miR2Trait: an integrated resource for investigating miRNA-disease associations

Poornima Babu¹, Ashok Palaniappan^{1*}

¹Department of Bioinformatics, School of Chemical and Biotechnology, SASTRA Deemed University, Thanjavur Tamilnadu 613401. India

*Corresponding author: apalania@scbt.sastra.edu

ABSTRACT:

MicroRNAs are key components of cellular regulatory networks, and breakdown in miRNA function could lead to cascading effects culminating in pathophenotypes. A better understanding of the role of miRNAs in diseases would aid 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 catalogues, uncovers new relationships, and includes analytical utilities to interrogate and extract knowledge from these datasets. The resource could aid in identifying the disease enrichment of a user-given set of miRNAs and analyzing the miRNA profile of a specified disease. miR2Trait is available as both a web-server (<https://sas.sastra.edu/pymir18>) and an open-source command-line interface (<https://github.com/miR2Trait>) under MIT license for both commercial and non-commercial use. The datasets are available for download at: <https://doi.org/10.6084/m9.figshare.8288825>.

INTRODUCTION:

MicroRNAs (miRNAs) are key elements of post-transcriptional regulation in the genomic architecture of both prokaryotes and eukaryotes. They are a class of diminutive non-coding RNAs of about 18-25 nucleotides, first observed in *C. elegans* [1]. In miRNA-guided 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 aetiology of complex diseases, especially cardio-vascular, auto-immune and neuro-degenerative diseases, and cancers [2-4]. About half of all miRNA genes are present in cancer-related genetic loci [3], and they are known to play key roles in tumorigenesis and cancer progression [5-7].

The naming of human miRNAs in the miRBase registry (<https://www.mirbase.org>) is related to the miRNA maturation pathway (Figure 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 protein translation [7].

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 [8]. HMDD was developed by Cui et al [9] 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, but is still missing annotation of the -5p or -3p source arm [10]. Since only one arm of the mature miRNA is predominant *in vivo*, this information could be vital in biological research. 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 accurate, 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.3342897>).

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,...).

MATERIALS AND 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 [11]; and (2) DisGeNET, a widely used and standardised knowledge management platform with >1,000,000 gene–disease associations [12]. 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 (Figure 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. In a similar vein, 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 disease associations for a given miRNA is indicative of the breadth of its regulatory impact, which could flag master regulator miRNAs (or regulatory hubs). The disease spectrum width of the i^{th} miRNA is calculated after Qui et al [13] :

$$\text{DSW}(i) = d_i/D_N \quad (1)$$

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 of the j^{th} disease is given by:

$$\text{MSW}(j) = m_j/m_N \quad (2)$$

where m_j is the number of miRNAs associated with the j^{th} disease and M_N is the 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 [14-17]. 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) [18]. The test follows the construction of the Fisher contingency table for each disease, as illustrated in Table 1.

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

The statistical significance (p-value) of enrichment is computed using Python scipy routines (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. 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, and then the miRNA abundance for this *diseasome* is computed in a manner similar to above.

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.

Figure 3. Three layer web-server implementation of miR2Trait accessible at <https://sas.sastra.edu/pymir18>. The miRNA enrichment analysis backend was implemented in Python.

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 (Fig. 3). 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. All data and code used in the work are available for download, either in the Resources section of the web server or in the github repository.

RESULTS

The mapping sizes extracted from the respective databases is given in Table 2. In summary, 2599 miRNAs were mapped to 15062 genes using miRTarbase, and 8819 genes were mapped to 13075 diseases using DisGeNET. Merging the two relations yielded a mapping between 2595 miRNAs and 11689 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 Figure 4.

Figure 4. Density function plots of **A)** miRNA-disease; and **B)** disease-miRNA mappings.

