1-18-2016

Bioinformatics Resources for MicroRNA Discovery

Alyssa C. Moore

Jonathan S. Winkjer

Tsai-Tien Tseng
Kennesaw State University, ttseng@kenensaw.edu

Bioinformatics Resources for MicroRNA Discovery

Alyssa C. Moore, Jonathan S. Winkjer and Tsai-Tien Tseng

Department of Molecular and Cellular Biology, Kennesaw State University, Kennesaw, GA, USA.

Supplementary Issue: Gene and Protein Expression Profiling in Disease

ABSTRACT: Biomarker identification is often associated with the diagnosis and evaluation of various diseases. Recently, the role of microRNA (miRNA) has been implicated in the development of diseases, particularly cancer. With the advent of next-generation sequencing, the amount of data on miRNA has increased tremendously in the last decade, requiring new bioinformatics approaches for processing and storing new information. New strategies have been developed in mining these sequencing datasets to allow better understanding toward the actions of miRNAs. As a result, many databases have also been established to disseminate these findings. This review focuses on several curated databases of miRNAs and their targets from both predicted and validated sources.

KEYWORDS: bioinformatics, microRNA, database

SUPPLEMENT: Gene and Protein Expression Profiling in Disease

CITATION: Moore et al. Bioinformatics Resources for MicroRNA Discovery. *Biomarker Insights* 2015:10(S4) 53–58 doi: 10.4137/BMI.S29513.

TYPE: Review

RECEIVED: July 21, 2015. **RESUBMITTED:** November 22, 2015. **ACCEPTED FOR PUBLICATION:** November 24, 2015.

ACADEMIC EDITOR: Karen Pulford, Editor in Chief

PEER REVIEW: Five peer reviewers contributed to the peer review report. Reviewers' reports totaled 1,082 words, excluding any confidential comments to the academic editor.

FUNDING: Alyssa C. Moore was supported by the S-STEM scholarship from the National Science Foundation. This material is based upon work supported by the National Science Foundation under Grant No. 1259954. The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

COMPETING INTERESTS: Authors disclose no potential conflicts of interest.

CORRESPONDENCE: ttseng@kennesaw.edu

Introduction

In 1956, Crick stated the central dogma of molecular biology describing the flow of information from DNA to RNA to protein.[1] Although the process of information transmission was oversimplified, the central dogma hinted at the wealth of information that can be extracted from every biological sequence. The mining of information from nucleotide and protein sequences prompted the development of bioinformatics, the science that interfaces biology and computer science to answer biological questions on a molecular level. Sequence-based discovery allows the elucidation of the relationships between structure, function, and evolution. Discovering the relationships between our genetic sequences and the various genetic actions, including the causes of diseases, is one of the main goals of bioinformatics.

The development of biomarker identifications is often associated with the diagnosis and evaluation of various diseases. Many biomarkers are macromolecules of nucleic acids, carbohydrates, and proteins in nature. The initial isolation of nucleic acid-based biomarkers requires the need for genomics as opposed to proteomics, which is needed to isolate protein-based biomarkers. These raw *–omic* outputs are often subjected to further analyses with bioinformatics techniques that focus on particular aspects of the dataset, specifically, in this discussion, biomarkers.[2] Recently, there has been an increased emphasis on the role of microRNA (miRNA) as a biomarker in the diagnosis and possible treatment for cancer.[3,4] miRNAs are small single-stranded noncoding RNAs that control gene expression at the posttranscriptional level. miRNAs act as posttranscription regulators of mRNA by binding to a specific miRNA-binding site on the 3′-untranslated region (3′-UTR) of mRNA.[5] They are often regarded as both predictive and prognostic biomarkers.[6] Sequence-level polymorphisms in miRNA or their target sites can have strong downstream effects in phenotype. These polymorphisms have been implicated in a number of diseases, ranging from cancer, diabetes, Parkinson's, and Alzheimer's. For example, miRNAs have been considered as a serum biomarker for cancer diagnosis and prognosis, particularly in B-cell lymphoma as noted by Lawrie et al.[7]

