Briefings in Bioinformatics, 2017, 1–25

doi: 10.1093/bib/bbx130
Paper

MicroRNAs and complex diseases: from experimental results to computational models

Xing Chen, Di Xie, Qi Zhao and Zhu-Hong You

Corresponding author: Xing Chen, School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China. Tel.: 86-0516-83590815; Fax: 86-0516-83590815; E-mail: xingchen@amss.ac.cn

Abstract

Plenty of microRNAs (miRNAs) were discovered at a rapid pace in plants, green algae, viruses and animals. As one of the most important components in the cell, miRNAs play a growing important role in various essential and important biological processes. For the recent few decades, amounts of experimental methods and computational models have been designed and implemented to identify novel miRNA–disease associations. In this review, the functions of miRNAs, miRNA–target interactions, miRNA–disease associations and some important publicly available miRNA-related databases were discussed in detail. Specially, considering the important fact that an increasing number of miRNA–disease associations have been experimentally confirmed, we selected five important miRNA-related human diseases and five crucial disease-related miRNAs and provided corresponding introductions. Identifying disease-related miRNAs has become an important goal of biomedical research, which will accelerate the understanding of disease pathogenesis at the molecular level and molecular tools design for disease diagnosis, treatment and prevention. Computational models have become an important means for novel miRNA–disease association identification, which could select the most promising miRNA–disease pairs for experimental validation and significantly reduce the time and cost of the biological experiments. Here, we reviewed 20 state-of-the-art computational models of predicting miRNA–disease associations from different perspectives. Finally, we summarized four important factors for the difficulties of predicting potential disease-related miRNAs, the framework of constructing powerful computational models to predict potential miRNA–disease associations including five feasible and important research schemas, and future directions for further development of computational models.

Key words: microRNA; complex disease; microRNA–disease association prediction; computational model; machine learning; biological network

Article structure

In this article, we first introduced the microRNA (miRNA), which contains the miRNA function, miRNA–target interactions and miRNA–disease associations as three subsections of 'MicroRNA'. Second, we introduced databases that have various data related to miRNAs and different functions for researchers. Third, we introduced different types of computational models for miRNA–disease association prediction developed in the recent years based on various data resources and algorithm theories. Finally, we provided the discussion and conclusion for all the contents in this review.

MicroRNA

There are numerous genes in human body. They work together to guarantee each function goes well [1]. Through translating

Xing Chen, PhD, is a professor of School of Information and Control Engineering, China University of Mining and Technology. He is also the Founding Director of Institute of Bioinformatics, China University of Mining and Technology. His research interests include disease and noncoding RNAs, network pharmacology, complex network, and machine learning.
Di Xie is a full-time master student of School of Mathematics, Liaoning University. His research interests include disease and microRNAs, machine learning and applied mathematics.
Qi Zhao, PhD, is an associate professor of School of Mathematics, Liaoning University. His research interests include disease and noncoding RNAs, complex network and machine learning.
Zhu-Hong You, PhD, is a professor of Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Science. His research interests include disease and noncoding RNAs, network pharmacology and machine learning.
Submitted: 6 July 2017; **Received (in revised form):** 13 August 2017

into proteins holding diverse functions, genes create a well-organized, smoothly running human body through various complicated procedures [1]. However, these protein-coding genes only take up an extremely minority of the human genome (approximately 1.5%) [2–11]. Noncoding RNAs (ncRNAs), which were falsely regarded as transcriptional noise and intermediary from gene to protein in the past [11–13], have been proved by accumulating evidence to have regulatory roles in various biological processes [11, 14]. One classic example is that Taft *et al.* found the increased complexity of organisms was in proportion to the increased percentage of ncRNAs [15]. Furthermore, based on whether the length of transcript is >200 nucleotides (nt), ncRNAs can be divided into two groups: small ncRNAs and long ncRNAs [5, 11, 16–30].

MiRNA is a class of single-stranded, endogenous, small, evolutionarily conserved ncRNAs belonging to the former group (21-24 nt), normally negatively regulating gene expression by sequence-specific base pairing with their target messenger RNAs (mRNAs) [31–33]. In the year 1993, Victor Ambros *et al.* discovered the first miRNA, lin-4, that has complementarity to binding sites of lin-28 [32]. Soon, the second miRNAs, let-7, was identified as well. Coincidently, these two miRNAs happened to be the first two known miRNAs acting as positive regulators, found out by the method of forward genetic screens [32, 34, 35]. As the first two known miRNAs were discovered, the door of research on miRNAs was opened. Many miRNAs were discovered in plants, green algae, viruses and animals [36]. Owing to the increasing crucial roles of miRNAs, more attention has been poured into miRNA-related research. According to the latest version of miRBase, 28 645 entries have been identified and 2588 mature miRNAs have been found in human (miRBase, Release 21) [36, 37]. However, further studies discovered that the majority of miRNAs play their roles negatively. For instance, miR-17–92 cluster is proved to be oncogenic, and further study showed it was associated with malignant lymphoma [38, 39]. Furthermore, experiments indicate miRNAs have high conservation across species and some of them are even linage specific.

MiRNA function

As one of the most important components in the cell, miRNAs play a growing important role in various essential and important biological processes, such as cell development, proliferation differentiation, apoptosis, signal transduction, viral infection and so on [40–47]. Some important functions of miRNA were summarized as follows [48]. For example, the first discovered miRNA lin-4 plays a critical regulatory role in the nematode larvae development by regulating the expression of its target genes lin-14 and lin-28 [34].

Experimental techniques summary

With the rapid development of research on the miRNA function, the experimental methods for exploring it have also experienced different stages of the development process [49]. Methods to induce a loss of function of miRNAs represent a powerful functional genomics tool [49]. Traditional screening for mutation of miRNA genes has not been proved successful, which is because miRNAs are often encoded by multigene family members and the loss of function of one miRNA member is often obscured by redundant functions of other miRNA members that have almost identical sequences and the ability to bind and regulate the same target gene transcripts [49]. Fortunately, we can choose to inhibit the regulatory functions of each miRNA family that we want to silence with the developed various methods and strategies. For example, the vitro chemically modified miRNA inhibitors such as antagomirs and the vivo target mimicry for miRNA sponge of mammalian cells have been recently proved to have effect for the inhibitive functions of specific miRNA families [50, 51]. In the recent decades, Castoldi *et al.* proposed the miChip, which is an array-based method for miRNA expression profiling using locked nucleic acid capture probes in 2008 [52]. Furthermore, Cui *et al.* used graphene oxide-protected DNA probes for multiplex miRNA analysis in complex biological samples based on a cyclic enzymatic amplification method in 2012 [53]. Additionally, though new techniques grow rapidly, deep sequencing is still a popular concept, which aims for high number of replicate reads of each region of a sequence. Despite the ability of technologies to sequence and to detect miRNAs with the previous high-throughput, deep sequencing presents formidable computational challenges and suffers from the errors caused when we prepare the small RNA libraries. Even mapping deep sequencing reads to the genome is not trivial, as no animal genome besides that of *Caenorhabditis elegans* has been sequenced completely. Moreover, sequencing errors and polymorphisms, as well as RNA editing and splicing, are just reasons of the ambiguity of the deep sequencing results. Although currently almost all of these problems remain mostly unsolved, deep sequencing can successfully survey the small RNA contents of animal genomes with unmatched sensitivity [54]. In fact, deep sequencing is still popular for current studies such as the study of Koshizuka *et al.* in 2017, which researched the deep sequencing-based miRNA expression signatures in head and neck squamous cell carcinoma (HNSCC) [55]. Meanwhile, functional miRNA screening methods were often implemented in the research for miRNA function in the recent period of time [56].

Function on nervous system

Recent studies demonstrated that miRNAs are associated with nervous system. MiRNAs are responsible for the development of early embryonic stem cell survival and differentiation. Furthermore, it was exciting to know that certain miRNAs also help maintaining the survival of mature neurons and carrying on their function [48]. Therefore, the dysfunction of miRNAs may cause various nervous system diseases [48]. For example, in the midbrain of patients with Parkinson's disease, the deficiency of miR-133b could be easily observed, which is regarded as a regulator to the maturation and function of midbrain dopaminergic neurons [48].

Function on cell differentiation and development

Evidence also suggested that miRNAs play significant roles in cell differentiation and development either *in vitro* or *in vivo*. For instance, overexpression of MiR-128 in glioma cells was proved to inhibit cell proliferation [48, 57]. Meanwhile, a conserved target site was identified within the 3′ untranslated region (UTR) of E2F3a, which could regulate the process of cell cycle progression [48].

Function on viral infection

When viruses attack human body, host cells use miRNAs to target viral functions and defend against RNA and DNA viruses [48, 58, 59]. The first identified anti-viral miRNA successfully controlled the increase of the retrovirus primate foamy virus type 1 in somatic cell [48, 60]. However, in return, viruses may take advantage of miRNAs to control host cell. For example, the SV40-encoded miRNA miR-S1 would help viruses hide infected

cell from immune system [48]. Therefore, the regulatory function could be useful for both viruses and host cells.

Function on immunity

MiRNAs also have close relationship with mammalian immune system. Genetic ablation, loss or deregulation of certain miRNAs will probably affect immune development, even worse, causing immune disorders [48].

MiRNA–target interactions

By base pairing between the seed region of a miRNA and the binding sites on its target mRNAs at 3′-UTR [61–65], miRNA complements with these binding sites based on complete or partial sequence complementarity [66]. The former complementarity leads to transcript degradation, while the latter one causes translational repression [66]. In addition, further studies showed, the number of binding sites in certain regions of the targets may have some relation with the degree of translational repression that target genes experience in the process of protein transcription. For example, experiment showed the influence acted by miRNA witnessed an exponential exacerbation when the number of biding sites increased [67–69].

From previous studies, some important conjectures have been obtained as follows: (1) miRNAs may consist of 1–4% genes in the human genomes [70]; (2) a miRNA could regulate as many as 200 mRNAs [70]; (3) one-third of human genes could be targeted by miRNAs [71]. Therefore, it could be concluded that a miRNA could target multiple mRNAs and a single mRNA could also be targeted by multiple miRNAs, which constitutes a complicated miRNA–target interactions network. This important posttranscriptional regulatory network could play critical roles in various biological processes.

Traditionally, scientists recognized miRNA–target interactions by means of genetic analyses. However, this method spent unnecessary and redundant time for experiments, and false results could happen as well. Therefore, computational models have been developed to predict candidate regulatory targets [72]. Core protocols for predicting regulatory sites can be concluded as the following three steps: (1) Recognize conserved Watson–Crick pairing to the 5′ region within miRNA concentrated on nucleotides 2–7, namely, miRNA ‘seed’ [72]. (2) Use existing whole-genome alignments to compile orthologous 3′-UTRs to eliminate the noise of false-positive predictions [72]. (3) Search in the range of the orthologous UTRs for conserved occurrence of either 7 nt match [72]. These three main rules could be used to develop corresponding computational models to help predict candidate targets. The algorithms can be divided into four groups. The first kind of models used the principle of base sequence complementarity to predict targets [73]. The tools based on this principle are miRanda [74], TargetScan [75] and PicTar. Second, thermal stability of the mRNA–miRNA duplex can be used either to be a crucial recognition method or to be a secondary filtering step for extracting the more likely targets [73]. In the tools of RNAhybrid [76] and DIANA micro-T [77], thermal stability plays the former role, while in miRanda [74], PicTar [78] and MicroInspector [79], it acts as a secondary filtering instead. Third, target site conservation can benefit the identification of mRNA–miRNA interactions [73]. In the same way as thermal stability, this method is used to reduce the proportion of false positive, even though sometimes it may increase the percentage of false negatives [73]. Also, the miRNA and gene profile analysis can be applied into explaining and understanding miRNA–target interaction. The score for miRNA and potential targets can be calculated based on multiple conditions such as location of target sites [73].

There are several databases that collect experimentally confirmed miRNA–target interactions or provided predicted miRNA–target interactions with high accuracy. MirWalk, a comprehensive database, provides predicted and validated information on miRNAs from human mouse and rat and their target genes (http://www.ma.uni-heidelberg.de/apps/zmf/mirwalk/documentation.html) [73]. Also, the tool miRDB created an online database for miRNA target prediction as well as functional annotations in human, mouse, rat, dog and chicken. More details can be found here: http://mirdb.org/miRDB/ [73]. StarBase, as a web server, provides predicted miRNA target for both animals and plant organisms (http://starbase.sysu.edu.cn/index.php) [73]. Other relevant databases include miRNAMap (http://miRNAMap.mbc.nctu.edu.tw/), TarBase (http://diana.imis.athena-innovation.gr/DianaTools), microRNA.org (http://www.microrna.org), miRBase (http://www.mirbase.org/) and so on.

Experimental techniques summary

In fact, experimental methods for investigating the relationships between miRNAs and their related targets are still at the initial stage. The first validations of miRNA–target interactions were reported in plants [80, 81]. Two experimental approaches were established for these validations: direct visualization of the target cleavage *in vitro* and cloning of the target cleavage products *in vivo* by 3′ or 5′ rapid-amplification of cDNA ends by Polymerase Chain Reaction (RACE-PCR) [82]. Traditional wheat germ extracts or newly established maize germ extracts are convenient cell-free systems for miRNA target validation. Both systems contain abundant endogenous miRNAs that have already been loaded on the RNA-induced silencing complexes (RISCs), and exogenous target mRNAs are readily cleaved by the existing miRNA-associated RISCs. However, these cell-free systems have limitations when validating miRNA–target interactions whose corresponding miRNAs do not exist in wheat or maize germ extracts. *In vitro* programming of active RISCs by synthetic miRNA duplexes in plant systems needs further exploration [81, 83]. In contrast, Drosophila embryo extracts are capable of RISC assembly programmed by various kinds of small RNAs and can thus serve as a platform for animal miRNA target validations, as well as a useful heterologous system for validation of plant miRNA targets [84]. Schmittgen *et al.* proposed a high-throughput method to monitor the expression of miRNA precursors [85]. Furthermore, Miranda *et al.* proposed a pattern-based method for the identification of miRNA binding sites and their corresponding heteroduplexes [86]. Meanwhile, Neely *et al.* proposed a single-molecule method for the quantitation of miRNA gene expression [87]. For the recent period, Aw *et al.* proposed a conformation-induced fluorescence method for overcoming the disadvantage of small size of miRNA when tagging or direct detecting the miRNA–target interactions [88]. Moreover, Zhang *et al.* used a two-stages duplex-specific nuclease (DSN)-assisted target recycling signal amplification method for the sensitive detection of miRNA in complex biological samples [89]. All the mentioned methods play important roles in the studies of miRNA–target interactions.

MiRNA–disease associations

One of the most fundamental and significant goal in biomedical research is to understand the molecular, physiological and pathological mechanisms underlying complex human diseases. Recently, the deeper understanding of functions of miRNA has

been revealed. Therefore, studies on the disease mechanisms have been extended from genes to miRNAs [90]. Accumulating evidence has demonstrated that almost each miRNA interacts with hundreds of targets, and plays a role as 'oncogenes' or 'tumor suppressor' gene in tumorigenesis, metastasis, proliferation and differentiation of certain cancer cells [91]. Therefore, miRNAs have become emerging cancer biomarkers, helping add an auxiliary to predict and analyze several kinds of cancers [91]. Deciphering the roles of miRNAs in human complex diseases would benefit not only the understanding of the molecular mechanism of these diseases, but also biomarker detections for human disease development, progression, prognosis, diagnosis and treatment responses evaluation [91].

Increasing evidence indicates miRNAs have close associations with various human complex diseases, such as various cancers, diabetes, Alzheimer's disease, acquired immune deficiency syndrome and cardiac hypertrophy [70, 92–97]. Especially, accumulated studies have shown that miRNAs can function as oncogenes or tumor suppressors in the initiation, progression and metastasis of various types of cancers [98], including breast cancer [95], prostate cancer [99], lung cancer [100], colon cancer [101], ovarian cancer [102]. The same miRNA may have different effects on the same disease. For example, miR-137 showed both cell multiplication by targeting cell division cycle 42 (Cdc42), cyclin-dependent kinases 6 (Cdk6) in lung cancer cells and down-regulation by DNA methylation [103].