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

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

Database / resource	Mapping relation	Mapping size
miRTarBase	miRNAs→Target genes	2599 → 15062
DisGeNET	Genes → Diseases	8819 → 13075
miR2Trait	miRNAs→ Diseases	2595 → 11689

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	2530
2	hsa-miR-26b-5p	0.20	2321
3	hsa-miR-124-3p	0.19	2175

4	hsa-miR-16-5p	0.18	2110
5	hsa-miR-92a-3p	0.17	1974
6	hsa-miR-1-3p	0.16	1913
7	hsa-miR-17-5p	0.16	1912
8	hsa-let-7b-5p	0.16	1864
9	hsa-miR-155-5p	0.16	1856
10	hsa-miR-93-5p	0.16	1766

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	2491
2	Schizophrenia	0.96	2466
3	Intellectual Disability	0.95	2452
4	Low intelligence	0.94	2421
5	Dull intelligence	0.94	2421
6	Poor school performance	0.94	2421
7	Mental deficiency	0.94	2421
8	Mental Retardation	0.94	2421
9	Mental and motor retardation	0.92	2381
10	Global developmental delay	0.92	2381

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 a query with an adj. p-value (also known as q-value) cutoff of 0.05. This returned a set of 4121 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 [19] could contribute to the emergence of cancer hallmarks [20] and could be tested likewise.

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 significance. 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

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 “Leukaemia”. 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 [21-33]. Databases devoted to miRNA associations in cancer have been developed [34-36], and our work here would augment efforts in this direction. Expression of specific microRNAs has been recognized as vital diagnostic / prognostic biomarker and therapeutic target in a variety of cancers [37-45].

Table 6. Diseasesome analysis of miRNA over-abundance. A summary of cancer-related keyword searches with miR2Trait.

'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

DISCUSSION

Benchmarking

Dedicated miRNA omics databases have been developed to aid researchers unravel the role of miRNAs in biological processes (for e.g, see [46, 47]). Computational methods have also been advanced for inferring miRNA-disease connections [48-50]. 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 aetiology, 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 'normalise' the miRNA nomenclature, by suppressing the

subtype information in miR2Trait. This contracted the number of miRNAs in miR2Trait to 1759. Following this, the databases were compared (Figure 5). Even after contraction, miR2Trait appears more comprehensive in the catalog of miRNA-disease associations. MiRNAs unique to miR2Trait showed a 2.5-fold change relative to the miRNAs unique to HMDDv3.2 (884 vs 334). The unique miRNA complements suggest directions for research. There is also a significant overlap between the two databases.

Figure 5. Benchmarking against HMDDv3.2. The analysis revealed a large overlap but also an unexpected asymmetry in the relative sizes of the unique complements. miR2Trait has a clearly larger unique complement.

Deployment

The miR2Trait web server 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 1445 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 [51]. 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 [52]. 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 [53, 54]. Such searches usually also return miRNAs whose links with the query would not be apparent nor readily ascertained from the literature. Search results for such miRNAs lacking documented disease connections might offer hypotheses for future investigations.

The web-server supports advanced searching with regular expressions (https://dev.mysql.com/doc/refman/8.0/en/regexp.html#operator_regexp). The search term is

case-insensitive, matches anywhere, and interpreted according to regex grammar. The following use-cases will aid the construction of appropriate queries:

- (i) To match at the beginning, the '^' symbol may be used. For e.g, '^essential.*' matches diseases beginning with “essential”.
- (ii) To match at the end, the '\$' symbol may be used. For e.g, “.*tension\$” matches diseases with “tension” at the end.
- (iii) To match multiple options, the '|' symbol may be used. For e.g, '(cancer)|(neoplasm)' fetches all 'cancer' *and* all 'neoplasm' database entries.
- (iv) To match anywhere, with zero or more occurrences of search term, the '*' wildcard may be used. For e.g., '.*tension' retrieves diseases containing “tension” anywhere in their names; here '.' serves to signify any character. The query 'hsa-miR-454*,' when searching by miRNA would match entries that contain “hsa-miR-45” anywhere.
- (v) The special character '+' mandates the occurrence of the preceding character (at least once). 'Neo+' would match diseases containing a prefix word 'Neo' (for e.g, any 'neoplasm'). As another example, when searching by miRNA, 'hsa-miR-454+' would match the following: hsa-miR-4540, hsa-miR-454-3p, hsa-miR-454-5p.
- (vi) A template for constructing queries that retrieve all subtypes of a specific miRNA would be: 'miR-([0-9][0-9]*)[ab]?-[1-9]-' where the parenthesis specifies the miRNA of interest. To retrieve only the 5p variants, we could modify the query thus: 'miR-([0-9][0-9]*)[ab]?-[1-9]-5p'.