With the advent of next-generation sequencing (NGS), the identification and quantitation of miRNA as biomarkers are becoming more precise. Many experiments of miRNA quantitation were the results of whole transcriptome sequencing, often referred to as RNA-seq.[8] The hallmark feature of NGS is the ability to elucidate millions of strands of nucleotides simultaneously, which results in an unprecedented amount of coverage for any genome. While NGS is of great interest to many readers, the technical detail is beyond the scope and the allotted space of this review. For users interested in NGS technology, review articles by Mardis,

Mutz et al, and Koboldt et al provide a thorough coverage on its usage and application.[9–12] For readers interested in NGS and classical methods of miRNA discovery, Eminaga et al, Tam et al, and Git et al provide an excellent overview for the processes.[13–15]

NGS is also known as massive parallel sequencing or deep sequencing due to its potential outputs. Consequently, the amount of data generated has also been unprecedented. This requires the establishment of corresponding protocols in processing miRNA data from RNA-seq experiments. For bioinformatics to contribute to the analysis of these RNA-seq datasets, protocols need to be created for finding the most relevant miRNA species. While the main goal of this review is to focus on various repositories of miRNAs and their interactions, it is worthy of note that efforts of computational approaches, such as miRClassify,[16] are also accelerating the overall annotation process of miRNAs. In addition, TargetScan,[17] miRanda,[18] and PicTar[19] are the leading programs in the field, as reflected by the number of citations. For other computational approaches, it is recommended that readers should review articles by Zou et al, Wang et al, and Wei et al.[16,20,21]

As one of the most important goals in bioinformatics, the proper storage and organization of data will lead to easy retrieval and dissemination of information. This review focuses on the specific aspect of databases in miRNA discovery. Several databases are discussed below. The inclusion of databases reviewed here must meet the following criteria: (1) clear documentation of updates and history, (2) recent updates in the past 12 months, and (3) not a simple derivative on data from another database. The major features of each database reviewed here are summarized in Table 1.

miRBase

miRBase (www.mirbase.org) combines the knowledge of miRNA and NGS to create a repository aimed at assigning stable and consistent names to novel miRNAs.[22] While it can be accessed via its web interface, bulk download via file transfer protocol is also available. Established in 2002, miRBase was originally called the miRNA Registry, which allowed submissions of novel miRNAs to be named in a consistent and organized fashion.[23] Its first release contained 218 miRNA loci from five species. As of June 2014, after continuous growth, release 21 contains 28,645 entries representing hairpin precursor miRNAs that expressed 35,828 mature miRNA products in 223 species. miRBase can be used for searching and browsing both hairpin and mature sequences.

Since the inception of miRBase, the annotation strategy was developed and continually improved to organize all the information associated with miRNA species. Its goal was to officialize identifiers as quickly as possible for publication in articles. For example, the prefix in dme-mir-100 designates the organism and is followed by sequentially assigned numbers. Recently, for sequences derived from the 5′ and 3′ arms of the hairpin precursor, names are assigned as dme-miR-100-5p and dme-miR-100-3p, respectively, to specify the mature sequences. This standardized scheme also includes a strategy where homologous miRNA loci are assigned the same number from different species. Two of the most recent developments for miRBase are associated with the advances of NGS technology and community-based contributions toward the textual and functional information on miRNAs.[22] The curators for miRBase attributed the most recent database additions to the next-generation or deep sequencing. This has led to more research groups participating in the process. Similar to many other knowledge bases, the annotation process is also

Table 1. Summarized features from databases reviewed.

DATABASE	MAIN FEATURE	ANNOTATION	DOWNLOAD OPTIONS	INTEGRATED TOOLS, API AND VISUALIZATION	DATA SOURCE
mirBase	Nomenclature assignment	Manual, automated (text mining)	EMBL, fasta, gff3	Stem-loop, deep-sequencing	SRA, GEO, PubMed, community
miRDB	Functional annotation	Automated (machine-learning)	Spreadsheet, flat file	n/a	PubMed, RNA-seq, miRBase
mirWalk	Predicted binding sites	Mutomated (multiple programs)	Search tables	n/a	Refseq 61, miRBase
mirTarBase	miRNA-target interactions	Manual, automated (NLP)	Spreadsheet, flat file	Word cloud, expression profile, structure of pre-miRNA, CytoscapeWeb	TCGA, GEO, CLIP-seq, CLASH-seq, Degradome-seq
mirCancer	miRNA expression	Manual, automated (text mining)	Flat file	n/a	miRBase, PubMed
doRiNA	RNA binding proteins (RBPs), miRNA	Automated pipeline	BED file	UCSC Genome Browser, REST, Python API	GEO, CLIP-seq, selected literature
SomamiR	Somatic and germline mutations	Automated, aided by KEGG	Spreadsheet	n/a	NHGRI GWAS, TCGA
EDRN	Biomarker information	Manual, automated in EDRN Catalog and Archive Service (eCAS)	Flat file	Biomarker Summary Information, BioMuta	Studies from participants