In the Human MiRNA-Disease Database [104], miRNA–disease associations were divided into four types according to the different supporting evidence, including miRNA–disease associations from the evidence of genetics, epigenetics, circulating miRNAs and miRNA–target interactions, respectively [37]. The same miRNA may also be associated with different diseases based on the same association type or the same disease based on different association types. Especially, deregulations of miRNAs can result in dysfunction of downstream mRNA [105], and further caused a wide spectrum of human diseases [92, 96], including glioblastoma [106, 107], coronary artery disease [108] and type 2 diabetes [98]. For example, mir-206 could slow down the progression of amyotrophic lateral sclerosis by sensing motor neuron injury [109]. Furthermore, mir-375 can regulate insulin secretion [110]; the miR-1 is involved in heart development; deletion of miRNA-1-2 interrupts the regulation of carcinogenesis [111, 112]. Moreover, miRNAs and environmental factors (EFs) could jointly affect human diseases [113–115].

Experimental techniques summary

Application of various miRNA array platforms reveals numerous miRNA biomarkers for a variety of human diseases including various kinds of cancers. The expression changes of these miRNA biomarkers indicate changes from normal to abnormal genetic and physiological conditions [116]. For example, a specific spectrum of miRNAs including miR-23, miR-24, miR-26, miR-27, miR-103, miR-107, miR-181, miR-210 and miR-213 was induced in neoplastic cells under a hypoxic environment compared with normal conditions [117]. Similarly, the method to compare the differences between miRNA profiles of tumor and normal tissues have revealed distinct miRNA biomarkers for various tumor cells over the past few years [118]. Exploration of these miRNA biomarkers will be particularly useful in early diagnostics of human diseases such as cancers, diabetes and Alzheimer's disease. Eventually, detailed miRNA atlases of humans, animals and plants will be available with the help of small RNA deep sequencing and miRNA array platforms for a wide variety of applications. As the recent development of experimental technique, Gottwein *et al.* used the method of releasing viral miRNA orthologs into infected cells to found the viral miRNA functions as an ortholog of cellular miR-155 in 2007 [119]. Furthermore, Cameron *et al.* used the method of changing the expression of miRNAs to found that Epstein–Barr virus latent membrane protein 1 induces cellular miRNA miR-146a, which is a modulator of lymphocyte signaling pathways [120]. Meanwhile, Su *et al.* proposed an *in vivo* method to identify miRNA targets not predicted by computation algorithms, which was the p21 targeting by miR-92a in cancer [121]. Additionally, Markou *et al.* took advantage of the approach of quantitative real-time reverse transcription-PCR (RT-PCR) to research the prognostic value of mature miRNA-21 and miRNA-205 overexpression in non-small-cell lung cancer (NSCLC) [122]. For the recent period, Bakhidze *et al.* carried out the analysis of expression of miRNA in cytological smears as a new method for the diagnosis and prognosis of preinvasive cervical carcinoma, which have shown promise for clinical value [123]. Moreover, Huang *et al.* proposed an improved method to quantitate mature plant miRNA in biological matrices using modified periodate treatment and inclusion of internal controls [124].

The following five miRNA-related diseases and five disease-related miRNAs are introduced as specific examples to illustrate the relationships between miRNAs and diseases.

Breast cancer

Breast cancer is one of the most commonly occurring female cancers in the world [91]. If the breast cancer could be detected and treated early, the better prognosis would be possible for the patient [91]. However, owing to its heterogeneous histopathological patterns and complex clinical behaviors, early detection and treatment is still a challenge [91]. To diagnose, manage and treat breast cancer, the molecular mechanisms of it should be fully exploited [91]. Increasing evidence showed that there are close association between breast cancer and several miRNAs. For example, in the breast cancer cell lines and tissue specimens sampled from chemotherapy-sensitive or -resistant patients, low expression of miR-195 could be observed easily [91]. Furthermore, miR-195 could reduce breast tumor cell survival and increase apoptosis through downregulation of Raf-1, Bcl-2 and P-glycoprotein expression [91]. Also, it will help Adriamycin treatment through increasing the sensitivity of breast cancer cells [91]. Therefore, the occurrence of breast cancer may be detected and prevented in advance by using related miRNAs as biomarkers after the association between miRNAs and breast cancer has been thoroughly understood.

Colorectal cancer

Colorectal cancer (CRC) is the third most common deadly cancer in the world after breast and lung cancer [125–127]. Unlike breast cancer, in which miRNAs regulate the cell biological processes like cell proliferation, death and apoptosis, some target genes associated with the metastatic process have been shown to be dysregulated by miRNAs in CRC [128]. In the experiment carried on by Vickers *et al.* [129], the differential expressions of several metastasis-associated miRNAs were observed. According to the result, the trend of expression level of miR-135a and miR-335 has been shown to be proportional to the progression of CRC, while miR-206 has an opposite trend [129]. Furthermore, a significant increase of let-7a expression has been witnessed in metastatic CRC [129].

Lung cancer

Lung cancer remains the commonest cause of cancer-related death with a high incidence every year. Despite of newly developed drugs and therapeutic approaches, the mortality of lung cancer is still high owing to late presentation, poor prognosis, low treatment rates and the high proportion of patient who cannot undergo curative resection. Increasing research shows that specific miRNAs can be considered as biomarkers for lung cancers. For example, the expression level of miR-29s was found to be inversely correlated to DNA methyltransferases 3A (DNMT3A) and DNA methyltransferases 3B (DNMT3B) in lung cancer tissues by controlling the reexpression of methylation-silenced tumor suppressor genes and inhibiting tumorigenicity [130]. In addition, it is reported that miR-145 could inhibit the proliferation of transfected lung adenocarcinoma cells with a significantly downregulated mRNA expression of epidermal growth factor receptor (EGFR) and Nudix hydrolase 1 (NUDT1) [131]. Furthermore, the overexpression of miR-138 was confirmed to have inhibiting effect on the cell proliferation of human NSCLC by binding to the 3′-UTR of enhancer of zeste homolog 2 (EZH2) oncogene and suppressing the expression level of EZH2 [132].

Prostate cancer

Prostate cancer has been witnessed to have an exceptionally fast increase of incidence all over the world for the past decade, which reaches about 3% per year [133]. Great progress has been made in clinical treatment of prostate cancer, including the approval of several new and effective drugs to improve the survival of advanced prostate cancer patients [134]. Approximately 30% of prostate cancer patients will relapse and require taxane-based chemotherapy, which can only provide a moderate survival benefit of 1–2 months. The roles of miRNAs in regulating common signaling pathways in prostate cancer carcinogenesis are extensively studied as drug targets and biomarkers [135]. For example, the promoter of miR-21 was found to bind to androgen receptors and upregulate the expression of miR-21, which consequently boosted the castration-resistant growth of prostate cancer cell line [136]. In addition, the partial methylation of the promoters of miR-29a and miR-1256 could cause their epigenetic deregulation, which contributes to the inhibition of prostate cancer cell growth and invasion [137]. Furthermore, miR-34b was reported to be silenced in human prostate cancer through CpG hypermethylation, which directly targeted methyltransferases and deacetylases [138].

Kidney cancer

Kidney cancer is one of the most common causes leading to death, which accounts for ∼2% of human malignancies with >250 000 new cases diagnosed and >40% mortality every year [139]. Owing to the refractory nature of kidney cancer, patients usually suffer from high rates of metastatic recurrences [∼30% of localized renal cell carcinoma (RCC) cases] and death (5-year survival rate of 60–70%) [139, 140]. Radical/partial nephrectomy, the initial treatment for kidney cancer, remains the main therapeutic approach for kidney cancer patients. Emerging targeted cancer therapies, including protein von HippelLindau-hypoxia-inducible factor (pVHL-HIF)-targeted cancer therapy, vascular endothelial growth factor (VEGF)-targeted cancer therapy and mammalian target of rapamycin (mTOR) signaling-targeted cancer therapy, have excited extensive interest and attention to identify novel molecular biomarkers (i.e. miRNAs) for early diagnosis, risk assessment and therapeutic intervention of kidney cancer [139, 141]. An increasing number of miRNAs have been reported to be involved in the development and progression of kidney cancer. For example, Wulfken *et al.* used an array technology (TaqMan Low Density Array) to identify miR-1233 as a potential biomarker for RCC and confirmed its distinctly high serum levels in RCC patients by using quantitative real-time polymerase chain reaction (PCR) [142]. Similarly, circulating miR-378 and miR-451 in serum were also identified and validated as potential biomarkers for RCC [143].

Hsa-let-7a

The contribution of miRNAs to various human pathological conditions including cancers has been well established with the accumulating reports and investigations. An extensive number of miRNAs, such as let-7 miRNA family and miRNA hsa-mir-1-1, have been well studied and found to be associated with a wide range of human diseases including various cancers, asthma, human immunodeficiency virus (HIV), obesity and so on [144–146]. As a good example, has-let-7a was reported to induce diseases with aberrant expression. For example, miRNA let-7a has lower expression levels in the blood of NSCLC patients and the cells of NSCLC tissues than the normal controls [147]. It was also observed that miRNA let-7a was underexpressed in colorectal adenocarcinomas cases [148]. Furthermore, miRNA let-7a was found to be an important regulator of integrin $\beta 3$ and therefore the loss of let-7a expression is involved in the development of malignant melanoma [149].

Hsa-let-7b

Accumulating research shows that miRNA let-7b serves as significant targets for the epigenetic machinery in various diseases including liver cancer, breast cancer, lung cancer and malignant melanoma and so on [59, 149–151]. It was reported that the transcriptional regulation of miRNA let-7b could suppress high-mobility group AT-hook 2 (HMGA2) expression via RNA interference pathways serving as a tumor suppressor [151]. Mouth cancer was also reported to be associated with the downregulation of let-7b [152]. Specifically, analysis in HNSCC cell lines by real-time PCR showed that the down-regulated let-7b was associated with elevated expression levels of Dicer, an RNase III endonuclease required for miRNA maturation, which could increase the proliferation of oral cancer cells [153]. In addition, the correlation between let-7b miRNA expression and cataracts was also confirmed by the previous research [154]. Let-7b played as an important regulator in cellular aging and tissue senescence and its unregulated expression level could make lens opacity more serious [154].

Hsa-mir-1

The associations between miRNA mir-1 and various complex human diseases have been also reported by recent research and investigations [155–157]. For example, when the DNA hypermethylation of miR-1-1 was first observed in hepatocellular carcinogenesis (HCC) cells and primary HCC, the ectopic expression of DNA-hypermethylated miR-1-1 was identified to inhibit the growth of tumor cells by targeting mesenchymal to epithelial transition factor (MET), forkhead box prote1 (FoxP1) and histone deacetylase 4 (HDAC4) [158]. The epigenetic silencing of miR-1 was also demonstrated in human prostate cancer following 5-Aza-dC treatment [159]. Furthermore, previous research also reported the frequent methylation of miR-1-1 in CRC and considered miR-1-1 functions as a tumor suppressor by controlling MET expression [160, 161].

Hsa-mir-25

MiRNA mir-25 has been also considered as an effective biomarker for different diseases [162–164]. Specially, mir-25 could boost the process of thyroid cancer progression and further contributes to the development of anaplastic carcinomas by

regulating the expression of the polycomb protein EZH2 [162]. The involvement of mir-25 in cholangiocarcinoma was also identified [163]. Specifically, the miR-25 level was reported to be positively related to the evasion of Tumor Necrosis Factor (TNF)-related apoptosis-inducing ligand-induced cholangiocarcinoma apoptosis by targeting Death Receptor 4 (DR4) [163]. Furthermore, the association between mir-25 and human colon cancer was identified by bioinformatics predictions and experimental validation carried out by Li *et al.* [164]. The down-regulated mir-25 could suppress the cell proliferation and migration in colon cancer by repressing Smad7, while the overexpression of mir-25 inhibited the xenografts of colon cancer cells *in vivo* [164].

Hsa-mir-499

MiRNA mir-499 has been found to regulate gene expression and promote or suppress the growth of cancer cells via posttranscriptional mechanisms [165–167]. For example, the overexpression of miR-499 was shown to be associated with the downregulation of ets1 mRNA and the inhibition of invasion and migration of HepG2 cells, which demonstrated that the miR-499 was able to repress the expression of the ets1 proto-oncogene getting involved in the pathogenesis of HCC [165]. It was also reported that miR-499 could serve as a biomarker of acute myocardial infarction [166]. For example, the plasma concentration of miR-499 is significantly increased in cases of acute myocardial infarction [166]. Furthermore, the ectopic expression of miR-499b could induce cell cycle arrests and the downregulations of CDK6, Cell Division Cycle 25A (CDC25A) and cell cycle protein A (CyclinA) in ovarian cancer cell lines SKOV3-ipl, which boosted the development of ovarian cancer [167].

Databases

To date, there are different kinds of databases that have been built to store various data related to miRNAs, including databases collecting comprehensive information of miRNAs such as miRBase [168, 169] and IntmiR [170], databases collecting miRNA-related interaction such as miREnvironment [114] and miRTarBase [171] and databases collecting miRNA–disease associations such as HMDD [172] and miR2Disease [173] (See Table 1).

Databases collecting comprehensive information of miRNAs

miRBase

(http://www.mirbase.org/) [168, 169]

The current version of miRBase (miRBase 21) contains over 28 645 entries, which represent hairpin precursor miRNAs. All these miRNAs express 35 828 mature miRNA products in 223 species. As a searchable database of published miRNA sequences and annotation, miRBase provides the users the information of target information, primary evidence and known miRNA sequences for recorded miRNAs. The entries in miRBase Sequence database provide the information of the location and mature miRNA sequence through the functions of searching and browsing. Users can also retrieve the entries with name, keyword, references and annotation. miRBase also provides a consistent naming system for miRNAs, i.e. the miRBase Registry.

miRGator

(http://mirgator.kobic.re.kr/) [174, 175]

The latest release of this database is miRGator v3.0, which collects 73 deep sequencing data sets form GEO, SRA and TCGA archives covering 4.1 billion short reads and 2.5 billion aligned reads. The data in miRGator are curated into 38 diseases and 71 anatomic categories. This database also compiles miRNA–mRNA target interactions and inverse correlation between gene expression and the miRNA–mRNA relationships according to three databases of validated targets and six databases of predicted targets. miRGator also provides the analysis for the coexpression between miRNAs and their target mRNAs, which is represented in two formats of correlation heat map and network views. The major feature in the latest update is the inclusion of deep sequencing data, which allows the users to examine short read alignment with secondary structure and read count information.

miRGen

(http://www.microrna.gr/mirgen/) [176, 177]

MiRGen aims at studying the associations between miRNA function and miRNA genomic organization, and offering useful tools for studies in miRNA genomic organization, co-transcription and targeting. Based on miRGen database, the users can acquire the information about positional relationships between miRNAs and genomic annotation sets as well as the potential miRNA targets predicted by popular prediction programs. Specifically, the whole-genome collections of miRNAs can be explored with the associated University of California Santa Cruz (UCSC) genome browser annotation sets (e.g. Known Genes, Refseq Genes, CpG islands) in the Genomics interface. The associations between miRNAs and confirmed target genes from TarBase as well as predicted target genes are presented in the Targets interface.

IntmiR. (http://rgcb.res.in/intmir/) [170]

IntmiR is a manually curated database that focuses on intronic miRNAs of human and mouse genome and provides the users with comprehensive miRNA-associated information including target genes, pathways and diseases. There are 426 intronic miRNA loci from human and 76 from mouse recorded in this database, expressing distinct target mRNA sequences. The entries in IntmiR can be retrieved based on the miRNA ID or target gene.