CONCLUSION

MicroRNAs play key roles in the development of disease by virtue of their important regulatory activities. To document such 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 have been developed. Taken together, this work constitutes a resource that would provide evidence for suspected associations, generate new hypotheses for experimental pursuit, and facilitate the investigation of the miRNA basis 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 identification of statistically overrepresented 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 potential uses of the resource. The resource has also been benchmarked with the best alternative miRNA-disease database (HMDDv3.2). Further, the

source code for the resource is made available, and includes an additional command-line tool for quantifying miRNA over-abundance in a user-given disease. All the data developed in the project are freely available as supplementary information. The resource could potentially contribute to and enhance our understanding of the role of miRNAs in disease.

ACKNOWLEDGMENTS

We would like to thank the Department of Bioinformatics, School of Chemical & Biotechnology, SASTRA University for infrastructure and computing support. This work was supported in part on DST-SERB grant EMR/2017/000470 to A.P.

REFERENCES

1. Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 1993 75(5):843-54.
2. Min Li, Christian Marin-Muller, Uddalak Bharadwaj, Kwong-Hon Chow, Qizhi Yao, and Changyi Chen. MicroRNAs: Control and Loss of Control in Human Physiology and Disease. *World J Surg* 2009, 33(4): 667–684.doi: 10.1007/s00268-008-9836-x
3. George Adrian Calin, Cinzia Sevignani, Calin Dan Dumitru, Terry Hyslop, Evan Noch, Sai Yendamuri, Masayoshi Shimizu, Sashi Rattan, Florencia Bullrich, Massimo Negrini, and Carlo M. Croce. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *PNAS* 2004, 101 (9) 2999-3004; <https://doi.org/10.1073/pnas.0307323101>.
4. Tai Y-H. MicroRNAs in Human Diseases: From Cancer to Cardiovascular Disease. *Immune Netw* 2011, 11(3):135-54. doi: 10.4110/in.2011.11.3.135.
5. Liu B, Shyr Y, Cai J, Liu Q. Interplay between miRNAs and host genes and their role in cancer. *Brief Funct Genomics*. 2018 Jul 22;18(4):255-266. doi: 10.1093/bfpg/elz002. PMID: 30785618; PMCID: PMC6609535.
6. Hayes J, Peruzzi PP, Lawler S. MicroRNAs in cancer: biomarkers, functions and therapy. *Trends Mol Med*. 2014 Aug;20(8):460-9. doi: 10.1016/j.molmed.2014.06.005. Epub 2014 Jul 12. PMID: 25027972.
7. Leigh-Ann MacFarlane and Paul R. Murphy. MicroRNA: Biogenesis, Function and Role in Cancer. *Curr Genomics* 2010, 11(7): 537–561. doi: 10.2174/138920210793175895
8. Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, Li M, Wang G, Liu Y. miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res*. 2009, 37(Database issue):D98-104. doi: 10.1093/nar/gkn714.
9. Yang Li, Chengxiang Qiu, Jian Tu, Bin Geng, Jichun Yang, Tianzi Jiang, and Qinghua Cui. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res*. 2014, 42(Database issue): D1070–D1074. doi: 10.1093/nar/gkt1023