community based in miRBase. Two major sources are involved in the annotation process: publications from PubMed and contribution of textual and functional annotations from the miRNA community. miRBase provides primary references for each miRNA sequence describing its discovery, links to evidence supporting the annotation, coordinates on the genome, and links to databases of predicted and validated target sites. miRBase can be searched with identifiers or keywords along with genomic location. miRNA sequences were also collected and mapped from the Gene Expression Omnibus (GEO) and the Short Read Archive, which are hosted by the National Center for Biotechnology Information (NCBI).

miRDB

While serving as an online resource for functional annotations, miRDB (www.mirdb.org) also functions as a repository for miRNA-target predictions with data downloaded from version 21 of miRBase.[24] Users can also submit their own sequences for prediction at miRDB. As of early 2015, 2.1 million predicted gene targets regulated by 6,709 miRNAs are included in miRDB. The above target prediction was performed with MirTarget.[24] MirTarget was developed by analyzing high-throughput expression profiling data in a support vector machine framework. The MirTarget algorithm also serves as the back-end for the web server interface in prediction. One of the most recent developments was the inclusion of integrated computational analyses with literature, resulting in a new strategy and a scoring system for the identification of functional miRNA with the following four selection criteria. First, PubMed literature mining was utilized to map NCBI gene database for the association of miRNAs with corresponding PubMed records. Second, sequence conservation among different species was considered as functionally important. Third, expression profiles from 81 RNA-seq experiments were used for functional miRNA identification. Fourth, functional annotations by miRBase resulted in the identification of *high confidence* human miRNAs with structural analysis and expression counts. Furthermore, to alleviate falsely identified miRNAs from high-throughput sequencing, the curators of miRDB used a combination of computational analyses and literature mining to identify 568 and 452 functional miRNAs in humans and mice, respectively, for the FuncMir collection in miRDB (http://mirdb.org/miRDB/FuncMir.html).

miRWalk

The third database reviewed is miRWalk (mirwalk.uni-hd.de), which hosts predicted and validated miRNA-binding sites along with information on all known genes of human, mouse, and rat.[25] Similar to miRDB, miRWalk also utilizes automated text mining searches of PubMed to extract information on miRNAs. It is designed as a comprehensive database for predicted and validated targets for miRNAs associated with genes, pathways, diseases, organs, cell lines, and transcription factors.

One of the goals for miRWalk is to use a computational approach to identify the longest consecutive complementary regions between miRNA and gene sequences. The identified miRNA binding sites are generated with the miRWalk algorithm and then combined with the results of many other established prediction programs and databases, including DIANA-microTv4.0,[26,27] DIANA-microT-CDS,[26] miRanda-rel2010,[18] mirBridge,[28] miRDB4.0,[24] miRmap,[29] miRNAMap,[30] doRiNA,[31] PicTar2,[19] RNA22v2,[32] RNAhybrid2.1,[33] and TargetScan6.2.[34] Continual updates and upgrades are the goals for improving miRWalk. Recently, the comparative platform of miRNA-binding sites within the mRNA 3′-UTR region was also upgraded with 13 miRNA-target prediction datasets. All results described above can be found via the web interface of miRWalk 2.0, containing two modules: predicted target module (PTM) and validated target module (VTM). The PTM provides novel comparative platforms of binding sites for the promoter, coding sequence (CDS), and 5′- and 3′-UTR regions. The VTM contains interaction information associated with genes, pathways, organs, diseases, cell lines, Online Mendelian Inheritance in Man (OMIM) disorders, and literature on miRNAs, in addition to information on proteins known to be involved in miRNA processing. The above modules are categorized into different search pages to allow users to retrieve miRNA-associated information using different identifiers.

miRTarBase

The miRTarBase (mirtarbase.mbc.nctu.edu.tw) aimed to provide "the most current and comprehensive information of experimentally validated miRNA-target interactions (MTIs)."[35] For its initial launch of version 1.0 in 2010, the database utilized over 100 published studies. As of September 15, 2015, version 6.0 is the most current iteration of miRTarBase containing 4,966 articles and 3,786 miRNAs. In comparison to databases that provide collections of miRNAs without deeper annotation, the uniqueness of miRTarBase is the curation on MTIs with both manual and computer-aided methods together with a robust suite of tools for the visualization of MTIs and diseases.