Databases collecting miRNA-related interactions

miReg

(http://iioab-mireg.webs.com/) [178]

MiReg is a small database that manually curates the regulatory relationships among miRNA-associated elements including validated upstream regulators (i.e. transcription factors, drugs, physical and chemical), downstream targets, associated biological process, experimental condition and disease state. This database records 47 human miRNAs, which contain 295 experimentally validated relationships of miRNAs with their upstream regulators and downstream targets collected from 190 PubMed references. Users can explore the database by five browsing modes (i.e. miRs, upstream regulators, drugs/other modulators, biological processes and diseases).

miRTarBase

(http://miRTarBase.mbc.nctu.edu.tw/) [171]

Accumulating evidence shows that miRNAs play crucial important roles in disease pathology by regulating key oncogenes, tumor suppressors or protein expression and subsequently controlling cell mechanisms. Therefore, the interactions between miRNAs and target genes are garnering increasing interest from researchers. miRTarBase is a manually curated database, which collects 366 181 miRNA–target interactions covering 3786 miRNAs and 22 563 target genes from 18

Table 1. Comparison list of various databases

Database	Function	URL
Databases collecting comprehensive information of miRNAs		
miRBase	Provide the users the information of target information, primary evidence and known miRNA sequences for recorded miRNA	http://www.mirbase.org/
miRGator	Compile miRNA–mRNA target interactions and inverse correlation between gene expression and the miRNA–mRNA relationships	http://mirgator.kobic.re.kr/
miRGen	Study the associations between miRNA function and miRNA genomic organization, and offering useful tools for studies in miRNA genomic organization, co-transcription and targeting	http://www.microrna.gr/mirgen/
IntmiR	Focus on intronic miRNAs of human and mouse genome and provide the users with comprehensive miRNA-associated information including target genes, pathways and diseases	http://rgcb.res.in/intmir/
Databases collecting miRNA-related interactions		
miReg	Curate the regulatory relationships among miRNA-associated elements including validated upstream regulators, downstream targets, associated biological process, experimental condition and disease state	http://iioab-mireg.webs.com/
miRTarBase	Collect 366 181 miRNA–target interactions covering 3786 miRNAs and 22 563 target genes from 18 species	http://miRTarBase.mbc.nctu.edu.tw/
miRecords	Provide the users with predicted miRNA targets produced by 11 previously proposed miRNA target prediction algorithms	http://c1.accurascience.com/miRecords/
miRWalk	Store the largest amount of predicted and experimentally verified miRNA–target interactions	http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/
TarBase	Indexed more than half a million experimentally validated miRNA–gene interactions covering 356 distinct cell types from 24 species	http://www.microrna.gr/tarbase
miRNAMap	Collect experimentally verified interactions between miRNAs and miRNA target, provide three established computational tools for identifying miRNA targets in 3′-UTR of genes	http://mirnamap.mbc.nctu.edu.tw/
microRNA.org	Provide miRNA expression profiles derived from tissues and cell lines	http://www.microrna.org
miRPathDB	Provide researchers easy access to the information about target pathways regulated by miRNAs	https://mpd.bioinf.uni-sb.de/
PEMDAM	Construct three bipartite association networks: EF–miRNA network, EF–disease network and miRNA–disease network	http://lmmd.ecust.edu.cn/database/pemdam/
MiREnvironment	Users can also perform bioinformatics analysis to predict cancer treatment as well as EF–disease associations through this web resource	http://210.73.221.6/miren
SomamiR 2.0	A better platform for functional analysis of somatic mutations altering miRNA–ceRNA interactions	http://compbio.uthsc.edu/SomamiR
Databases collecting miRNA–disease associations		
MiRCancer	Provide 878 miRNA–cancer associations, which includes 236 miRNAs and 79 human cancers	http://mircancer.ecu.edu/
MiR2Disease	Provide a data resource of miRNA deregulation in various human diseases	http://www.mir2disease.org/
HMDD	Detailed and comprehensive annotations for miRNA–disease associations are presented in each entity	http://www.cuilab.cn/hmdd
MiREC	Focus on the miRNA–disease associations that are specific for endometrial cancer	http://www.mirecdb.org
DbDEMC	Record the miRNA expression information in 14 kinds of cancers	http://www.picb.ac.cn/dbDEMC/
OncomiRDB	Store information on experimentally validated oncogenic and tumor-suppressive miRNAs	http://bioinfo.au.tsinghua.edu.cn/oncomirdb/
OncomiRdbB	Collect the associations between miRNAs and breast cancer as well as the target genes	http://tdb.ccmb.res.in/OncomiRdbB/index.htm

species. The data in miRTarBase were collected from 4966 articles involving in reporter assay, western blot, microarray and next-generation sequencing experiments.

miRecords

(http://c1.accurascience.com/miRecords/) [179]

MiRecords is an integrated database, which includes 1135 experimentally validated miRNA–target interactions between 301 miRNAs and 902 target genes from seven animal species. In addition, this database also provides the users with predicted miRNA targets produced by 11 previously proposed miRNA target prediction algorithms.

miRWalk

(http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/) [180, 181]

The latest release of this database is miRWalk2.0, which is another integrated database storing the largest amount of predicted and experimentally verified miRNA–target interactions. Specifically, it hosts miRNA binding site interactions on 18 394 gene ontology terms and miRNA–target interactions on 2035 disease ontologies, 6727 human phenotype ontologies and 4980 Online Mendelian Inheritance in Man (OMIM) disorders. The miRWalk2.0 not only documents miRNA binding sites within the complete sequence of a gene, but also combines this information with a comparison of binding sites resulting from 12

existing miRNA–target prediction programs to build novel comparative platforms of binding sites for the promoter, cds, 5′-UTR and 3′-UTR regions. It also documents experimentally verified miRNA–target interaction information collected via an automated text-mining search and data from existing resources.

TarBase
(http://www.microrna.gr/tarbase) [182, 183]

TarBase is a comprehensive database of animal miRNA targets, whose current version indexes more than half a million experimentally validated miRNA–gene interactions covering 356 distinct cell types from 24 species. This database records the experimental results with the corresponding information including utilized experimental methodology, experimental conditions and treatment. The advanced information on the binding site location and the primer sequences used for cloning experiments is also provided in this database.

miRNAMap
(http://mirnamap.mbc.nctu.edu.tw/) [184]

MiRNAMap is miRNA–target interaction database that focuses on metazoan genomes including target genes in two insects, nine vertebrates and one worm. The current version (miRNAMap2.0) not only collects experimentally verified interactions between miRNAs and miRNA target genes but also provides three established computational tools (i.e. miRanda, RNAhybrid and TargetScan) for identifying miRNA targets in 3′-UTR of genes. In addition, miRNAs expression profiles provided by microRNAMap database can offer valuable information for users to investigate the properties of miRNA, such as tissue specificity and aberrant expression in cancer cell.

microRNA.org
(http://www.microrna.org) [185]

MicRNA.org is a comprehensive database of miRNA–target interactions and expression profiles of miRNAs. Specifically, this database utilizes the miRanda prediction algorithm to identify miRNA targets based on the up-to-date compendium of mammalian miRNAs, and uses mirSVR algorithm to compute target downregulation scores. MicRNA.org resource also provides miRNA expression profiles derived from tissues and cell lines. Specially, the latest version of micRNA.org released in August 2010 records 1100 human miRNAs.

miRPathDB
(https://mpd.bioinf.uni-sb.de/) [186]

miRPathDB was constructed to complement available web servers by providing researchers easy access to the information about target pathways regulated by miRNAs. The database contains 2595 human miRNAs, different miRNA target sets (14 773 experimentally validated target genes as well as 19 281 predicted targets genes) and a broad selection of functional biochemical categories (KEGG, WikiPathways, BioCarta, SMPDB, PID, Reactome pathways, functional categories from gene ontology, protein families from Pfam and chromosomal locations totaling 12 875 categories). In addition to *Homo sapiens*, *Mus musculus* data are also stored and can be compared with human target pathways.

PEMDAM
(http://lmmd.ecust.edu.cn/database/pemdam/) [187]

Besides miRNA–gene interactions, the relationship between miRNAs and EFs is another kind of valuable information for miRNA-associated studies because EFs can act as connections between miRNAs and various diseases. PEMDAM database provides this information by constructing three bipartite association networks: EF-miRNA network, EF-disease network and miRNA–disease network, which may facilitate the understanding on toxicology mechanisms and disease etiologies.

MiREnvironment
(http://210.73.221.6/miren) [114]

MiREnvironment is another database that describes the complex interactions between EFs and miRNAs. The latest version of miREnvironment database contains 3857 entries covering 1242 miRNAs, 394 EFs, 304 phenotypes and 24 species from 557 publications. Users can also perform bioinformatics analysis to predict cancer treatment as well as EF–disease associations through this web resource.

SomamiR 2.0
(http://compbio.uthsc.edu/SomamiR) [188]

SomamiR 2.0 is a database of cancer somatic mutations in miRNAs and their target sites that potentially alter the interactions between miRNAs and competing endogenous RNAs (ceRNA) including mRNAs, circular RNAs (circRNA) and long noncoding RNAs (lncRNA). SomamiR 2.0 expanded the scope of the database by including somatic mutations that impact the interactions between miRNAs and two classes of ncRNAs, circRNAs and lncRNAs. SomamiR 2.0 has mapped 388 247 somatic mutations to the experimentally identified miRNA target sites. It also includes a list of somatic mutations in the miRNA seed regions, which contain the most important guiding information for miRNA target recognition. Data and functions from multiple sources including biological pathways and genome-wide association studies were updated and integrated with SomamiR 2.0 to make it a better platform for functional analysis of somatic mutations altering miRNA–ceRNA interactions.

Databases collecting miRNA–disease associations

MiRCancer
(http://mircancer.ecu.edu/) [189]

MiRCancer is a miRNA–cancer association database, which was constructed through text mining on PubMed literature and manual confirmation. There are 878 miRNA–cancer associations documented in miRCancer database, which includes 236 miRNAs and 79 human cancers. These empirical records would offer valuable information for researchers interested in miRNA biomarkers.

MiR2Disease
(http://www.mir2disease.org/) [173]

MiR2Disease is a manually curated database that provides a data resource of miRNA deregulation in various human diseases. The current version of miR2Disease contains 3273 entries covering 349 miRNAs and 163 diseases. Each entry records detailed information including miRNA expression pattern, detection method and literature reference.

HMDD
(http://www.cuilab.cn/hmdd) [172]

HMDD is a database that collects experimentally supported evidence for miRNA–diseases associations. The current release (HMDDv2.0) has recorded 10 368 entries including 572 miRNAs and 378 human diseases from 3511 articles. Detailed and comprehensive annotations for miRNA–disease associations are presented in each entity including the data from evidence of

genetics, epigenetics, circulating miRNAs and miRNA–target interactions.

MiREC
(http://www.mirecdb.org) [190]

MiREC is a database focusing on the miRNA–disease associations that are specific for endometrial cancer. This database contains 228 miRNAs and 920 target genes. MiREC database would boost the studies regarding the role of miRNAs in pathogenic mechanism of endometrial cancer.

DbDEMC
(http://www.picb.ac.cn/dbDEMC/) [191]

DbDEMC is database that records the miRNA expression information in 14 kinds of cancers and all the information was collected from 48 microarray data sets. There are 607 miRNAs including 590 mature miRNAs and 17 precursor miRNAs recorded in dbDEMC database. Users can browse miRNA annotations through specific cancer types.

OncomiRDB
(http://bioinfo.au.tsinghua.edu.cn/oncomirdb/) [192]

OncomiRDB database stores information on experimentally validated oncogenic and tumor-suppressive miRNAs. This database contains 2259 entries including >300 miRNAs and 829 target genes in 25 different cancer tissues by exploring ∼9000 articles. Users can search and browse the miRNA annotations through graphical and verbal interfaces provided by OncomiRDB.

OncomiRdbB
(http://tdb.ccmb.res.in/OncomiRdbB/index.htm.) [193]

OncomiRdbB database collects the associations between miRNAs and breast cancer as well as the target genes. This database was created based on different existing databases and lately validated experiment results. The numbers of human and mouse breast cancer-related miRNAs in OncomiRdbB database reach 782 and 246, which are significantly larger than other established databases. This database is expected to facilitate the knowledge regarding the roles of miRNAs in breast cancer.

Computational models

The close involvement of mutations and deregulations of miRNAs in pathological conditions through various complex mechanisms has been indicated by an increasing number of research evidence, which gives rise to increasing attention on identification of miRNA biomarkers as well as miRNA–disease associations. As a good complement to experiment-based methods, which are expensive and time-consuming, computational prediction models could effectively search the most potential candidates for further validation experiments, and therefore decrease the time and money for miRNA–disease association identification. In this section, we introduced the frameworks of the state-of-the-art computational models to predict novel miRNA–disease associations.

Specifically, these prediction models can be divided into four categories, namely, score function-based, complex network algorithm-based, machine learning-based and multiple biological information-based models. Score function-based models adopt probability distribution or statistics analysis on the miRNA- and disease-related training data to construct score functions for prioritizing potential miRNA–disease associations. Complex network algorithm-based models are mainly based on various miRNA similarity networks and disease similarity networks from different perspectives. Machine learning-based prediction models aim to utilizing powerful machine learning algorithms to make reliable predictions by extracting effective features or solving specific optimization problems [194]. Multiple biological information-based models consider diverse types of miRNA-related and disease-related associations like miRNA–gene and disease–protein associations, and try to construct associations between miRNAs and diseases through all these intermediate medium associations.

Score function-based models

Prioritizing algorithm for miRNA similarity based on cumulative hypergeometric distribution
Jiang *et al.* have proposed the first computational approach to search the most potential miRNA candidates involved in diseases of interest [195]. This model is mainly based on a scoring system, which assumes that functional-related miRNAs are more likely to be associated with phenotypically similar diseases (Figure 1). Specifically, this approach contains two major steps. First, the disease phenotype similarity scores downloaded from MimMiner are used to build disease phenotype network *DN* and the functionally related miRNA network *MN* is constructed based on the co-regulated target gene. In the second step, a scoring approach is implemented on the constructed disease and miRNA networks, aiming at searching the most similar miRNAs to those known to be associated with diseases of interest. Given a disease d and its similar disease set D, the scores of miRNAs in *MN* are computed simply based on the cumulative hypergeometric distribution:

$$score_k = 1 - \sum_{i=m}^{M} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

Here, N denotes the total number of miRNAs in *MN* and M denotes number of miRNAs associated with the diseases in D; n denotes the neighbor number of miRNAs in *MN* and m denotes the number of the neighbors of miRNAs in *MN*, which have links with the members of D, respectively. As a result, this model achieved the Area Under Receiver Operating Characteristic (ROC) Curve (AUC) value of 0.7580 on the benchmark data set by implementing the leave-one-out cross-validation (LOOCV). However, the information source for miRNA network suffers from high rates of false-positive and false-negative results because most of the miRNA–target interactions have not been experimentally supported, which would influence the performance of this model to some extent.

WBSMDA
Chen *et al.* have developed the model of Within and Between Score for MiRNA–Disease Association prediction (WBSMDA), which integrates miRNA functional similarity, disease semantic similarity and Gaussian interaction profile kernel similarity of diseases and miRNAs [196]. Specifically, the Gaussian interaction profile kernel similarity is integrated with disease semantic similarity and miRNA functional similarity by replacing them when they are absent. The integrated disease similarities allow WBSMDA to be applied to the new disease without any associated miRNA by investigating its similar diseases with known associated miRNAs. To measure the extent to which a miRNA is

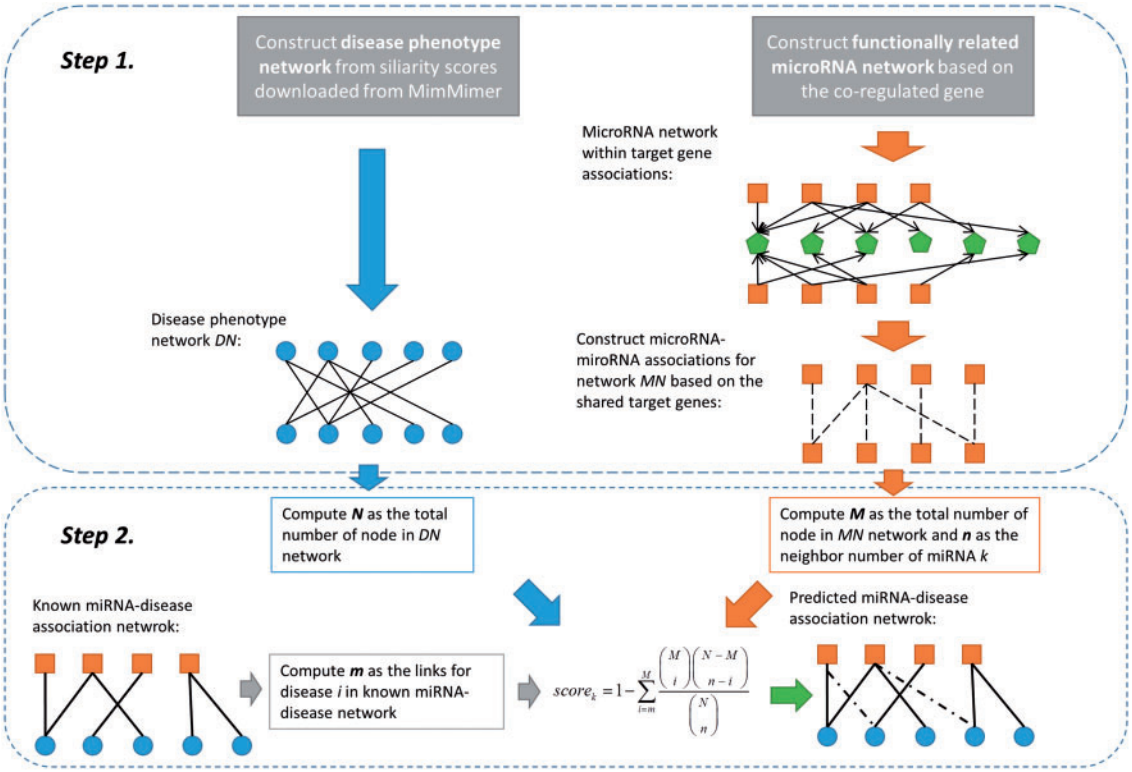

Figure 1. The flowchart of Jiang's method, which is mainly constructed by two steps. (Step 1) The disease phenotype network and functionally related miRNA network are constructed in Step 1; (Step 2) a prioritizing algorithm based on hypergeometric distribution is adopted in Step 2 to calculate the relevant scores between candidate miRNAs and disease of interest.