10. Huang et al. HMDD v3.0: a database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res*. 2019 Jan 8;47(D1):D1013-D1017.
11. 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. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* 2018, 46(D1):D296-D302. doi: 10.1093/nar/gkx1067
12. Piñero J1, Queralt-Rosinach N1, Bravo À1, Deu-Pons J1, Bauer-Mehren A1, Baron M1, Sanz F1, Furlong LI2. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database* 2015, doi: 10.1093/database/bav028.
13. Qiu, C., Chen, G. & Cui, Q. Towards the understanding of microRNA and environmental factor interactions and their relationships to human diseases. *Sci Rep* 2, 318 (2012). <https://doi.org/10.1038/srep00318>
14. Li J, Han X, Wan Y, Zhang S, Zhao Y, Fan R, Cui Q, Zhou Y. TAM 2.0: tool for MicroRNA set analysis. *Nucleic Acids Res.* 2018 Jul 2;46(W1):W180-W185. doi: 10.1093/nar/gky509. PMID: 29878154; PMCID: PMC6031048.
15. Christina Backes, Quratulain T. Khaleeq, Eckart Meese, Andreas Keller, miEAA: microRNA enrichment analysis and annotation, *Nucleic Acids Research*, Volume 44, Issue W1, 8 July 2016, Pages W110–W116, <https://doi.org/10.1093/nar/gkw345>
16. Çorapçıoğlu ME, Oğul H. miSEA: microRNA set enrichment analysis. *Biosystems.* 2015 Aug;134:37-42. doi: 10.1016/j.biosystems.2015.05.004. Epub 2015 Jun 17. PMID: 26093049.
17. Rivals I, Personnaz L, Taing L, Potier MC. Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics* 2007, 23(4):401-407.
18. R. A. Fisher. On the Mathematical Foundations of Theoretical Statistics. *Phil. Trans. R. Soc. Lond. A* 1922, 222, 309-368. doi: 10.1098/rsta.1922.0009
19. Chan, E, Prado, DE, Weidhaas, JB. Cancer microRNAs: From subtype profiling to predictors of response to therapy. *Trends in Molecular Medicine* 2011, 17 (5):235–243. doi:10.1016/j.molmed.2011.01.008
20. Hanahan D & Weinberg RA. Hallmarks of Cancer: The Next Generation. *Cell* 2011, 144: 646 - 674.
21. 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. MicroRNA-335-5p is a potential suppressor of metastasis and invasion in gastric cancer. *Clin Epigenetics* 2017, 9:114. doi: 10.1186/s13148-017-0413-8.
22. K Miyamoto, N Seki, R Matsushita, M Yonemori, H Yoshino, M Nakagawa & H Enokida. Tumour-suppressive miRNA-26a-5p and miR-26b-5p inhibit cell aggressiveness by regulating PLOD2 in bladder cancer. *Br J Cancer* 2016, 115(3): 354–363.
23. Jingpu Zhang, Yang Song, Chunlei Zhang, Xiao Zhi, Hualin Fu, Yue Ma, Yunsheng Chen, Fei Pan, Kan Wang, Jian Ni, Weilin Jin, Xianli He, Haichuan Su, and Daxiang Cui. Circulating MiR-16-5p and MiR-19b-3p as Two Novel Potential Biomarkers to Indicate Progression of Gastric Cancer. *Theranostics* 2015, 5(7): 733–745. doi: 10.7150/thno.10305
24. Jie-Yu Wang, Jia-Cheng Huang, Gang Chen, Dan-Ming Wei. Expression level and potential target pathways of miR-1-3p in colorectal carcinoma based on 645 cases from 9 microarray datasets. *Molecular Medicine Reports* 2018, 5013-5020. <https://doi.org/10.3892/mmr.2018.8532>