In the most recent release, over 360,000 MTIs were collected by manual review after applying natural language processing (NLP) on literature text. In comparison to others, the application of an artificial intelligence approach by the curators of miRTarBase, such as NLP, is a unique feature and should increase the number of relevant articles in the database. Unlike other miRNA databases, miRTarBase contains many robust features of graphical visualization. For instance, the word cloud is a new feature to visualize relationships between individual miRNA and medical conditions. For interactions between miRNAs and their respective targets, Cytoscape Web can be integrated to aid the understanding of miRNA-target regulation.[36] Beyond the usage of Cytoscape Web, the curators also used the Database for Annotation, Visualization and Integrated

Discovery (DAVID) gene annotation tool to perform gene ontology and Kyoto Encyclopedia of Genes and Genome (KEGG) pathway enrichment annotation to further examine the functions of the target genes involved in MTIs.[37–39] These MTIs and associated annotations can be searched by users via the interfaces of the species browser and search utility. The above two interfaces have recently undergone enhancement and redesign. This allows basic MTI searches by miRNA, target gene symbol, validation method, or PubMed ID.

Other than user interface and visualization tools, miRTarBase sets itself apart from similar databases by incorporating datasets from NCBI GEO (www.ncbi.nlm.nih.gov/geo/) and The Cancer Genome Atlas (TCGA) (cancergenome.nih.gov/) to provide miRNA-target gene expression profiles. Specifically, TCGA provides clinical aspects of miRNA and gene expression profiles. Gene expression profiles from the above two data sources are currently considered as a method for experimental validation with the NGS technology. Several specific approaches involving NGS technology are currently being utilized by the curators, including cross-linking and immunoprecipitation (CLIP)-seq,[40] crosslinking, ligation, and sequencing of hybrids (CLASH-seq),[41] and degradome-seq.[42] Overall, miRTarBase contains 21 human CLIP-seq datasets, 5 mouse CLIP-seq datasets, 6 nematode datasets, and 1 human CLASH-seq dataset.

miRCancer

For readers specifically interested in miRNA and cancer, miRCancer (mircancer.ecu.edu) provides a comprehensive collection on the expression of miRNAs via text mining of PubMed.[43] The components for this approach are literature collection, named entity and expression recognition, rule matching, voting, manual verification, and recording. Regular expressions were first used to identify miRNA in literature for miRCancer with miR and miR- for locating miRNA names. Species prefixes, such as hsa- and mmu-, were also used as a part of the regular expressions in searching for related literature. For recognition of cancer names, a cancer name dictionary was compiled from the International Classification of Diseases for Oncology (codes.iarc.fr). The curators also established a dictionary for miRNA expression with 28 terms to include common keywords and phrases for upregulation and downregulation. The text mining approach for miRCancer further relies on 75 rules constructed by the curators using sentence structures commonly found in describing miRNA expressed in cancer cells. These rules are hard-coded sentence structures. Manual revision is then carried out to improve automated extraction. As of March 2015, 44,353 miRNAs for 173 cases of human cancer are associated with 2,073 publications in miRCancer.

doRiNA 2.0

The main goal of doRiNA is to create a single framework for the systematic curation, storage, and integration of RNA-binding proteins (RBPs) and miRNAs from different species.[31,44] It is a database of RNA interactions in posttranscriptional regulation, with predictions carried out by PicTar.[19] Unlike other miRNA databases, doRiNA 2.0 (dorina.mdc-berlin.de) stands out with a strong capability for local implementation, allowing integration into third-party pipelines. Furthermore, doRiNA 2.0 solicits user feedback, can be implemented locally, and operates on an open-source model. As a part of the upgraded version 2.0, the developers also reworked the user interface and expanded the database to improve the usability of the website. It therefore should be considered as one of the most unique and technically sophisticated databases.