related to a disease, this model combines within-score and between-score as follows:

$$F(m(j),d(i)) = \frac{C_m^w(m(j),d(i)) \times C_d^w(m(j),d(i))}{C_m^b(m(j),d(i)) \times C_d^b(m(j),d(i))}$$

Here, C_m^w and C_d^w denote the within-scores as follows:

$$C_m^w(m(j),d(i)) = \max_{pm_p^i \in pm^i} S_m(m(j),pm_p^i)$$

$$C_d^w(m(j),d(i)) = \max_{pd_u^j \in pd^j} S_d(d(i),pd_u^j)$$

where pm^i is the miRNA group that has known relation with disease $d(i)$, pd^j is the disease group that is associated with miRNA $m(j)$ in the known miRNA–disease association data set. C_m^b and C_d^b denote the between-scores as follows:

$$C_m^b(m(j),d(i)) = \max_{nm_q^i \in nm^i} S_m(m(j),nm_q^i)$$

$$C_d^b(m(j),d(i)) = \max_{nd_v^j \in nd^j} S_d(d(i),nd_v^j)$$

where nm^i is the miRNA group that does not have known relation with disease $d(i)$, nd^j was the disease group that is not proved to be associated with miRNA $m(j)$ in the known miRNA–disease association data set. What is more, WBSMDA can predict potential miRNAs for new disease d as follows:

$$F(m(j),d) = \frac{C_d^w(m(j),d)}{C_d^b(m(j),d)}$$

and predict potential diseases for new miRNA m as follows:

$$F(m,d(i)) = \frac{C_m^w(m,d(i))}{C_m^b(m,d(i))}$$

As a result, WBSMDA obtained reliable performance with the AUC of 0.8031 based on the LOOCV. However, the miRNA functional similarity used in this study is not experimentally supported and needs to be improved by further research.

Complex network algorithm-based models

RWRMDA

Chen *et al.* have developed a random walk-based computational model of Random Walk with Restart for MiRNA–Disease Association (RWRMDA) for predicting novel human miRNA–disease associations (Figure 2) [64]. Specifically, RWRMDA implements random walk on the miRNA functional similarity network (MFSN). RWRMDA regards the experimentally supported disease-related miRNAs as seed miRNAs and searches most potential miRNA–disease associations by using random walk with restarts (RWR):

$$p(t+1) = (1-r)Wp(t) + rp(0)$$

Here, $p(t+1)$ and $p(t)$ denotes the probability vectors for walker at the time $t+1$ and t; $p(0)$ is the initial probability vector for walker; r denotes the restart probability; W is the column-normalized miRNA–miRNA functional similarity matrix. As a result, RWRMDA achieved an AUC of 0.8617 based on LOOCV. The main limitations of RWRMDA lie in the invalid application for new disease without known associated miRNAs.

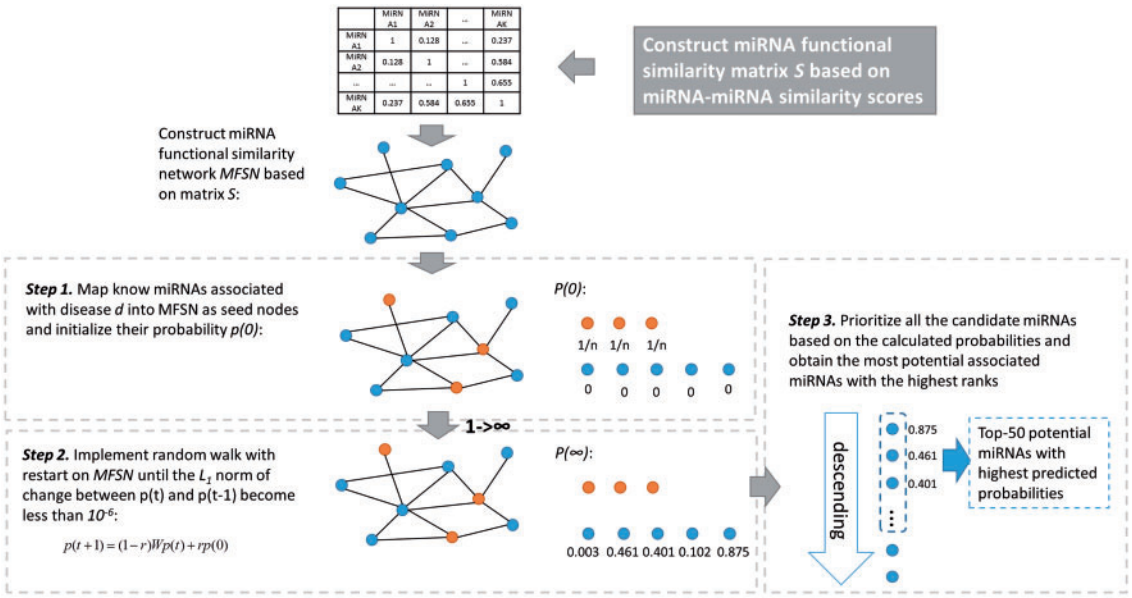

Figure 2. The flowchart of RWRMDA, which mainly depicts three main steps: (Step 1) map known miRNA associated with disease onto MFSN network and initialize their probabilities; (Step 2) implement RWR on MFSN until the convergence condition is reached; (Step 3) Rank the candidate miRNA and pick the high-ranked miRNAs as most potential associations.

MIDP

Xuan *et al.* have proposed the model of MIDP by implementing random walk on the miRNA–disease bilayer network [197]. This model hypothesizes that functionally similar miRNAs are normally implicated in similar disease and vice versa. Therefore, it computes functional similarity of miRNAs by measuring the semantic similarities of associated diseases and constructs a MFSN termed Mnet. The transition matrix M is then constructed based on Mnet and separated into two transition matrixes (i.e. M_Q and M_U) for labeled nodes and unlabeled nodes. Finally, the random walk is implemented to estimate the relevance scores for disease-related miRNAs:

$$S(t+1) = r_Q M_Q^T S(t) + p_Q(1-r_Q)X + r_U M_U^T S(t) + p_U(1-r_U)X$$

Here, $S(t+1)$ and $S(t)$ denote the probability vectors for walkers to reach the next vertex at the time $t+1$ and t; r_Q and r_U denote the weights of labeled nodes and unlabeled nodes; p_Q and p_U are the sum of the probability that the walker arrives at each labeled node and unlabeled node at time t. X is the choice vector for walker to go back or to restart walking. As a result, this model achieved prediction performance with AUCs ranges from 0.786 to 0.945 by being applied to 18 human data sets. Considering the newly confirmed miRNA–disease association, the proportion of labeled nodes will keep increasing and therefore the values of r_Q and r_U need manual intervention for adjustment in future application.

NTSMDA

By adopting a novel measure approach for network topological similarity, Sun *et al.* proposed the computational model of NTSMDA to infer miRNA–disease associations [198]. This model considers the topological information of known miRNA–disease association network by using a diffusion algorithm and constructs the integrated adjacent matrixes for miRNAs and diseases, *SD* and *SM*, by using the Gaussian interaction profile kernel. The relevance scores for each miRNA–disease pair are then computed by using a two-step inference algorithm:

$$f'(d_j) = \alpha \sum_{i=1}^{N_m} \frac{sm_{ji} f_m(m_i)}{k_m(m_i)} + (1-\alpha) \sum_{i=1}^{N_m} \frac{sd_{ji} f_d(m_i)}{k_d(m_i)}$$

$$f''(m_i) = \beta \sum_{l=1}^{N_d} \frac{sm_{li} f'(d_l)}{k_m(d_l)} + (1-\beta) \sum_{l=1}^{N_d} \frac{sd_{li} f'(d_l)}{k_d(d_l)}$$

Here, $f_m(m_i)$ and $f_d(m_i)$ denote the initial resource vectors of miRNA m_i in *SM* and *SD*, respectively; $k_m(m_i)$ and $k_d(d_l)$ denote the sum of row i in *SM* and that of row j in *SD*; α and β are two damping factors for contribution balance between sm_{ij} and sd_{ij}, which are the elements of *SM* and *SD* in the i^{th} row and j^{th} column. Based on the computed $f'(m_i)$, which depicts the relevance scores of diseases to miRNA m_i, the potential miRNA–disease associations could be obtain by ranking. As a result, this model obtained good performance with an AUC of 0.894 by implementing the LOOCV experiment. However, this model largely depends on the reliability of known miRNA–disease associations and fails to remain the prior information on similarities among miRNAs and diseases. In addition, NTSMDA cannot be applied to new disease without any known related miRNAs.

HGIMDA

Chen *et al.* have recently proposed a model of Heterogeneous Graph Inference for MiRNA–Disease Association prediction (HGIMDA) based on the combination of known miRNA–disease association network, integrated miRNA similarity network and disease similarity network, which reflected the path relations within these three networks [199]. Specifically, the integrated miRNA–miRNA and disease–disease similarities are computed by integrating Gaussian interaction profile kernel similarity with miRNA functional similarity and disease semantic similarity, respectively. Based on the integrated miRNA and disease similarity, the potential association probability of each disease–miRNA pair is further defined as follows:

$$P(d,m) = \sum_{i=1}^{nm} \sum_{j=1}^{nd} SM(m_i, m) * A(m_i, d_j) * SD(d_j, d)$$

Here, SM, SD and A denote the integrated similarity matrix of miRNAs and disease and miRNA–disease adjacency matrix. To calculate $P(d,m)$ by an iterative approach, HGIMDA normalizes SM and SD and then obtains association probability matrix P based on the following equation:

$$P(i+1) = \alpha SM \times P(i) \times SD + (1-\alpha)A$$

Here, α denotes the decay factor. This interactive equation would finally reach stable state with a cutoff of 10^{-6}. As a result, this model achieved AUCs of 0.8781 and 0.8077 based on global and local LOOCV, respectively. However, HGIMDA has two main limitations: it fails to additionally integrate other types of biological information and needs manual intervention for parameter adjustment.

MFSP

Ding *et al.* developed a novel path-based calculation method of miRNA functional similarity based on miRNA–disease associations, called MFSP [200]. First, the hierarchical structure about disease obtained from Medical Subject Heading (MeSH) descriptor was transferred into features of diseases. Second, the semantic similarity of disease pair S_{DD} was calculated by cosine similarity and the disease similarity network was constructed. Third, the weight sum of paths among diseases was achieved based on the different transferring times respectively. Fourth, the miRNA–miRNA path matrix P was formed based on the weight sum of paths among disease sets as follows:

$$P = \frac{\sum_{i=0}^{b} a^i R_{MD} M_i R_{MD}^T}{\sum_{i=0}^{b} a^i}$$

where $M_i = (S_{DD})^i$ represented the transferring matrix for transferring times i when the maximum transferring times b was given, R_{MD} represented the adjacency matrix of miRNA–disease association and a^i was the weight. Finally, they measured the functional similarity by miRNA–miRNA path matrix calculated as:

$$MFSP(m_i, m_j) = \frac{2 * P(m_i, m_j)}{P(m_i, m_i) + P(m_j, m_j)}$$

PBMDA

You *et al.* have proposed a novel and effective Path-Based computational model for MiRNA–Disease Association (PBMDA) by integrating known human miRNA–disease associations, integrated miRNA similarity and integrated disease similarity (Figure 3) [201]. Specifically, they constructed a heterogeneous graph with lots of paths, which consisted of three weighted matrixes as follows:

$$W_{miRNA-miRNA} = \begin{cases} 0 & S_m(m(i), m(j)) < T \\ S_m(m(i), m(j)) & otherwise \end{cases}$$

$$W_{disease-disease} = \begin{cases} 0 & S_d(d(i), d(j)) < T \\ S_d(d(i), d(j)) & otherwise \end{cases}$$

$$W_{miRNA-disease} = Y(m_i, d_j) \qquad 0 \le i \le nm, 0 \le j \le nd$$

where T is a threshold variable (experimentally set as 0.5), $S_m(m(i), m(j))$ is the integrated similarity between miRNAs $m(i)$ and $m(j)$, $S_d(d(i), d(j))$ is the integrated similarity between diseases $d(i)$ and $d(i)$, Y is the known association matrix, n_m and n_d are the numbers of miRNAs and diseases, respectively. Accordingly,

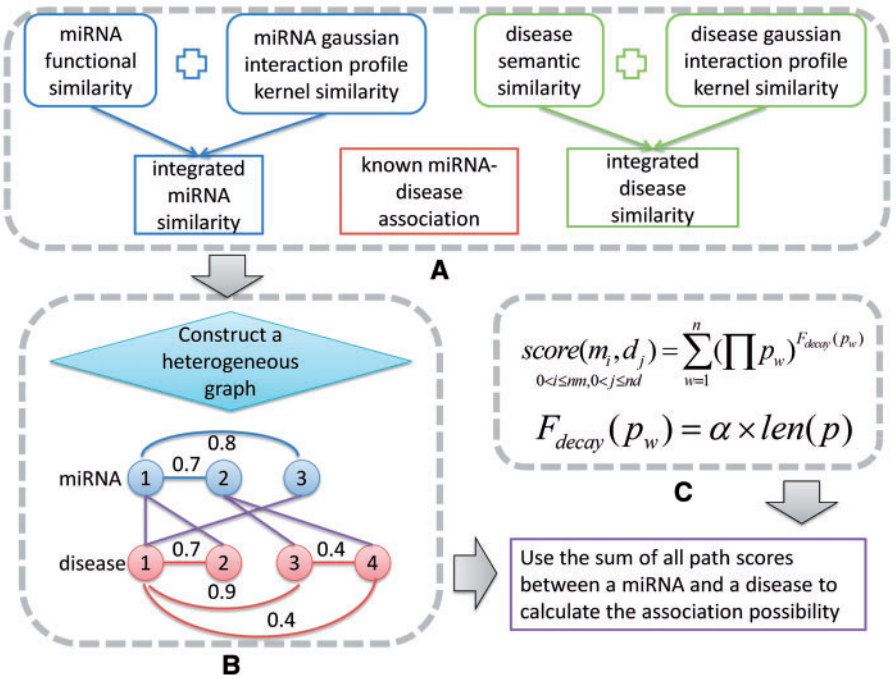

Figure 3. The flowchart shows the three steps of PBMDA: (**A**) combine known miRNA–disease association, integrated miRNA similarity and integrated disease similarity as the input data of section (**B**); (**B**) construct a heterogeneous graph consisting of three typical weighted networks; (**C**) calculate the potential association probability between miRNA and disease by the scoring system.

the scoring formula related to miRNA m_i and disease d_j was defined as follows:

$$\underset{0<i\le nm,0<j\le nd}{score(m_i,d_j)} = \sum_{w=1}^{n}\left(\prod p_w\right)^{F_{decay}(p_w)}$$

where $p=\{p_1,p_2,\ldots,p_n\}$ is a set of paths linking up a miRNA m_i and a disease d_j, $\prod p_w$ represents the product of the weight of all the edges in path p_w obtained from three weighted matrixes mentioned above, $F_{decay}(p_w)$ is the decay function calculated as follows:

$$F_{decay}(p_w)=\alpha\times len(p)$$

where parameter α is a decay factor, and *len(p)* is the length of path *p*. As a result, this model achieved AUCs of 0.9169 and 0.8341 based on global and local LOOCV, respectively. However, PBMDA has limitations that the distance-decay function in their approach is relatively simple and the problem of sparsity in known miRNA–disease association matrix still influences the performance of PBMDA.