25. Nicole Cloonan, Mellissa K Brown, Anita L Steptoe, Shivangi Wani, Wei Ling Chan, Alistair RR Forrest, Gabriel Kolle, Brian Gabrielli and Sean M Grimmond. The miR-17-5p microRNA is a key regulator of the G1/S phase cell cycle transition. *Genome Biology* 2008, 9:R127.
26. Honghai Xu, Cong Liu, Yuelin Zhang, Xiong Guo, Zongzhi Liu, Zhenqun Luo, Yanhai Chang, Shizhang Liu, Zhengming Sun, Xiaoqing Wang. Let-7b-5p regulates proliferation and apoptosis in multiple myeloma by targeting IGF1R. *Acta Biochimica et Biophysica Sinica* 2014, 46 (11): 965–972.
<https://doi.org/10.1093/abbs/gmu089>
27. E. Douglas Robertson, Christine Wasylyk, Tao Ye, Alain C. Jung, Bohdan Wasylyk, The Oncogenic MicroRNA Hsa-miR-155-5p Targets the Transcription Factor ELK3 and Links It to the Hypoxia Response. *PLoS One* 2014, doi: 10.1371/journal.pone.0113050
28. Enrica Fabbri, Giulia Montagner, Nicoletta Bianchi, Alessia Finotti, Monica Borgatti, Ilaria Lampronti, Giulio Cabrini, Roberto Gambari. MicroRNA miR-93-5p regulates expression of IL-8 and VEGF in neuroblastoma SK-N-AS cells. *Oncology Reports* 2016, 2866-2872. <https://doi.org/10.3892/or.2016.4676>
29. Garzon, R.; Garofalo, M.; Martelli, M. P.; 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, M. F.; Mecucci, C.; Croce, C. M.; Falini, B. Distinctive microRNA signature of acute myeloid leukemia bearing cytoplasmic mutated nucleophosmin. *Proceedings of the National Academy of Sciences* 2008, 105(10): 3945–3950. doi:10.1073/pnas.0800135105
30. Marcucci, Guido; Radmacher, Michael D.; Maharry, Kati; Mrózek, Krzysztof; Ruppert, Amy S.; Paschka, Peter; Vukosavljevic, Tamara; Whitman, Susan P.; Baldus, Claudia D.; Langer, Christian; Liu, Chang-Gong; Carroll, Andrew J.; Powell, Bayard L.; Garzon, Ramiro; Croce, Carlo M.; Koltz, Jonathan E.; Caligiuri, Michael A.; Larson, Richard A.; Bloomfield, Clara D. MicroRNA Expression in Cytogenetically Normal Acute Myeloid Leukemia. *New England Journal of Medicine* 2008, 358(18): 1919–1928. doi:10.1056/NEJMoa074256
31. Calin, George Adrian; Ferracin, Manuela; Cimmino, Amelia; Di Leva, Gianpiero; Shimizu, Masayoshi; Wojcik, Sylwia E.; Iorio, Marilena V.; Visone, Rosa; Sever, Nurettin Ilfer; Fabbri, Muller; Iuliano, Rodolfo; Palumbo, Tiziana; Pichiorri, Flavia; Roldo, Claudia; Garzon, Ramiro; Sevignani, Cinzia; Rassenti, Laura; Alder, Hansjuerg; Volinia, Stefano; Liu, Chang-gong; Kipps, Thomas J.; Negrini, Massimo; Croce, Carlo M. A MicroRNA Signature Associated with Prognosis and Progression in Chronic Lymphocytic Leukemia. *New England Journal of Medicine* 2005, 353(17): 1793–1801. doi:10.1056/NEJMoa050995
32. Caramuta, Stefano; Egyházi, Suzanne; Rodolfo, Monica; Witten, Daniela; Hansson, Johan; Larsson, Catharina; Lui, Weng-Onn. MicroRNA Expression Profiles Associated with Mutational Status and Survival in Malignant Melanoma. *Journal of Investigative Dermatology* 2010, 130 (8): 2062–2070. doi:10.1038/jid.2010.63
33. Jima, D. D.; Zhang, J.; Jacobs, C.; Richards, K. L.; Dunphy, C. H. *et al.* Deep sequencing of the small RNA transcriptome of normal and malignant human B-cells identifies hundreds of novel microRNAs. *Blood* 2010, 116(23): e118-e127. doi:10.1182/blood-2010-05-285403.
34. Yang Z, Wu L, Wang A, Tang W, Zhao Y, Zhao H, Teschendorff AE. dbDEMC 2.0: updated database of differentially expressed miRNAs in human cancers. *Nucleic*