Developers of doRiNA 2.0 collected and integrated all available data on miRNA and RBP target sites from the public domain. More than 67 new publicly available RBP datasets have been added into doRiNA 2.0. In the latest version of doRiNA, miRNA and their targets were identified with both computational predictions and new experimental techniques by chimeric sequencing reads. Due to the lack of reliable *in silico* predictions of RBP target sites, the curators have decided to focus on high-resolution, transcriptome-wide CLIP experiments. All candidate miRNA target sites are still subject to probabilistic scoring by a hidden Markov model. Data from various cell lines of human, mouse, roundworm, and fly are available in Browser Extensible Data (BED) formats, allowing integration of coordinates and annotation tracks with the UCSC Genome Browser.[45]

Recent updates in version 2.0 provide various improvements from the previous version. Developers of doRiNA paid special attention toward the infrastructure and interoperability surrounding their repository. doRiNA 2.0 can now achieve high query speed and complexity by precomputing several important data characteristics. External developers can easily integrate doRiNA 2.0 into third-party analysis piplines via a representational state transfer application program interface (API), while the Python API can be used for local queries by users. Documentations for the above two APIs can be found at http://dorina.mdc-berlin.de/docs. The developers have also migrated away from the traditional Common Gateway Interface (CGI) and Structured Query Language (MySQL) implementations and instead used a fast key-value cache and store (redis.io) as well as in-memory caching of frequent queries for faster access. Mirrored sites and database servers are utilized by doRiNA 2.0 to achieve high service availability. Both the web application and the APIs are available under an open-source license approved by the Open Source Interconnection that permits research and commercial access and reuse. The developers at doRiNA essentially created an *ecosystem* that provides a user-friendly environment while encouraging external developers to adapt this miRNA repository.

SomamiR

SomamiR (compbio.uthsc.edu/SomamiR/) was created to integrate heterogeneous datasets to investigate the impact of somatic and germline mutations on miRNA function in

cancer.[46] It specifically contains experimentally determined germline and somatic miRNA mutations associated with cancer, along with their target sites. A total of 15 sources of somatic mutations that have been identified from whole-genome sequencing of paired normal and cancer samples were analyzed and incorporated into SomamiR.

Three methods were used to predict how mutations may impact target sites in SomamiR. First, a comprehensive list of how somatic mutations may alter miRNA-binding sites was created with methods established by Ellwanger et al.[47] Second, two popular miRNA-target prediction algorithms, TargetScan[17] and PITA,[48] were used to determine mutations that are more likely to alter functional binding sites. Third, five major types of information were used to annotate miRNAs, genes, and target locations in SomamiR: results of association studies, gene pathways, sequence conservation, expression of miRNAs in cancer, and germline mutations. For association studies, high scoring markers from genome-wide association studies (GWAS) of cancer in National Human Genome Research Institute (NHGRI) GWAS catalog were collected. The data on meta-analysis of cancer candidate gene association studies from the Cancer GAMAdb[49] were also collected. Developers also carried out functional annotation of genes containing somatic mutations that alter miRNA target sites with the KEGG. They further highlighted genes with somatic mutations from miRNA target sites in each pathway. To improve miRNA-target prediction, the conservation of a target site sequence across species has been used. A 46-way multiZ[50] alignment of vertebrate genomes was utilized to determine whether the sequence of a predicted target site was conserved. To better understand somatic cell mutations associated with cancer, miRNA expression data from various cancer genome sequencing projects deposited at TCGA were also collected. In addition to somatic cell mutations, germline mutations that alter predicted and experimental miRNA target sites were collected from PolymiRTS.[51] The name PolymiRTS derives from polymorphisms in miRNAs and their target sites. PolymiRTS is a database for tracking and identifying sequence polymorphisms in miRNAs or their target sites to possibly reveal links to molecular, physiological, and behavioral disease phenotypes.