Machine learning-based models

Support vector machine-based classification model

Xu *et al.* have developed a supervised learning-based classification model to predict the potential miRNA related to prostate cancer [202]. This model first constructs the miRNA target-dysregulated network (MTDN) and then represents candidate miRNAs by using four-dimension features, which are finally used as feature vectors for support vector machine (SVM) to learn classifier and make prediction. Specifically, this model measures the dysregulation degrees of target genes by comparing tumor and nontumor subgroups to construct MTDN, which is a bipartite graph containing two kinds of vertices (i.e. miRNA nodes and target gene nodes). There are four measures defined to retain topological information of MTDN: the number of dysregulated target genes and coregulators, the proportion of disease-related miRNAs in coregulator set and the fraction of coregulated targets. As a result, this model achieved an average prediction accuracy of 0.8872 in 5-fold cross-validation tests. However, the negative samples are hard to obtain from the data resource so that the effect of this model has been restricted. In addition, this model has been only performed on the data of prostate cancer. And therefore whether these four measures are suitable for other disease or not is unknown because diverse diseases could have different mechanisms and underlying MTDNs.

HDMP

Xuan *et al.* have proposed a prediction model for inferring the potential disease-related miRNAs by searching the most similar miRNA neighbors [203]. The novelty of this model lies in considering the information from miRNA family or cluster based on the observation that miRNAs from the same family/cluster tend to be involved in similar diseases. HDMP computes disease similarity by incorporating information content of disease terms and disease phenotype similarity based on the disease MeSH Directed Acyclic Graphs (DAGs). And then the functional similarity of each miRNA pair (say u and v) is measured based on their associated disease groups (say DTu and DTv):

$$Misim(u,v)=\frac{\sum_{1\le i\le|DT_u|}S(d_i,DT_v)+\sum_{1\le j\le|DT_v|}S(d_j,DT_u)}{|DT_u|+|DT_v|}$$

Here, $S(d,DT_v)$ and $S(d,DT_u)$ denote the similarity between disease *d* and disease groups DT_v and DT_u, respectively. Based on the constructed MFSN, HDMP then assigns miRNAs from the same family/cluster with higher weights when calculating relevance scores for these miRNAs because the members in the same miRNA family/cluster are more probably associated with similar diseases. Finally, the relevance score of unlabeled miRNAs was computed and highly ranked miRNAs are considered as potential candidates. As a result, HDMP achieved average AUC of 0.825 when being applied to 18 human diseases. However, the model is invalid for the new diseases that have no known associated miRNAs. In addition, HDMP could not perform better than most of the previous models, which were calculated based on the global network similarity measure.

RLSMDA

The previous supervised learning-based model relied heavily on negative samples, which are impossible to obtain currently. Therefore, a semi-supervised learning-based computational model of Regularized Least Squares for MiRNA–Disease Association (RLSMDA) has been developed by Chen *et al.* for inferring human miRNA–disease associations [204]. In the model of RLSMDA, the putative miRNA–disease associations are yielded by a combined classifier in the disease and miRNA space (Figure 4). Specifically, RLSMDA applied the method of Regularized Least Squares (RLS) to construct two optimal classifiers based on miRNA functional similarity and disease semantic similarity, which is measured by the DAG of disease MeSH descriptors. The RLS-based optimal classifiers for disease and miRNA function with the following solutions for optimization problems:

$$F_M^*=SM*(SM+\eta_M*I_M)^{-1}*A^T$$

$$F_D^*=SD*(SD+\eta_D*I_D)^{-1}*A$$

Here, η_M and η_D denote the trade-off parameters; *SM* and *SD* denote the matrixes of miRNA functional similarity and disease semantic similarity; I_M and I_D are the identity matrixes with the same sizes of *SM* and *SD*, respectively; *A* denotes adjacency matrix of disease–miRNA associations. Based on the optimal classifiers F_M^* and F_D^*, a combined classifier is finally constructed for inferring miRNA–disease associations:

$$F^*=w*F_M^{*T}+(1-w)*F_D^*$$

where *w* is the damping parameter for balancing the contribution of two optimal classifiers and the entity of *F(i,j)* denote the relevance scores for disease *i* and miRNA *j*. As a result, RLSMDA achieved AUC values of 0.8450 and 0.9511 based on the frameworks of local and global LOOCV. With the application of disease similarity information, RLSMDA can be applied to the new disease without any associated miRNAs. However, this model

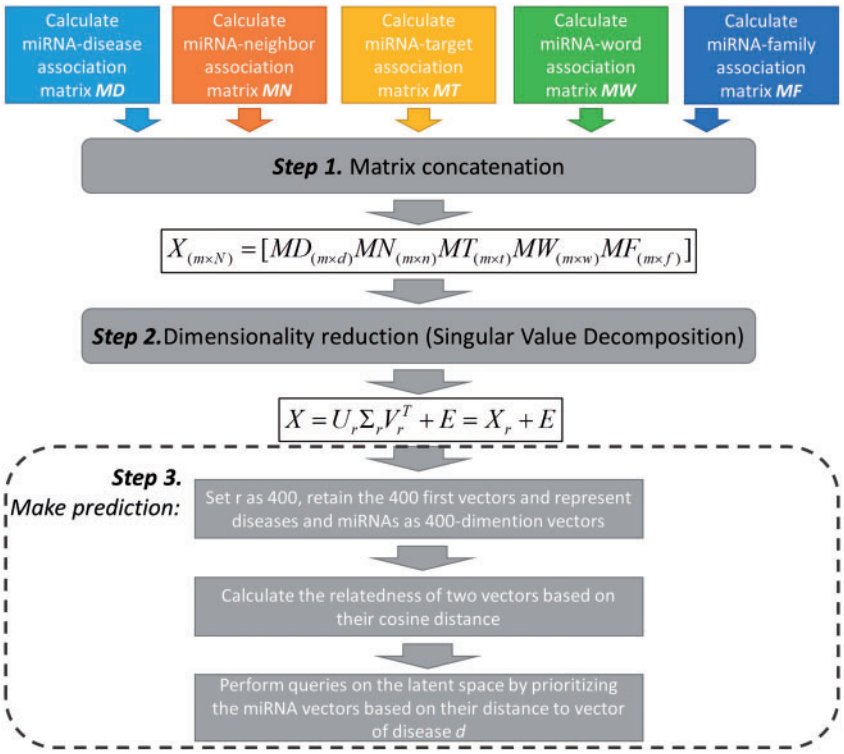

Figure 4. The flowchart shows the three steps of miRAI: (Step 1) combine five distinct miRNA-associated matrixes into one by simple matrix concatenation; (Step 2) use SVD for dimensionality reduction; (Step 3) calculate the relatedness between miRNA vectors and disease vectors and rank the distances for query performance.

needs manual intervention for the damping parameter for balancing the contribution of two optimal RLS-based classifiers.

RBMMMDA

Chen *et al.* have developed the model of Restricted Boltzmann Machine for Multiple types of MiRNA–Disease Association prediction (RBMMMDA) based on Restricted Boltzmann Machine (RBM) for inferring multiple types of miRNA–disease associations [37]. RBMMMDA used the RBM, which is the core of deep learning to provide a self-contained framework to obtain competitive classifiers directly. Different from other computational model for miRNA–disease association prediction, there are four types of miRNA–disease associations (i.e. miRNA–target interactions, circulation, epigenetics and genetics) predicted by RBMMMDA, which can provide more comprehensive information for inferred miRNA–disease associations. Specifically, RBMMMDA applied RBM by constructing a two-layer undirected graph in which the potential features of both disease and miRNA are represented as visible and hidden units, respectively. Given n visible units, m hidden units and t types of miRNA–disease associations, each visible unit is a binary vector $v_i = (v_i^1, \ldots, v_i^k, \ldots, v_i^t), i = 1, \ldots, n$ denoting the state of unit and hidden layer is encoded as $h = (h_1, \ldots, h_m)$. RBM defined the energy of it joint configuration (v, h) as follow:

$$E(v,h) = -\sum_{i=1}^{n}\sum_{k=1}^{n} a_i^k v_i^k - \sum_{j=1}^{m} b_j h_j - \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{t} W_{ij}^k h_j v_i^k$$

Here, a_i^k and b_j denote bias weights of visible and hidden units; W_{ij}^k denotes the weight between variable v_i^k and h_j. RBMMMDA further adopts a mean-field version of Contrastive Divergence algorithm to train RBM and makes prediction based on conditional probabilities. As a result, RBMMMDA achieved reliable prediction performance with the AUC of 0.8606 by implementing LOOCV. However, this model fails to consider additional biological information like miRNA functional similarity and disease similarity, and cannot be applied to the new diseases without any associated miRNAs.

MiRAI

Pasquier *et al.* have developed a singular value decomposition-based (SVD-based) vector space model for inferring miRNA–disease associations by considering multiple miRNA-associated information sources (Figure 5) [205]. This method integrates five distinct matrixes from multiple miRNA-associated information including miRNA–disease, miRNA–neighbor, miRNA–target, miRNA–word and miRNA–family associations. The feature vectors are extracted by using SVD on the combined matrix, which is simply gathered by five matrixes:

$$X = [MD_{(m\times d)} MN_{(m\times n)} MT_{(m\times t)} MW_{(m\times w)} MF_{(m\times f)}]$$

Here, *MD*, *MN*, *MT*, *MW* and *MF* denote the matrixes of miRNA–disease, miRNA–neighbor, miRNA–target, miRNA–word and miRNA–family associations, respectively. To achieve the best prediction performance, this model chooses to use 400 dimensions as the vector space size according to the test result between 50 and 500. In other words, only the top 400 vectors of the approximate matrix yielded by SVD are retained to represent the diseases and miRNAs. Finally, this model yielded the ranked list for miRNAs related to disease d by prioritizing the cosine distances of miRNA vectors to the vector of disease d. As a result, an average AUC value of 0.875 was achieved by MiRAI model when being applied to 27 human diseases. However, this

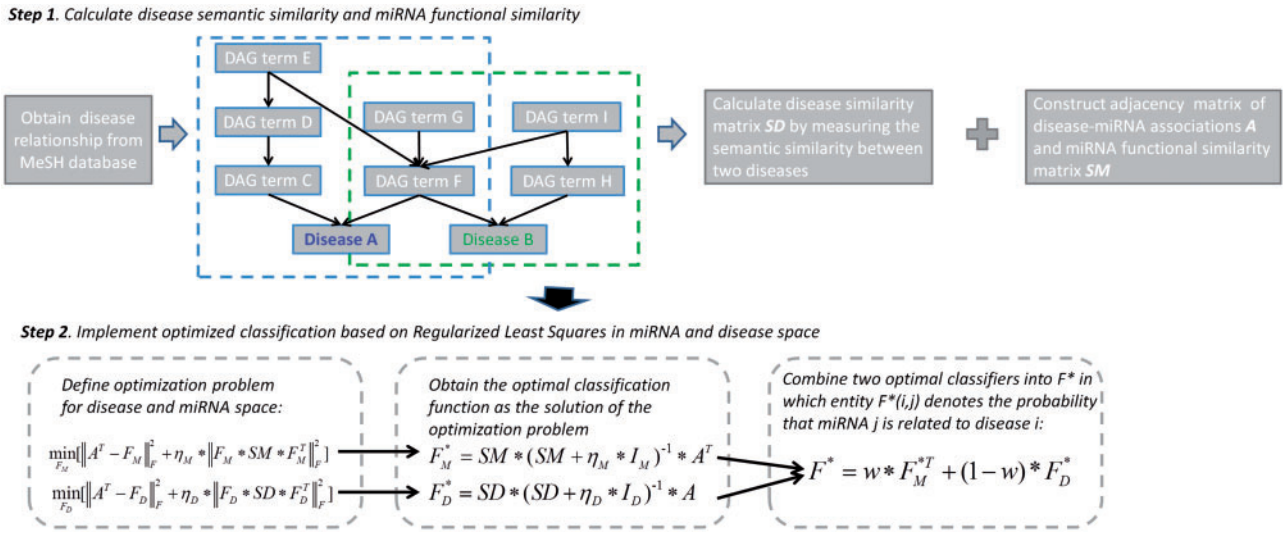

Figure 5. The flowchart of RLSMDA, which demonstrates two main steps: (Step 1) calculate disease semantic similarity and miRNA functional similarity; (Step 2) construct optimized classifiers based on RLS in miRNA and disease space.

model needs manual parameter adjustment to achieve its best performance because the dimension of vector space largely determines the capability of describing miRNAs and diseases. And the optimal parameters could be different for different data resources.

MCMDA

Li *et al.* developed a Matrix Completion for MiRNA–Disease Association prediction model (MCMDA) based on the known miRNA–disease associations in HMDD database (Figure 6) [206]. MCMDA uses the singular value thresholding algorithm to accomplish the matrix completion procedure. Specifically, the algorithm aims to obtain the final approximate complete matrix *X* through an iterative process that minimizes the difference between the scores of predicted associations in *X* and known miRNA–disease association matrix *M*, just as the following optimization problem:

$$\min_{X} f_{\tau}(X)$$

$$s.t. P_{\Omega}(X) = P_{\Omega}(M)$$

where P_{Ω} is the orthogonal projector onto the span of matrices vanishing outside of Ω so that the *(i,j)*th component of $P_{\Omega}(X)$ is equal to *X(i,j)* if $(i,j) \in \Omega$ or zero otherwise. $f_{\tau}(X)$ is a nonlinear function of *X*. To solve this problem effectively, the SVD of matrix X with rank r was adopted as follows:

$$X = U\Sigma V^{*}, \Sigma = diag(\{\sigma_i\}_{1 \le i \le r})$$

where *U* and *V* are $n_m \times r$ and $n_d \times r$ matrices, n_m and n_d are the numbers of known miRNAs and diseases. $\Sigma = diag(\{\sigma_i\}_{1 \le i \le r})$mean that Σ is a $r \times r$ diagonal matrix with positive singular values $\{\sigma_i\}_{1 \le i \le r}$ on its main diagonal. Then, an iterative process is implemented to obtain the final approximate complete matrix *X* of prediction scores. As a result, MCMDA achieved AUCs of 0.8749 in global LOOCV, 0.7718 in local LOOCV and average AUC of 0.8767 ± 0.0011 in 5-fold cross validation (CV). The result of case studies on the four human diseases illustrates that MCMDA achieved excellent prediction performance. Specially, MCMDA was effective and superior to the previous models in predicting potential miRNA–disease associations although it only depends on known miRNA–disease associations. However, MCMDA could not predict the potential miRNAs associated with the new diseases without any known related miRNAs and potential diseases associated with new miRNAs.

RKNNMDA

Chen *et al.* developed the computational model of Ranking-based k-nearest-neighbors for MiRNA–Disease Association prediction (RKNNMDA) to predict potential miRNA–disease associations by integrating k-nearest-neighbors (KNN) algorithm and SVM Ranking model (Figure 7) [207]. Specifically, they introduced the SVM Ranking model, which was a variant of the SVM algorithm to rank the previously sorted neighbors by extracting special features from training data sets. They calculated the inconsistent proportion of two examples to obtain the Hamming loss, which acted as an essential training data set for the SVM Ranking model to learn with. Given the selected miRNA *m(i)*, then the *m(i)*-associated disease label sets were obtained for each neighbor *m(j)*, marked as *md(j)*, and the *m(i)*-associated disease label set for *m(i)* itself, marked as *md(i)*, from adjacency matrix A, respectively. Then the Hamming loss could be defined as follows:

$$HammingLoss(m(i), m(j)) = \frac{|md(i) \Delta md(j)|}{|md(i) \cup md(j)|}$$

They constructed the corresponding weight score for each miRNA–disease association, which was obtained from the last step for the final possibility sorting by means of weighted voting. The weight score (WS) between *m(i)* and disease *d* was defined as the following formula:

$$WS(m(i), d) = \sum_{j=1}^{k} disease(neim'(i,j)) * 2^{k-j}$$

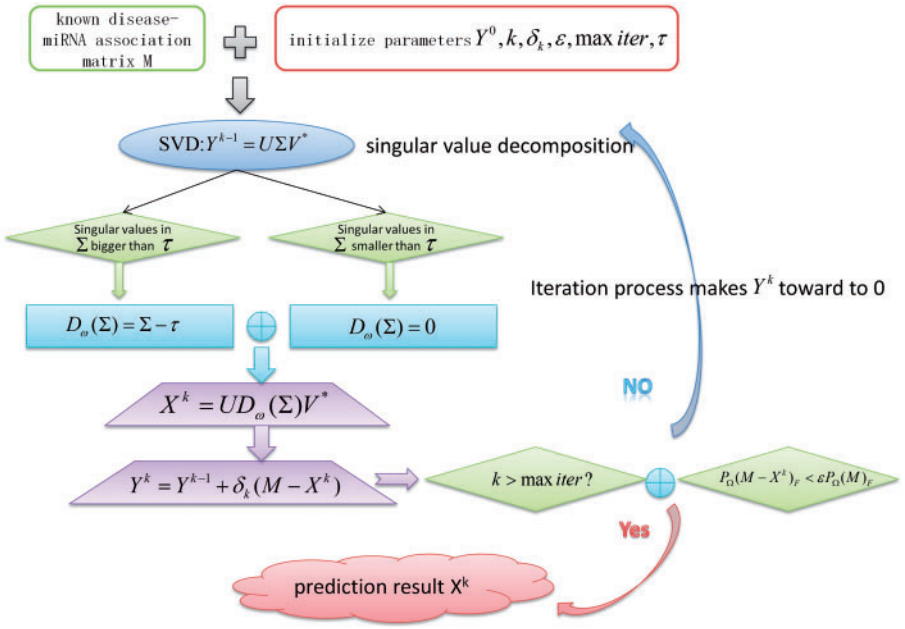

Figure 6. The Flowchart of MCMDA to predict the potential miRNA–disease associations. First, take known miRNA–disease association and initialize parameters as input data. Second, adopt SVD for a further iterative process of the matrix completion algorithm. Finally, the potential association score matrix is obtained when the stopping criteria of the iterative process is satisfied.