- Acids Res. 2017 Jan 4;45(D1):D812-D818. doi: 10.1093/nar/gkw1079. Epub 2016 Nov 28. PMID: 27899556; PMCID: PMC5210560.
35. Sarver AL, Sarver AE, Yuan C, Subramanian S. OMCD: OncomiR Cancer Database. BMC Cancer. 2018 Dec 6;18(1):1223. doi: 10.1186/s12885-018-5085-z. PMID: 30522456; PMCID: PMC6282392.
36. Ahmed M, Nguyen H, Lai T, Kim DR. miRCancerdb: a database for correlation analysis between microRNA and gene expression in cancer. BMC Res Notes. 2018 Feb 7;11(1):103. doi: 10.1186/s13104-018-3160-9. PMID: 29471873; PMCID: PMC5822626.
37. Blenkiron, Cherie; Goldstein, Leonard D; Thorne, Natalie P; Spiteri, Inmaculada; Chin, Suet-Feung; Dunning, Mark J;Barbosa-Morais, Nuno L; Teschendorff, Andrew E; Green, Andrew R; Ellis, Ian O; Tavaré,Simon; Caldas, Carlos;Miska, Eric A. MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. Genome Biology. 2007, 8(10): R214. doi:10.1186/gb2007-8-10-r214
38. Sempere, L. F.; Christensen, M.; Silahatoglu, A.; Bak, M.; Heath, C. V.; Schwartz, G.; Wells, W.; Kauppinen, S.;Cole, C. N. Altered MicroRNA Expression Confined to Specific Epithelial Cell Subpopulations in Breast Cancer. Cancer Research 2007, 67 (24): 11612–11620. doi:10.1158/0008-5472.CAN-07-5019
39. Mattie, Michael D; Benz, Christopher C; Bowers, Jessica; Sensinger, Kelly; Wong, Linda; Scott, Gary K; Fedele, Vita; Ginzinger, David; Getts, Robert; Haqq, Chris. Optimized high-throughput microRNA expression profiling provides novel biomarker assessment of clinical prostate and breast cancer biopsies. Molecular Cancer 2006, 5(1):24. doi:10.1186/1476-4598-5-24
40. Iorio, M. V.; Ferracin, M.; Liu, C.-G.; Veronese, A.; Spizzo, R.; Sabbioni, S.; Magri, E.; Pedriali, M.; Fabbri, M.;Campiglio, M.; Menard, S.; Palazzo, J. P.; Rosenberg, A.; Musiani, P.; Volinia, S.; Nenci, I.; Calin, G. A.; Querzoli, P.;Negrini, M.; Croce, C. M. MicroRNA Gene Expression Deregulation in Human Breast Cancer. Cancer Research 2005, 65 (16): 7065–7070. doi:10.1158/0008-5472.CAN-05-1783
41. Lowery, Aoife J; Miller, Nicola; Devaney, Amanda; McNeill, Roisin E; Davoren, Pamela A; Lemetre, Christophe;Benes, Vladimir; Schmidt, Sabine; Blake, Jonathon; Ball, Graham; Kerin, Michael J. MicroRNA signatures predict oestrogen receptor, progesterone receptor and HER2/neu receptor status in breast cancer. Breast Cancer Research 2009, 11 (3): R27. doi:10.1186/bcr2257
42. 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. Diagnostic Assay Based on hsa-miR-205 Expression Distinguishes Squamous From Nonsquamous Non-Small-Cell Lung Carcinoma. Journal of Clinical Oncology 2009, 27(12): 2030–2037. doi:10.1200/JCO.2008.19.4134
43. Ueda, Tetsuya; Volinia, Stefano; Okumura, Hiroshi; Shimizu, Masayoshi; Taccioli, Cristian; Rossi, Simona; Alder,Hansjuerg; Liu, Chang-gong; Oue, Naohide; Yasui, Wataru; Yoshida, Kazuhiro; Sasaki, Hiroki; Nomura, Sachiyo;Seto, Yasuyuki; Kaminishi, Michio; Calin, George A; Croce, Carlo M. Relation between microRNA expression and progression and prognosis of gastric cancer:a microRNA expression analysis The Lancet Oncology 2010, 11(2): 136–146. doi:10.1016/S1470-2045(09)70343-2
44. Fridman, Eddie; Dotan, Zohar; Barshack, Iris; David, Miriam Ben; Dov, Avital; Tabak, Sarit; Zion, Orit; Benjamin,Sima; Benjamin, Hila; Kuker, Hagit; Avivi,