In SomamiR, each gene is represented by a single web page to provide all somatic mutations that alter miRNA target sites in the gene, as well as associate with specific types of cancer. Each web page representing a gene can also be accessed through several browsable tables that are linked from the database homepage. These browsable tables contain somatic mutations in miRNAs and respective target sites. Furthermore, experimental evidence linking these mutations to various cancer types is also incorporated into these tables. Two additional tables can be used to browse database entries in the context of association studies and KEGG gene pathways. SomamiR also allows the following criteria for searching against the database: miRNA, gene symbol, RefSeq ID, and chromosome location. The search can be performed using the form on the website or by uploading a batch file with multiple terms. For users who are interested in parsing the database for further analysis, the complete content of SomamiR is also available for download at http://compbio.uthsc.edu/SomamiR/download/.

Early Detection Research Network (EDRN)

While the above-described databases are exclusively for the discovery and understanding of miRNAs, other repositories can contain similar information from various types of biomarkers. One such effort in categorizing data related to multiple types of biomarkers is EDRN from the National Cancer Institute[52] (edrn.nci.nih.gov). While EDRN is not exclusively designated as a sequence-level repository, biomarker data, including miRNAs, can be found under the section of *informatics*. Several tools are under the *informatics* section with a link to *Biomarker Database* being the most relevant to this review. The database is further divided into five sections: biomarkers, studies, publications, terms/glossary, and sites. Some data can be visualized using BioMuta, a curated single nucleotide variation and disease association database where the variations are mapped to the genome/protein/gene. Finally, EDRN is also aiming at establishing biomarker bioinformatics standards and ontology for the community.

Conclusion

While data repositories were the main focus for this review, miRNA-target prediction also presents other interesting questions in bioinformatics. It is more challenging to predict miRNA targets in animals than in plants, due to imperfect base pairing with target sites. This demonstrates the potential limitation for any prediction algorithms due to the complexity of many biological systems. There will be a strong need for further improvements to develop accurate predictions for miRNA targets. In addition to the goal of predicting miRNA targets, the selected miRNA databases reviewed above share the commonality of relying on textual information, mostly from PubMed, in the retrieval of relevant literature. It is also important for readers to note that one of the most important efforts is the standardization of biomarker nomenclature, including various miRNAs by EDRN. Standardization will improve the interoperability among different research groups and databases. Furthermore, nearly all curators for the above repositories recognized that major growth of data will result from sequencing. With the advent of new technologies, there is no doubt that more miRNAs will be discovered, resulting in an exciting new era for researchers.

Acknowledgment

We thank Ashley Pedicini for her assistance in the preparation of this article.

Author Contributions

Wrote the first draft of the manuscript: ACM, JSW, T-TT. Contributed to the writing of the manuscript: ACM, JSW,

T-TT. Jointly developed the structure and arguments for the article: ACM, JSW, T-TT. Made the critical revisions and approved the final version: ACM, JSW, T-TT. All authors reviewed and approved the final article.

