acta Biochimica et Biophysica Sinica* **46**(11):965–972 DOI 10.1093/abbs/gmu089.

Xuan P, Han K, Guo M, Guo Y, Li J, Ding J, Liu Y, Dai Q, Li J, Teng Z, Huang Y. 2013. Prediction of microRNAs associated with human diseases based on weighted k most similar neighbors. *PLOS ONE* **8**(9):e70204 DOI 10.1371/journal.pone.0070204.

Yang Z, Wu L, Wang A, Tang W, Zhao Y, Zhao H, Teschendorff AE. 2017. dbDEMC 2.0: updated database of differentially expressed miRNAs in human cancers. *Nucleic Acids Research* **45**(D1):D812–D818 DOI 10.1093/nar/gkw1079.

Zhang J, Song Y, Zhang C, Zhi X, Fu H, Ma Y, Chen Y, Pan F, Wang K, Ni J, Jin W, He X, Su H, Cui D. 2015. Circulating MiR-16-5p and MiR-19b-3p as two novel potential biomarkers to indicate progression of gastric cancer. *Theranostics* **5**(7):733–745 DOI 10.7150/thno.10305.