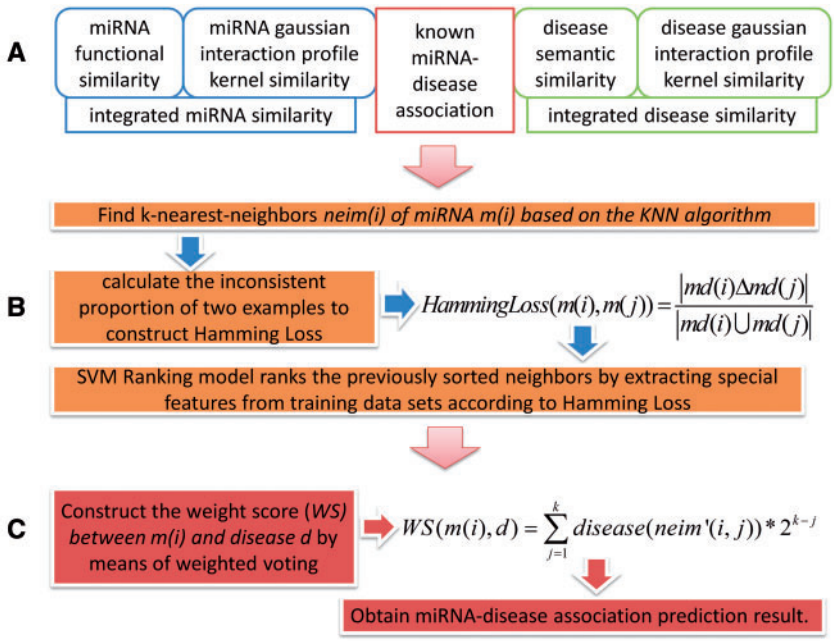

Figure 7. The flowchart of RKNNMDA, which demonstrates three main steps: (**A**) calculate integrated disease similarity, integrated miRNA similarity and input known miRNA–disease association; (**B**) find KNN of miRNA based on KNN algorithm and implement SVM Ranking model based on Hamming loss; (**C**) construct weight score (WS) to predict miRNA–disease association.

Here, *neim'(i,j)* refers to the jth neighboring miRNA of *m(i)* and *disease(neim'(i,j))* refers to the feature score of disease *d* extracted using miRNA functional similarity score between *m(i)* and *m(j)*, which could overcome the lack of real disease feature data to some extent. As the weight score value increases, the miRNA *m(i)* is more likely to be associated with disease *d*. As a result, RKNNMDA achieved AUC of 0.8221 based on LOOCV. In case studies, 96%, 80% and 94% of predicted top 50 potential related miRNAs for Colon Neoplasms, Esophageal Neoplasms and Prostate Neoplasms, respectively, had been confirmed by experimental literature. However, RKNNMDA may cause bias to miRNAs with more known associated diseases. Additionally, it still needs further studies on integrating similarity networks, KNN algorithm and SVM Ranking model for a much better performance of miRNA–disease association prediction.

KRLSM

Luo *et al.* recently proposed a model named KRLSM for predicting miRNA–disease associations using Kronecker RLS based on heterogeneous omics data [208]. They first adopted the algebraic properties of Kronecker product and combine the miRNA space and the disease space into a whole miRNA–disease space for predictions with the Kronecker product similarity matrix $S = S_M \otimes S_D$. With the combination of Kronecker product and the RLS classier, the optimization problem of KRLSM can be concluded in the miRNA–disease space as follows:

$$\min_{vec(\widehat{R}^T)} (\|vec(R^T) - vec(\widehat{R}^T)\|_F^2 + \sigma \|vec(R^T)^T \cdot S \cdot vec(R^T)\|_F^2)$$

The value of entity $\widehat{R}^*(i,j)$, which stands for the correlation predictive score between miRNA i and disease j can be obtained. As a result, the experiment results demonstrate that the proposed method outperforms the other state-of-the-art approaches. In addition, case studies of several common diseases further indicate the effectiveness of KRLSM to identify potential miRNA–disease associations.

Multiple biological information-based models

The two kinds of prediction models for miRNA–disease associations reviewed in the above two subsections only use single information directly related to miRNAs or diseases including experimentally supported miRNA–disease associations. However, the amount of known miRNA–disease is still insufficient owing to the difficulty of experiment-based identification. Considering this limited data volume, there are some prediction models that have been proposed by considering other types of prior biological information, such as protein- and target gene-associated networks, which can provide valuable insights for inferring miRNA–disease associations.

MiRPD

Mørk *et al.* have developed a computational model to infer miRNA–Protein–Disease associations (miRPD) based on the miRNA–protein–disease heterogeneous association network (Figure 8) [209]. The protein-related associations are used as bridges to link miRNAs and diseases. Specifically, miRPD uses text mining to obtain miRNA–protein relevance scores by mirSVR and Context +scores based on MiRanda and TargetScan, and protein–disease relevance scores based on Medline abstracts and quality scores assignment. The relevance score between miRNA M and disease D is computed as follows:

$$U(M,D) = \sum_{P^* \in P_M \cap P_D} T(M,P^*)Z(P^*,D)$$

Here, $T(M,P^*)$ and $Z(P^*,D)$ denote the relevance scores of miRNA M and disease D to their co-related protein P^*. However, there is an unbalance existing in research degree and known information for different proteins, which could cause biased prediction of miRPD. In addition, new diseases without any known associated miRNAs cannot be performed by this model.

Prediction model based on links among protein interaction network

Shi *et al.* have proposed a prediction model for inferring miRNA–disease associations, which mainly considers the associations between miRNA targets and disease genes on the protein–protein interaction (PPI) network (Figure 9) [210]. This model maps

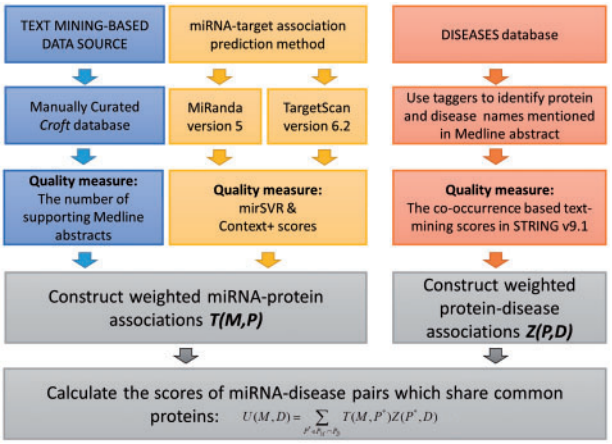

Figure 8. The flowchart of MiRPD, which depicts a complex process to construct weighted miRNA–protein and protein–disease associations, following by a scoring method based on common proteins.

disease genes and miRNA targets onto PPI network and measures enrichment scores, ES_1 and ES_2, by implementing RWR on PPI network with disease gene and miRNA targets as seed nodes, respectively. The enrichment score based on PPI network is computed as follows:

$$ES = \max_{1 \le i \le N} \left(\sum_{g_j \in G, j \le i} \sqrt{(N-n)/n} - \sum_{g_j \in G, j \le i} \sqrt{n/(N-n)} \right)$$

Here, $G = \{g_1, g_2, \ldots, g_n\}$ denotes the set of miRNA target or disease genes; N is the number of genes in PPI network. Based on the integrated enrichment scores, the P-value is computed and used for measuring the potential regulatory associations between miRNAs and diseases. As a result, this model explored nine cancer types and yielded AUCs ranging from 0.713 to 0.913. The model not only infers the most potential miRNA–disease association but also revealed the protein pathways and co-regulated module involved in the miRNA–disease association network. However, this model suffers from the biased prediction caused by the unbalanced protein information and cannot be applied to the new disease without any associated miRNAs.

Prediction model based on target gene-associated interactions

Xu *et al.* have proposed a prediction model to prioritize and identify the most potential miRNAs related to multiple diseases by constructing the interaction network between miRNAs and target genes, and that between target genes and diseases [98]. Specifically, this model collects miRNA–mRNA interactions from TCGA and GEO databases and the interactions with inverse correlations are further used for screen context-dependent miRNA–target gene interactions by implementing seven prediction algorithms. To obtain the relevance scores for miRNA–disease associations, this model integrates three types of biological information including GO sub-ontology biological processes, KEGG pathway information and average shortest path (ASP) in protein interaction network:

$$Score = \exp^{\sum_{j=1}^{N} p_{ij} \left(\frac{w_1 \times NTO_j^{GO} + w_2 \times NTO_j^{KEGG} + w_3 \times ASP_j}{w_1 + w_2 + w_3} \right)}$$

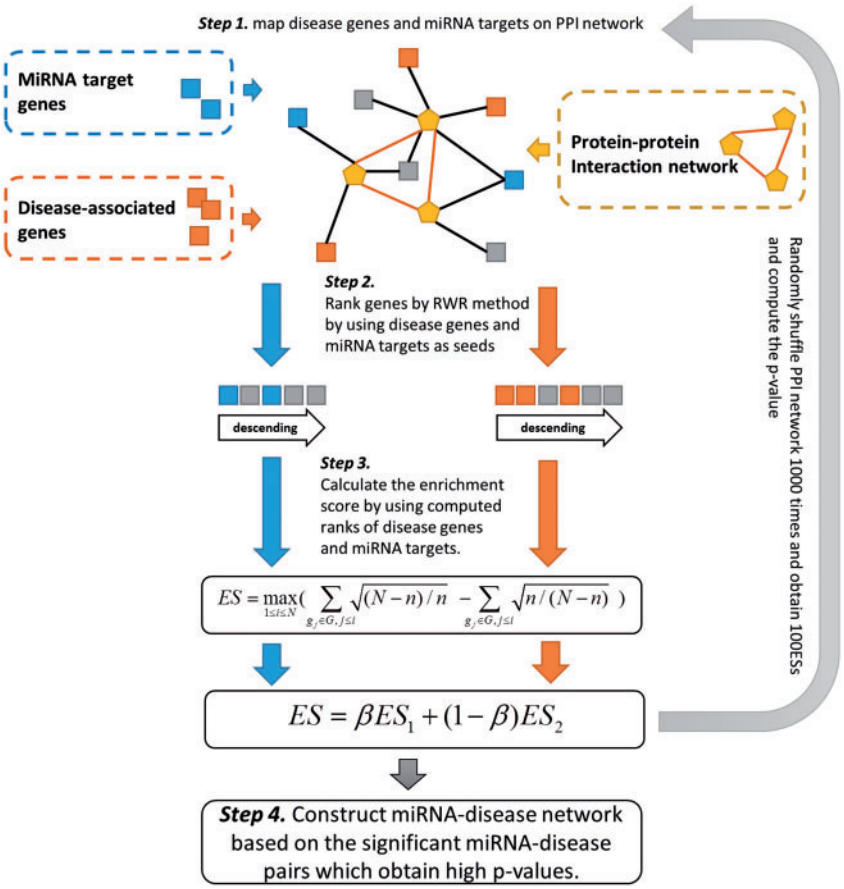

Figure 9. The flowchart of Shi's work, which mainly depicts four basic steps: (Step 1) map disease genes and miRNA targets onto PPI network; (Step 2) apply RWR on the heterogeneous network by using disease genes and miRNA targets as seeds and then implement probability ranking; (Step 3) repeat Steps 1 and 2 for 1000 times and calculate ES scores and the final P-value; (Step 4) construct predicted miRNA–disease network.

Here, w_1, w_2 and w_3 denote the contribution weights; p_{ij} is the disease phenotype similarity between disease i and j, which is derived from the MimMiner by computing correlation scores between disease phenotypes; NTO denotes the normalized term overlap score for biological processes and KEGG pathway information. Finally, all unlabeled miRNAs are prioritized based on the integrative scores and those with highest ranks are considered as the most potential candidates. As a result, this model yielded an average AUC of 0.7584 when exploring 11 cancer types. However, the interaction network between miRNAs and target genes constructed in this model is unreliable, most of whose interactions have not been confirmed by experimental observations. Furthermore, additional manual intervention is needed to optimize the parameters (i.e., w_1, w_2 and w_3) for better performance.

KBMFMDI

Lan *et al.* propose a computational framework named KBMFMDI to infer the relationship between miRNA and disease by integrating multiple data resources to measure disease similarity and miRNA similarity. In addition, a global method based on multiple kernel learning is used to predict potential miRNA–disease relationship [211]. Specifically, they exacted feature of miRNA domain by projecting kernel matrix $K_{M,a}$ into the R-dimension subspace with projection matrix A_M. Then, the composite components of different domain were obtained by linearly combining the kernel-specific components of different domain.

$$H_M = \sum_{a=1}^{P_M} e_{M,a} G_{M,a}$$

$$G_{M,a} = A_M^T K_{M,a}$$

where e_M and P_M denotes kernel weight value and the number of kernel matrix of miRNA. Similar to miRNA domain, the kernel-specific components of disease domain also can be obtained. Then latent score matrix $F = H_M^T H_D$ can be calculated. The experimental results demonstrate that this method can be effective to predict potential miRNA–disease association, especial in some diseases with few related miRNAs.

Discussion and conclusion

New knowledge about miRNA function may benefit the identification of disease-related miRNAs, which will further accelerate the understanding of disease pathogenesis at the molecular level and molecular tools design for disease diagnosis, treatment and prevention [92, 93, 97, 212]. Computational models have become an important means for novel miRNA–disease association identification, which could select the most promising miRNA–disease pair for experimental validation and significantly reduce the time and cost of the biological experiments.

The difficulties of predicting potential disease-related miRNAs lie in the following factors: (1) known miRNA–disease associations are still rare; (2) negative miRNA–disease associations are unavailable; (3) because of relatively limited biological data sets about miRNAs, miRNA similarity calculation and integration still have some difficulties; (4) considering the fact that new miRNAs are discovered each year, computational models must be applied to miRNAs without any known associated diseases.

MiRNA–disease association prediction models could be developed based on the following several ways. First, disease module principle (similar diseases tend to be related with functional similar miRNAs) could be used to predict novel disease–miRNA associations based on known experimentally verified associations [37, 64]. Second, potentially related miRNAs for the given disease could be identified by calculating the functional similarity between the target genes of the candidate miRNAs and known disease genes. Third, considering the topological properties of MFSN, computational models used for social network analysis could be extended to miRNA–disease association prediction. Furthermore, considering the fact that members in the same miRNA family or cluster tend to be related with the same or similar diseases, miRNA family and cluster information could be useful for model construction. Moreover, almost all the computational models can only predict whether there is association between investigated diseases and miRNAs, only RBMMMDA could predict the association types between miRNAs and diseases [37]. Therefore, more efforts should be made to solve these important biological models. Finally, the computational methods that will be proposed in the future should not only predict miRNA–disease associations, but also predict disease-related miRNA–EF interactions and disease-related miRNA–target interactions. Specially, Chen *et al*. has developed a novel disease-related miRNA–EF interactions prediction method called miRNA-EF interactions inference based on the random walk with restart (miREFRWR) by implementing random walks on miRNA similarity network and EF similarity network to uncover the hidden disease-related miRNA–EF interactions [213]. They also proposed a method of miRNA-EF interaction prediction with semi-supervised classifier (miREFScan) to predict disease-related interactions between miRNAs and EFs based on a semi-supervised classifier [115]. These studies have provided a direction of the further studies for disease-related miRNA–EF interactions and miRNA–target interactions.