REFERENCES

1. Crick F. Central dogma of molecular biology. *Nature*. 1970;227(5258):561–3.
2. Azuaje F, Devaux Y, Wagner D. Computational biology for cardiovascular biomarker discovery. *Brief Bioinform*. 2009;10(4):367–77. doi: 10.1093/bib/bbp008.
3. Jeffrey SS. Cancer biomarker profiling with microRNAs. *Nat Biotechnol*. 2008; 26(4):400–1. doi: 10.1038/nbt0408-400.
4. Melo SA, Esteller M. Dysregulation of microRNAs in cancer: playing with fire. *FEBS Lett*. 2011;585(13):2087–99. doi: 10.1016/j.febslet.2010.08.009.
5. Doench JG, Sharp PA. Specificity of microRNA target selection in translational repression. *Genes Dev*. 2004;18(5):504–11. doi: 10.1101/gad.1184404.
6. Kosaka N, Iguchi H, Ochiya T. Circulating microRNA in body fluid: a new potential biomarker for cancer diagnosis and prognosis. *Cancer Sci*. 2010;101(10): 2087–92. doi: 10.1111/j.1349-7006.2010.01650.x.
7. Lawrie CH, Chi J, Taylor S, et al. Expression of microRNAs in diffuse large B cell lymphoma is associated with immunophenotype, survival and transformation from follicular lymphoma. *J Cell Mol Med*. 2009;13(7):1248–60. doi: 10.1111/j.1582-4934.2008.00628.x.
8. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10(1):57–63. doi: 10.1038/nrg2484.
9. Mardis ER. Next-generation sequencing platforms. *Annu Rev Anal Chem*. 2013;6(1):287–303. doi: 10.1146/annurev-anchem-062012-092628.
10. Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends Genet*. 2008;24(3):133–41. doi: 10.1016/j.tig.2007.12.007.
11. Mutz K-O, Heilkenbrinker A, Lönne M, et al. Transcriptome analysis using next-generation sequencing. *Curr Opin Biotechnol*. 2013;24(1):22–30. doi: 10.1016/j.copbio.2012.09.004.
12. Koboldt DC, Steinberg KM, Larson DE, et al. The next-generation sequencing revolution and its impact on genomics. *Cell*. 2013;155(1):27–38. doi: 10.1016/j.cell.2013.09.006.
13. Eminaga S, Christodoulou DC, Vigneault F, et al. Quantification of microRNA expression with next-generation sequencing. *Curr Protoc Mol Biol*. 2013;4:4.17.
14. Tam S, de Borja R, Tsao M-S, et al. Robust global microRNA expression profiling using next-generation sequencing technologies. *Lab Invest*. 2014;94(3):350–8.
15. Git A, Dvinge H, Salmon-Divon M, et al. Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. *RNA*. 2010;16(5):991–1006.
16. Zou Q, Mao Y, Hu L, et al. miRClassify: an advanced web server for miRNA family classification and annotation. *Comput Biol Med*. 2014;45:157–60. doi: 10.1016/j.compbiomed.2013.12.007.
17. Agarwal V, Bell GW, Nam J-W, et al. Predicting effective microRNA target sites in mammalian mRNAs. *eLife*. 2015;4:e05005. doi: 10.7554/eLife.05005.
18. Enright AJ, John B, Gaul U, et al. MicroRNA targets in *Drosophila*. *Genome Biol*. 2003;5(1):R1. doi: 10.1186/gb-2003-5-1-r1.
19. Krek A, Grün D, Poy MN, et al. Combinatorial microRNA target predictions. *Nat Genet*. 2005;37(5):495–500. doi: 10.1038/ng1536.
20. Wang C, Wei L, Guo M, et al. Computational approaches in detecting non-coding RNA. *Curr Genomics*. 2013;14(6):371–7. doi: 10.2174/13892029113149990005.
21. Wei L, Liao M, Gao Y, et al. Improved and promising identification of human MicroRNAs by incorporating a high-quality negative set. *IEEE/ACM Trans Comput Biol Bioinform*. 2014 Jan–Feb;11(1):192–201. doi: 10.1109/TCBB.2013.146.
22. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*. 2014;42(D1):D68–73. doi: 10.1093/nar/gkt1181.
23. Griffiths-Jones S. The microRNA registry. *Nucleic Acids Res*. 2004;32(suppl 1): D109–11. doi: 10.1093/nar/gkh023.
24. Wong N, Wang X. miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res*. 2015;43(D1):D146–52. doi: 10.1093/nar/gku1104.
25. Dweep H, Gretz N. miRWalk2.0: a comprehensive atlas of microRNA-target interactions. *Nat Methods*. 2015;12(8):697–7. doi: 10.1038/nmeth.3485.
26. Paraskevopoulou MD, Georgakilas G, Kostoulas N, et al. DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Res*. 2013;41(Web Server issue):W169–73. doi: 10.1093/nar/gkt393.
27. Reczko M, Maragkakis M, Alexiou P, et al. Functional microRNA targets in protein coding sequences. *Bioinforma Oxf Engl*. 2012;28(6):771–6. doi: 10.1093/bioinformatics/bts043.