- Camila; Rosenblatt, Kinneret; Polak-Charcon, Sylvie; Ramon, Jacob; Rosenfeld, Nitzan; Spector, Yael. Accurate Molecular Classification of Renal Tumors Using MicroRNA Expression. *The Journal of Molecular Diagnostics* 2010, 12(5): 687-696. doi:10.2353/jmoldx.2010.090187
45. Gutiérrez, N C; Sarasquete, M E; Misiewicz-Krzeminska, I; Delgado, M; De Las Rivas, J; Ticona, F V; Fermián, E; Martín-Jiménez, P; Chillón, C; Risueño, A; Hernández, J M; García-Sanz, R; González, M; San Miguel, J F. Deregulation of microRNA expression in the different genetic subtypes of multiple myeloma and correlation with gene expression profiling. *Leukemia* 2010, 24(3): 629-637. doi:10.1038/leu.2009.274
46. Ruepp A, Kowarsch A, Schmidl D, Buggenthin F, Brauner B, Dunger I, Fobo G, Frishman G, Montrone C, Theis FJ. PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. *Genome Biol.* 2010 Jan 20;11(1):R6. doi: 10.1186/gb-2010-11-1-r6. PMID: 20089154; PMCID: PMC2847718.
47. Keller A, Leidinger P, Bauer A, Elsharawy A, et al. Toward the blood-borne miRNome of human diseases. *Nat Methods.* 2011 Sep 4;8(10):841-3. doi: 10.1038/nmeth.1682. PMID: 21892151.
48. Chen X, Liu MX, Yan GY. RWRMDA: predicting novel human microRNA-disease associations. *Mol Biosyst.* 2012 Oct;8(10):2792-8. doi: 10.1039/c2mb25180a. PMID: 22875290.
49. Xuan P, Han K, Guo M, Guo Y, Li J, Ding J, Liu Y, Dai Q, Li J, Teng Z, Huang Y. Prediction of microRNAs associated with human diseases based on weighted k most similar neighbors. *PLoS One.* 2013 Aug 8;8(8):e70204. doi: 10.1371/journal.pone.0070204. Erratum in: *PLoS One.* 2013 Sep 23;8(9):null. PMID: 23950912; PMCID: PMC3738541.
50. Qu Y, Zhang H, Lyu C, Liang C. LLCMDA: A Novel Method for Predicting miRNA Gene and Disease Relationship Based on Locality-Constrained Linear Coding. *Front Genet.* 2018 Nov 28;9:576. doi: 10.3389/fgene.2018.00576. PMID: 30555511; PMCID: PMC6282048.
51. Teng Sun, Yan-Han Dong, Wei Du, Chun-Ying Shi, Kun Wang, Muhammad-Akram Tariq, Jian-Xun Wang, and Pei-Feng Li. The Role of MicroRNAs in Myocardial Infarction: From Molecular Mechanism to Clinical Application. *Int J Mol Sci.* 2017, 18(4): 745. doi: 10.3390/ijms18040745.
52. 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. A blood based 12-miRNA signature of Alzheimer disease patients. *Genome Biol.* 2013 Jul 29;14(7):R78. doi: 10.1186/gb-2013-14-7-r78. PMID: 23895045; PMCID: PMC4053778.
53. Kim M, Zhang X. The Profiling and Role of miRNAs in Diabetes Mellitus. *J Diabetes Clin Res.* 2019;1(1):5-23. doi: 10.33696/diabetes.1.003. PMID: 32432227; PMCID: PMC7236805.

54. Guay C, Roggli E, Nesca V, Jacovetti C, Regazzi R. Diabetes mellitus, a microRNA-related disease? Transl Res. 2011 Apr;157(4):253-64. doi: 10.1016/j.trsl.2011.01.009. PMID: 21420036.