In this article, we summarized the biological functions of miRNAs and introduced the link between miRNAs and diseases with specific examples of five miRNA-related diseases and five disease-related miRNAs. Some publicly available databases were also introduced, which collect the comprehensive information on miRNAs including sequences, function, expression and target genes. In addition, we reviewed some previously proposed computational models, which aim at inferring miRNA–disease association on a large scale and can yield state-of-the-art prediction performance. Based on the work principles, these models were further classified into four categories, namely, score function-based, complex network algorithm-based, machine learning-based and multiple biological information-based models. The effective performances of all models reviewed in this article have been evaluated by exploring different data sets, which largely depends on similarity measures and the integration of various kinds of information for different purposes. However, different ways to construct computational models can bring different limitations. We therefore summarized the advantages and disadvantages of each category and discussed the potential directions for the future computational miRNA-associated research.

Score function-based models usually adopt probabilistic algorithms such as cumulative hypergeometric distribution and some other special proposed algorithms like within and between scores. The advantage of this category lies in the simplicity and reliability for the understanding of the algorithm theory and the computational operation of the algorithm experiment. However, the models of this category often suffer from the unreliability of data source from which the disease similarity and miRNA similarity are derived. For example, the functionally related miRNA similarity constructed in Jiang's work [195] was obtained based on Probability of Interaction by Target Accessibility, most of whose results have not been experimentally confirmed. In addition, for the probability distribution, it always has no prior knowledge about the distribution of the sample data. The category usually makes assumptions for the probability distribution, which may lead to an unsatisfied prediction results when the data resource does not coincide with the assumption.

The main feature of complex network algorithm-based models lies in the way that they construct two similarity networks (i.e. disease network and miRNA network) based on which the final prediction is made by specific global network-based measurement. Specifically, the miRNA similarity can be computed based on the co-regulated target genes, Gaussian interaction profile kernel and common associated disease groups; the disease similarity can be measured based on Gaussian interaction profile kernel, disease phenotype similarity scores from MimMiner, information content and semantic similarity from disease MeSH DAGs. Based on the constructed similarity networks, the models use different network-based algorithms such as random walk and diffusion algorithm to infer possible associations. The main advantage of these methods lies in the full use of topological information of the known miRNA–disease bilayer network. The framework of this category has a good transplantable character, which allows the corresponding models to be easily applied to other kinds of disease/miRNA similarity. And their basic hypothesis that functionally similar miRNAs tend to be involved in similar disease and vice versa accords with the observations from biological experiments. However, most of network similarity measurement-based models cannot be applied to the new disease without any known miRNAs because the miRNA similarities can only work through the known miRNAs associated with the given disease based on the basic assumption. The improvement on this kind of computational models is mainly owing to the more reliable and comprehensive construction of miRNA/disease function interaction network by considering additional experimental data.

Machine learning is garnering increasing research attentions and interests in bioinformatics and the field of prediction model construction for miRNA–disease associations is no exception. There are some popular machine learning-based algorithms (e.g. SVD, SVM, RBM and RLS) explored for inferring potential miRNA–disease associations, which can be classified into two work modes. The first mode is based on feature extraction. MiRAI [205] and Xu's method [98], which extract, respectively, 400-dimension and 4-dimension features, belong to this mode and face the common problem on parameter selecting on feature dimension and classifier configuration. And the second mode aims at optimizing a specific object function. For example, the models of RBMMMDA and RLSMDA define their specific object functions and solve the optimization problem based on different solutions. However, most of supervised learning models have a strict requirement on the data set used for learning because collecting negative miRNA–disease association

samples is difficult and even impossible. Furthermore, they often fail to retain the topological information of the known miRNA–disease bilayer network. The major advantage of machine learning-based models lies in that they can predict the potential miRNAs for new disease, which has no associated miRNA recorded. In addition, machine learning-based models have a large flexibility with different choices for feature extraction or optimization function. For example, the model of miRAI presents a SVD-based method to considering five types of miRNA-related associations by simply concatenating five matrices [205]. By modifying the local modules like adding or deleting features, machine learning-based model can consider various kinds of information for the final prediction and therefore have great potential to improve the prediction performance.

MiRNAs serve as a part of regulatory networks in epigenetic machinery and keep a close relationship with other types of biological molecules like enzymes and target genes. Considering the limited number of experimentally supported miRNA–disease associations, integrating multiple miRNA-associated networks would be a feasible and effective approach to provide a more comprehensive and reliable information source for constructing prediction model for miRNA–disease associations. There are some computational models proposed based on a heterogeneous association network constructed by additional interaction networks like PPI network and miRNA–genes–disease network. One major advantage of multiple biological information-based models is that these models can not only refer the most disease-related miRNA candidates but also can reveal how miRNAs influence diseases through the mediation of proteins or genes. However, the introduction of additional data source also brings the challenges for computational models to reasonably combine the various kinds of data source. Furthermore, the number of known miRNA-related interaction data is still insufficient, most of which were yielded by prediction methods and have not been experimentally confirmed. For example, the model of miRPD used two popular target prediction methods, MiRanda and TargetScan, to obtain two sets of miRNA–protein association [66]. Therefore, the insufficient amount of experimentally supported data limits the application ranges and the performance of these models can be largely influenced by the adopted prediction algorithms.

Nowadays, one of the major problems commonly faced by proposed prediction models for miRNA–disease associations is that the data volume of known miRNA–disease associations is still limited, and constructing a comprehensive and reliable association network for model learning even becomes the major task and difficulty for those models based on heterogeneous biological interaction networks. Therefore, the perspectives for further development of prediction models for miRNA–disease associations could include collecting a larger scale of experimentally validated data and considering other associated biological data. However, the introduction of other types of data can also bring additional challenge. Proposing a more effective and reasonable framework for data integration is expected to benefit the performance of future prediction models. It is worth noting that the prediction performance yielded by applying the same computational model to various diseases can be different. For example, the AUC values yielded by the model of miRAI achieved a maximum of 0.987 for Hypertrophic Cardiomyopathy but a minimum of 0.705 for Rectal Neoplasms [205]; Shi's method yielded the highest AUC of 0.913 for Nasopharyngeal Cancer but the lowest AUC of 0.713 for Breast Cancer [210]. Therefore, it is worthy to further analyze what kind of information is more meaningful and effective for referring potential miRNAs associated with specific diseases. Considering the advantage of ensemble learning and that there are some computational models have been explored on the same data sets, the relevant scores yielded by different computation models for a miRNA–disease pair can be further integrated by an ensemble model, which is expected to obtain more accurate and stable prediction performance.

Key Points

- As one of the most important components in the cell, miRNAs play a growing important role in various essential and important biological processes.
- MiRNAs have close associations with various human complex diseases.
- The functions of miRNAs, miRNA–target interactions and miRNA–disease associations were discussed in detail, respectively.
- Computational models have become an important means for novel miRNA–disease association identification, which could select the most promising miRNA–disease pair for experimental validation and significantly reduce the time and cost of the biological experiments.
- We summarized four important factors for the difficulties of predicting potential disease-related miRNAs.
- Based on the work principles, we further classified all the models into four categories, namely, score function-based, complex network algorithm-based, machine learning-based and multiple biological information-based models.

Funding

Fundamental Research Funds for the Central Universities (2017XKQY083 to X.C.).

References

1. Le TD, Liu L, Zhang J, *et al*. From miRNA regulation to miRNA-TF co-regulation: computational approaches and challenges. *Brief Bioinform* 2015;**16**:475–96.
2. Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* 2009;**23**:1494–504.
3. Taft RJ, Pang KC, Mercer TR, *et al*. Non-coding RNAs: regulators of disease. *J Pathol* 2010;**220**:126–39.
4. Lander ES, Linton LM, Birren B, *et al*. Initial sequencing and analysis of the human genome. *Nature* 2001;**409**:860–921.
5. Kapranov P, Cheng J, Dike S, *et al*. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 2007;**316**:1484–8.
6. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 2008;**322**:1845–8.
7. Claverie JM. Fewer genes, more noncoding RNA. *Science* 2005;**309**:1529–30.
8. Carninci P, Sandelin A, Lenhard B, *et al*. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* 2006;**38**:626–35.

9. Birney E, Stamatoyannopoulos JA, Dutta A, *et al*. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 2007;**447**: 799–816.
10. Bertone P, Stolc V, Royce TE, *et al*. Global identification of human transcribed sequences with genome tiling arrays. *Science* 2004;**306**:2242–6.
11. Chen X, Yan CC, Luo C, *et al*. Constructing lncRNA functional similarity network based on lncRNA-disease associations and disease semantic similarity. *Sci Rep* 2015;**5**:11338.
12. Chen G, Wang ZY, Wang DQ, *et al*. LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res* 2013;**41**:D983–6.
13. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell* 2009;**136**:629–41.
14. Esteller M. Non-coding RNAs in human disease. *Nat Rev Genet* 2011;**12**:861–74.
15. Taft RJ, Pheasant M, Mattick JS. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* 2007;**29**:288–99.
16. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Mol Cell* 2011;**43**:904–14.
17. Harrow J, Frankish A, Gonzalez JM, *et al*. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 2012;**22**:1760–74.
18. Guttman M, Russell P, Ingolia NT, *et al*. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 2013;**154**:240–51.
19. Derrien T, Johnson R, Bussotti G, *et al*. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 2012;**22**: 1775–89.
20. Wapinski O, Chang HY. Long noncoding RNAs and human disease (vol 21, pg 354, 2011). *Trends Cell Biol* 2011;**21**: 561.
21. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet* 2009;**10**:155–9.
22. Chen X. KATZLDA: KATZ measure for the lncRNA-disease association prediction. *Sci Rep* 2015;**5**:16840.
23. Chen X. Predicting lncRNA-disease associations and constructing lncRNA functional similarity network based on the information of miRNA. *Sci Rep* 2015;**5**:13186.
24. Chen X, Yan GY. Novel human lncRNA–disease association inference based on lncRNA expression profiles. *Bioinformatics* 2013;**29**:2617–24.
25. Liu MX, Chen X, Chen G, *et al*. A computational framework to infer human disease-associated long noncoding RNAs. *PLoS One* 2014;**9**:e84408.
26. Chen X, Yan CC, Zhang X, *et al*. Long non-coding RNAs and complex diseases: from experimental results to computational models. *Brief Bioinform* 2016;**18**:558–76.
27. Chen X, You ZH, Yan GY, *et al*. IRWRLDA: improved random walk with restart for lncRNA-disease association prediction. *Oncotarget* 2016;**7**:57919–31.
28. Chen X, Huang YA, Wang XS, *et al*. FMLNCSIM: fuzzy measure-based lncRNA functional similarity calculation model. *Oncotarget* 2016;**7**:45948–58.
29. Huang YA, Chen X, You ZH, *et al*. ILNCSIM: improved lncRNA functional similarity calculation model. *Oncotarget* 2016;**7**: 25902–14.
30. Chen X, Peng H, Yin Z. Current computational models for prediction of the varied interactions related to noncoding RNAs. *Biomed Res Int* 2016;**2016**:4183574.
31. Ambros V. The functions of animal microRNAs. *Nature* 2004; **431**:350–5.
32. 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**:843–54.
33. Lee RC, Ambros V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 2001;**294**:862–4.
34. Reinhart BJ, Slack FJ, Basson M, *et al*. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 2000;**403**:901–6.
35. Pasquinelli AE, Ruvkun G. Control of developmental timing by microRNAs and their targets. *Annu Rev Cell Dev Biol* 2002; **18**:495–513.
36. Griffiths-Jones S, Saini HK, van Dongen S, *et al*. miRBase: tools for microRNA genomics. *Nucleic Acids Res* 2008;**36**: D154–8.
37. Chen X, Yan CC, Zhang X, *et al*. RBMMMDA: predicting multiple types of disease-microRNA associations. *Sci Rep* 2015;**5**: 13877.
38. Tagawa H, Seto M. A microRNA cluster as a target of genomic amplification in malignant lymphoma. *Leukemia* 2005;**19**:2013–16.
39. He L, Thomson JM, Hemann MT, *et al*. A microRNA polycistron as a potential human oncogene. *Nature* 2005;**435**: 828–33.
40. Li J, Zhang YP, Wang YY, *et al*. Functional combination strategy for prioritization of human miRNA target. *Gene* 2014;**533**: 132–41.
41. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004;**116**:281–97.
42. Jiang W, Chen X, Liao M, *et al*. Identification of links between small molecules and miRNAs in human cancers based on transcriptional responses. *Sci Rep* 2012;**2**:282.
43. Cheng AM, Byrom MW, Shelton J, *et al*. Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis. *Nucleic Acids Res* 2005; **33**:1290–7.
44. Karp X, Ambros V. Developmental biology: enhanced: encountering MicroRNAs in cell fate signaling. *Science* 2005; **310**:1288.
45. Miska EA. How microRNAs control cell division, differentiation and death. *Curr Opin Genet Dev* 2005;**15**:563–8.
46. Xu P, Guo M, Hay BA. MicroRNAs and the regulation of cell death. *Trends Genet* 2004;**20**:617–24.
47. Cui Q, Yu Z, Purisima EO, *et al*. Principles of microRNA regulation of a human cellular signaling network. *Mol Syst Biol* 2006;**2**:46.
48. Huang Y, Shen XJ, Zou QA, *et al*. Biological functions of microRNAs: a review. *J Physiol Biochem* 2011;**67**:129–39.
49. Franco-Zorrilla JM, Valli A, Todesco M, *et al*. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet* 2007;**39**:1033–7.
50. Ebert MS, Neilson JR, Sharp PA. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat Methods* 2007;**4**:721–6.
51. Krutzfeldt J, Rajewsky N, Braich R. Silencing of microRNAs in vivo with 'antagomirs. *Nature* 2005;**438**:685–9.
52. Castoldi M, Schmidt S, Benes V, *et al*. miChip: an array-based method for microRNA expression profiling using locked nucleic acid capture probes. *Nat Protoc* 2008;**3**:321–9.
53. Cui L, Lin X, Lin N, *et al*. Graphene oxide-protected DNA probes for multiplex microRNA analysis in complex biological samples based on a cyclic enzymatic amplification method. *Chem Commun* 2012;**48**:194–6.