28. Tsang JS, Ebert MS, van Oudenaarden A. Genome-wide dissection of microRNA functions and co-targeting networks using gene-set signatures. *Mol Cell*. 2010;38(1):140–53. doi: 10.1016/j.molcel.2010.03.007.
29. Vejnar CE, Zdobnov EM. MiRmap: comprehensive prediction of microRNA target repression strength. *Nucleic Acids Res*. 2012;40(22):11673–83. doi: 10.1093/nar/gks901.
30. Hsu PWC, Huang H-D, Hsu S-D, et al. miRNAMap: genomic maps of microRNA genes and their target genes in mammalian genomes. *Nucleic Acids Res*. 2006;34(suppl 1):D135–9. doi: 10.1093/nar/gkj135.
31. Anders G, Mackowiak SD, Jens M, et al. doRiNA: a database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res*. 2012;40(D1):D180–6. doi: 10.1093/nar/gkr1007.
32. Miranda KC, Huynh T, Tay Y, et al. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell*. 2006;126(6):1203–17. doi: 10.1016/j.cell.2006.07.031.
33. Krüger J, Rehmsmeier M. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res*. 2006;34(Web Server issue):W451–4. doi: 10.1093/nar/gkl243.
34. 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. doi: 10.1016/j.cell.2004.12.035.
35. Hsu S-D, Tseng Y-T, Shrestha S, et al. miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Res*. 2014;42(Database issue):D78–85. doi: 10.1093/nar/gkt1266.
36. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504. doi: 10.1101/gr.1239303.
37. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2008;4(1):44–57. doi: 10.1038/nprot.2008.211.
38. Kanehisa M, Sato Y, Kawashima M, et al. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2015:gkv1070. doi: 10.1093/nar/gkv1070. http://nar.oxfordjournals.org/papfaq
39. Consortium GO. The gene ontology (GO) database and informatics resource. *Nucleic Acids Res*. 2004;32(suppl 1):D258–61. doi: 10.1093/nar/gkh036.
40. König J, McGlincy NJ, Ule J. Analysis of protein–RNA interactions with single-nucleotide resolution using iCLIP and next-generation sequencing. In: Harbers M, Kahl G, eds. *Tag-Based Next Generation Sequencing*. Wiley-VCH Verlag GmbH & Co. KGaA; 2011:153–69. http://onlinelibrary.wiley.com/doi/10.1002/9783527644582.ch10/summary.
41. Helwak A, Kudla G, Dudnakova T, et al. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*. 2013;153(3): 654–65. doi: 10.1016/j.cell.2013.03.043.
42. German MA, Pillay M, Jeong D-H, et al. Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat Biotechnol*. 2008; 26(8):941–6. doi: 10.1038/nbt1417.
43. Xie B, Ding Q, Han H, et al. miRCancer: a microRNA–cancer association database constructed by text mining on literature. *Bioinformatics*. 2013;29(5):638–44. doi: 10.1093/bioinformatics/btt014.
44. Blin K, Dieterich C, Wurmus R, et al. doRiNA 2.0–upgrading the doRiNA database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res*. 2015;43(Database issue):D160–7. doi: 10.1093/nar/gku1180.
45. Rosenbloom KR, Armstrong J, Barber GP, et al. The UCSC genome browser database: 2015 update. *Nucleic Acids Res*. 2015;43(Database issue):D670–81. doi: 10.1093/nar/gku1177.
46. Bhattacharya A, Ziebarth JD, Cui Y. SomamiR: a database for somatic mutations impacting microRNA function in cancer. *Nucleic Acids Res*. 2013;41(D1): D977–82. doi: 10.1093/nar/gks1138.
47. Ellwanger DC, Büttner FA, Mewes H-W, et al. The sufficient minimal set of miRNA seed types. *Bioinformatics*. 2011;27(10):1346–50. doi: 10.1093/bioinformatics/btr149.
48. Kertesz M, Iovino N, Unnerstall U, et al. The role of site accessibility in microRNA target recognition. *Nat Genet*. 2007;39(10):1278–84. doi: 10.1038/ng2135.
49. Schully SD, Yu W, McCallum V, et al. Cancer GAMAdb: database of cancer genetic associations from meta-analyses and genome-wide association studies. *Eur J Hum Genet EJHG*. 2011;19(8):928–30. doi: 10.1038/ejhg.2011.53.
50. Blanchette M, Kent WJ, Riemer C, et al. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res*. 2004;14(4):708–15. doi: 10.1101/gr.1933104.
51. Bhattacharya A, Ziebarth JD, Cui Y. PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Res*. 2014;42(D1):D86–91. doi: 10.1093/nar/gkt1028.
52. Reynolds T. Validating biomarkers: early detection research network launches first phase III study. *J Natl Cancer Inst*. 2003;95(6):422–3. doi: 10.1093/jnci/95.6.422.