54. Bar M, Wyman SK, Fritz BR, *et al*. MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries. *Stem Cells* 2008;**26**:2496–505.
55. Koshizuka K, Nohata N, Hanazawa T, *et al*. Deep sequencing-based microRNA expression signatures in head and neck squamous cell carcinoma: dual strands of pre-miR-150 as antitumor miRNAs. *Oncotarget* 2017;**8**:30288–304.
56. Rebustini IT. A functional MicroRNA screening method for organ morphogenesis. *Curr Protoc Cell Biol* 2017;**74**: 19.19.1–19.19.17.
57. Zhang Y, Chao TF, Li R, *et al*. MicroRNA-128 inhibits glioma cells proliferation by targeting transcription factor E2F3a. *J Mol Med* 2009;**87**:43–51.
58. O'Connell RM, Taganov KD, Boldin MP, *et al*. MicroRNA-155 is induced during the macrophage inflammatory response. *Proc Natl Acad Sci USA* 2007;**104**:1604–9.
59. Williams AE. Functional aspects of animal microRNAs. *Cell Mol Life Sci* 2008;**65**:545–62.
60. Lecellier CH, Dunoyer P, Arar K, *et al*. A cellular MicroRNA mediates antiviral defense in human cells. *Science* 2005;**308**: 557–60.
61. Meister G, Tuschl T. Mechanisms of gene silencing by double-stranded RNA. *Nature* 2004;**431**:343–9.
62. Zhou BY, Wang S, Mayr C, *et al*. miR-150, a microRNA expressed in mature B and T cells, blocks early B cell development when expressed prematurely (vol 104, pg 7080, 2007). *Proc Natl Acad Sci USA* 2008;**105**:18070.
63. Ambros V. microRNAs: tiny regulators with great potential. *Cell* 2001;**107**:823–6.
64. Chen X, Liu MX, Yan G. RWRMDA: predicting novel human microRNA-disease associations. *Mol Biosyst* 2012;**8**:2792–8.
65. Lu M, Zhang Q, Deng M, *et al*. An analysis of human microRNA and disease associations. *PLoS One* 2008;**3**:e3420.
66. Mork S, Pletscher-Frankild S, Caro AP, *et al*. Protein-driven inference of miRNA-disease associations. *Bioinformatics* 2014;**30**:392–7.
67. Hobert O. Common logic of transcription factor and microRNA action. *Trends Biochem Sci* 2004;**29**:462–8.
68. Krek A, Grun D, Poy MN, *et al*. Combinatorial microRNA target predictions. *Nat Genet* 2005;**37**:495–500.
69. Doench JG, Petersen CP, Sharp PA. siRNAs can function as miRNAs. *Genes Dev* 2003;**17**:438–42.
70. Esquela-Kerscher A, Slack FJ. Oncomirs—microRNAs with a role in cancer. *Nat Rev Cancer* 2006;**6**:259–69.
71. Bandyopadhyay S, Mitra R, Maulik U, *et al*. Development of the human cancer microRNA network. *Silence* 2010;**1**:6.
72. Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell* 2009;**136**:215–33.
73. Zotos P, Roubelakis MG, Anagnou NP, *et al*. Overview of microRNA target analysis tools. *Curr Bioinform* 2012;**7**:310–23.
74. John B, Enright AJ, Aravin A, *et al*. Human MicroRNA targets. *PLoS Biol* 2004;**2**:e363.
75. 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**:15–20.
76. Krüger J, Rehmsmeier M. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res* 2006;**34**: W451–4.
77. Maragkakis M, Reczko M, Simossis VA, *et al*. DIANA-microT web server: elucidating microRNA functions through target prediction. *Nucleic Acids Res* 2009;**37**:W273–6.
78. Grün D, Wang YL, Langenberger D, *et al*. microRNA target predictions across seven Drosophila species and comparison to mammalian targets. *PLoS Comput Biol* 2005;**1**:e13.
79. Baev V, Daskalova E, Minkov I. Computational identification of novel microRNA homologs in the chimpanzee genome. *Comput Biol Chem* 2009;**33**:62–70.
80. Llave C, Xie Z, Kasschau KD, *et al*. Cleavage of scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science* 2002;**297**:2053–6.
81. Tang G, Reinhart BJ, Bartel DP, *et al*. A biochemical framework for RNA silencing in plants. *Genes Dev* 2003;**17**:49–63.
82. Matranga C, Zamore PD. Plant RNA interference *in vitro*. *Cold Spring Harb Symp Quant Biol* 2004;**69**:403–8.
83. Qi Y, Denli AM, Hannon GJ. Biochemical specialization within Arabidopsis RNA silencing pathways. *Mol Cell* 2005; **19**:421–8.
84. Haley B, Tang G, Zamore PD. *In vitro* analysis of RNA interference in Drosophila melanogaster. *Methods* 2003;**30**:330–6.
85. Schmittgen TD, Jiang J, Liu Q, *et al*. A high-throughput method to monitor the expression of microRNA precursors. *Nucleic Acids Res* 2004;**32**:e43.
86. 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**:1203–17.
87. Neely LA, Patel S, Garver J, *et al*. A single-molecule method for the quantitation of microRNA gene expression. *Nat Methods* 2006;**3**:41–6.
88. Aw SS, Tang MX, Teo YN, *et al*. A conformation-induced fluorescence method for microRNA detection. *Nucleic Acids Res* 2016;**44**:e92.
89. Zhang K, Wang K, Zhu X, *et al*. Sensitive detection of microRNA in complex biological samples by using two stages DSN-assisted target recycling signal amplification method. *Biosens Bioelectron* 2017;**87**:358–64.
90. He L, Hannon GJ. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* 2004;**5**:522–31.
91. Goh JN, Loo SY, Datta A, *et al*. microRNAs in breast cancer: regulatory roles governing the hallmarks of cancer. *Biol Rev* 2015;**91**:409–28.
92. Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer* 2006;**6**:857–66.
93. Duisters RF, Tijsen AJ, Schroen B, *et al*. miR-133 and miR-30 regulate connective tissue growth factor. *Circ Res* 2009;**104**: 170–8.
94. Huang Q, Gumireddy K, Schrier M, *et al*. The microRNAs miR-373 and miR-520c promote tumour invasion and metastasis. *Nat Cell Biol* 2008;**10**:202–10.
95. Iorio MV, Ferracin M, Liu CG, *et al*. MicroRNA gene expression deregulation in human breast cancer. *Cancer Res* 2005;**65**: 7065–70.
96. Latronico MV, Catalucci D, Condorelli G. Emerging role of microRNAs in cardiovascular biology. *Circ Res* 2007;**101**: 1225–36.
97. Markou A, Tsaroucha EG, Kaklamanis L, *et al*. Prognostic value of mature microRNA-21 and microRNA-205 overexpression in non–small cell lung cancer by quantitative real-time RT-PCR. *Clin Chem* 2008;**54**:1696–704.
98. Xu C, Ping Y, Li X, *et al*. Prioritizing candidate disease miRNAs by integrating phenotype associations of multiple diseases with matched miRNA and mRNA expression profiles. *Mol Biosyst* 2014;**10**:2800–9.
99. Porkka KP, Pfeiffer MJ, Waltering KK, *et al*. MicroRNA expression profiling in prostate cancer. *Cancer Res* 2007;**67**: 6130–5.
100. Yanaihara N, Caplen N, Bowman E, *et al*. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* 2006;**9**:189–98.

101. Akao Y, Nakagawa Y, Naoe T. MicroRNA-143 and -145 in colon cancer. *DNA Cell Biol* 2007;**26**:311–20.
102. Yang H, Kong W, He L, *et al*. MicroRNA expression profiling in human ovarian cancer: miR-214 induces cell survival and cisplatin resistance by targeting PTEN. *Cancer Res* 2008;**68**: 425–33.
103. Zhu X, Li Y, Shen H, *et al*. miR-137 inhibits the proliferation of lung cancer cells by targeting Cdc42 and Cdk6. *FEBS Lett* 2013;**587**:73–81.
104. Li Y, Qiu C, Tu J, *et al*. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res* 2014;**42**:D1070–4.
105. Xiao Y, Guan J, Ping Y, *et al*. Prioritizing cancer-related key miRNA-target interactions by integrative genomics. *Nucleic Acids Res* 2012;**40**:7653–65.
106. Ye S, Yang L, Zhao X, *et al*. Bioinformatics method to predict two regulation mechanism: TF-miRNA-mRNA and lncRNA-miRNA-mRNA in pancreatic cancer. *Cell Biochem Biophys* 2014;**70**:1849–58.
107. Li Y, Xu J, Chen H, *et al*. Comprehensive analysis of the functional microRNA-mRNA regulatory network identifies miRNA signatures associated with glioma malignant progression. *Nucleic Acids Res* 2013;**41**:e203.
108. Taurino C, Miller WH, McBride MW, *et al*. Gene expression profiling in whole blood of patients with coronary artery disease. *Clin Sci* 2010;**119**:335–43.
109. Williams AH, Valdez G, Moresi V, *et al*. MicroRNA-206 delays ALS progression and promotes regeneration of neuromuscular synapses in mice. *Science* 2009;**326**:1549–54.
110. Poy MN, Eliasson L, Krutzfeldt J, *et al*. A pancreatic islet-specific microRNA regulates insulin secretion. *Nature* 2004; **432**:226–30.
111. Yang B, Lin H, Xiao J, *et al*. The muscle-specific microRNA miR-1 regulates cardiac arrhythmogenic potential by targeting GJA1 and KCNJ2. *Nat Med* 2007;**13**:486–91.
112. Zhao Y, Ransom JF, Li A, *et al*. Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. *Cell* 2007;**129**:303–17.
113. Qiu C, Chen G, Cui Q. Towards the understanding of microRNA and environmental factor interactions and their relationships to human diseases. *Sci Rep* 2012;**2**:318.
114. Yang Q, Qiu C, Yang J, *et al*. miREnvironment database: providing a bridge for microRNAs, environmental factors and phenotypes. *Bioinformatics* 2011;**27**:3329–30.
115. Chen X, Liu MX, Cui QH, *et al*. Prediction of disease-related interactions between MicroRNAs and environmental factors based on a semi-supervised classifier. *PLoS One* 2012;**7**: e43425.
116. Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer* 2006;**6**:857–66.
117. Kulshreshtha R, Ferracin M, Wojcik SE, *et al*. A MicroRNA signature of hypoxia. *Mol Cell Biol* 2007;**27**:1859–67.
118. Calin GA, Ferracin M, Cimmino A, *et al*. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *New Engl J Med* 2005;**353**:1793–801.
119. Gottwein E, Mukherjee N, Sachse C, *et al*. A viral microRNA functions as an ortholog of cellular miR-155. *Nature* 2007; **450**:1096–9.
120. Cameron JE, Yin Q, Fewell C, *et al*. Epstein-Barr virus latent membrane protein 1 induces cellular MicroRNA miR-146a, a modulator of lymphocyte signaling pathways. *J Virol* 2008; **82**:1946–58.
121. Su X, Wang H, Ge W, *et al*. An *in vivo* method to identify microRNA targets not predicted by computation algorithms: p21 targeting by miR-92a in cancer. *Cancer Res* 2015;**75**: 2875–85.
122. Markou A, Tsaroucha EG, Kaklamanis L, *et al*. Prognostic value of mature microRNA-21 and microRNA-205 overexpression in non-small cell lung cancer by quantitative real-time RT-PCR. *Clin Chem* 2008;**54**:1696–704.
123. Bakhidze E, Arkhangelskaya P, Mikhet'ko A, *et al*. Analysis of expression of microRNA in cytological smears as a new method for the diagnosis and prognosis of preinvasive cervical carcinoma. *Eur J Cancer* 2017;**72**:S91–2.
124. Huang H, Roh J, Davis CD, *et al*. An improved method to quantitate mature plant microRNA in biological matrices using modified periodate treatment and inclusion of internal controls. *PLoS One* 2017;**12**:e0175429.
125. Shivapurkar N, Weiner LM, Marshall JL, *et al*. Recurrence of early stage colon cancer predicted by expression pattern of circulating microRNAs. *PLoS One* 2014;**9**:e84686.
126. Drusco A, Nuovo GJ, Zanesi N, *et al*. MicroRNA profiles discriminate among colon cancer metastasis. *PLoS One* 2014;**9**: e96670.
127. Parkin DM, Bray F, Ferlay J, *et al*. Global cancer statistics, 2002. *CA Cancer J Clin* 2005;**55**:74–108.
128. Krijger ID, Mekenkamp LJ, Punt CJ, *et al*. MicroRNAs in colorectal cancer metastasis. *J Pathol* 2011;**224**:438–47.
129. Vickers MM, Bar J, Gorn-Hondermann I, *et al*. Stage-dependent differential expression of microRNAs in colorectal cancer: potential role as markers of metastatic disease. *Clin Exp Metastasis* 2012;**29**:123–32.
130. Grose D, Morrison DS, Devereux G, *et al*. The impact of comorbidity upon determinants of outcome in patients with lung cancer. *Lung Cancer* 2015;**87**:186–92.
131. Cho WC, Chow AS, Au JS. MiR-145 inhibits cell proliferation of human lung adenocarcinoma by targeting EGFR and NUDT1. *RNA Biol* 2011;**8**:125–31.
132. Zhang H, Zhao M, Lv Z, *et al*. MiR-138 inhibits tumor growth through repression of EZH2 in non-small cell lung cancer. *Cell Physiol Biochem* 2013;**31**:56–65.
133. Kunsbaeva G, Gilyazova I, Pavlov V, *et al*. The role of miRNAs in the development of prostate cancer. *Russ J Genet* 2015;**51**: 627–41.
134. Attard G, Parker C, Eeles RA, *et al*. Prostate cancer. *Lancet* 2015;**387**:70–82.
135. Bonci D, Coppola V, Patrizii M, *et al*. A microRNA code for prostate cancer metastasis. *Oncogene* 2016;**35**:1180–92.
136. Ribas J, Ni X, Haffner M, *et al*. miR-21: an androgen receptor–regulated microRNA that promotes hormone-dependent and hormone-independent prostate cancer growth. *Cancer Res* 2009;**69**:7165–9.
137. Li Y, Kong D, Ahmad A, *et al*. Epigenetic deregulation of miR-29a and miR-1256 by isoflavone contributes to the inhibition of prostate cancer cell growth and invasion. *Epigenetics* 2012; **7**:940–9.
138. Majid S, Dar AA, Saini S, *et al*. miRNA-34b inhibits prostate cancer through demethylation, active chromatin modifications, and AKT pathways. *Clin Cancer Res* 2013;**19**:73–84.
139. Saini S, Arora S, Majid S, *et al*. MicroRNAs in the development and progression of kidney cancer. In: *MicroRNA in Development and in the Progression of Cancer*. New York: Springer, 2014, 213–41.
140. Gottardo F, Liu CG, Ferracin M, *et al*. Micro-RNA profiling in kidney and bladder cancers. *Urol Oncol* 2007;**25**:387–92.
141. Zhou S, Wang J, Zhang Z. An emerging understanding of long noncoding RNAs in kidney cancer. *J Cancer Res Clin Oncol* 2014;**140**:1989–95.

142. Wulfken LM, Moritz R, Ohlmann C, *et al*. MicroRNAs in renal cell carcinoma: diagnostic implications of serum miR-1233 levels. *PLoS One* 2011;**6**:e25787.
143. Redova M, Poprach A, Nekvindova J, *et al*. Circulating miR-378 and miR-451 in serum are potential biomarkers for renal cell carcinoma. *J Transl Med* 2012;**10**:55.
144. Polikepahad S, Knight JM, Naghavi AO, *et al*. Proinflammatory role for let-7 microRNAS in experimental asthma. *J Biol Chem* 2010;**285**:30139–49.
145. Mishra R, Singh SK. HIV-1 Tat C modulates expression of miRNA-101 to suppress VE-cadherin in human brain microvascular endothelial cells. *J Neurosci* 2013;**33**:5992–6000.
146. Foley NH, O'Neill LA. miR-107: a toll-like receptor-regulated miRNA dysregulated in obesity and type II diabetes. *J Leukoc Biol* 2012;**92**:521–7.
147. Jeong HC, Kim EK, Lee JH, *et al*. Aberrant expression of let-7a miRNA in the blood of non-small cell lung cancer patients. *Mol Med Rep* 2011;**4**:383–7.
148. Earle JS, Luthra R, Romans A, *et al*. Association of microRNA expression with microsatellite instability status in colorectal adenocarcinoma. *J Mol Diagn* 2010;**12**:433–40.
149. Muller DW, Bosserhoff AK. Integrin beta 3 expression is regulated by let-7a miRNA in malignant melanoma. *Oncogene* 2008;**27**:6698–706.
150. Quesne JL, Jones J, Warren J, *et al*. Biological and prognostic associations of miR-205 and let-7b in breast cancer revealed by in situ hybridization analysis of micro-RNA expression in arrays of archival tumour tissue. *J Pathol* 2012;**227**:306–14.
151. Di Fazio P, Montalbano R, Neureiter D, *et al*. Downregulation of HMGA2 by the pan-deacetylase inhibitor panobinostat is dependent on hsa-let-7b expression in liver cancer cell lines. *Exp Cell Res* 2012;**318**:1832–43.
152. Jakymiw A, Patel RS, Deming N, *et al*. Overexpression of dicer as a result of reduced let-7 MicroRNA levels contributes to increased cell proliferation of oral cancer cells. *Genes Chromosomes Cancer* 2010;**49**:549–59.
153. Chang SS, Jiang WW, Smith I, *et al*. MicroRNA alterations in head and neck squamous cell carcinoma. *Int J Cancer* 2008;**123**:2791–7.
154. Peng CH, Liu JH, Woung LC, *et al*. MicroRNAs and cataracts: correlation among let-7 expression, age and the severity of lens opacity. *Br J Ophthalmol* 2012;**96**:747–51.
155. Wei W, Hu Z, Fu H, *et al*. MicroRNA-1 and microRNA-499 downregulate the expression of the ets1 proto-oncogene in HepG2 cells. *Oncol Rep* 2012;**28**:701–6.
156. Kojima S, Chiyomaru T, Kawakami K, *et al*. Tumour suppressors miR-1 and miR-133a target the oncogenic function of purine nucleoside phosphorylase (PNP) in prostate cancer. *Br J Cancer* 2012;**106**:405–13.
157. Pavicic W, Perkio E, Kaur S, *et al*. Altered methylation at microRNA-associated CpG islands in hereditary and sporadic carcinomas: a methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA)-based approach. *Mol Med* 2011;**17**:726–35.
158. Datta J, Kutay H, Nasser MW, *et al*. Methylation mediated silencing of MicroRNA-1 gene and its role in hepatocellular carcinogenesis. *Cancer Res* 2008;**68**:5049–58.
159. Hudson RS, Yi M, Esposito D, *et al*. MicroRNA-1 is a candidate tumor suppressor and prognostic marker in human prostate cancer. *Nucleic Acids Res* 2012;**40**:3689–703.
160. Suzuki H, Takatsuka S, Akashi H, *et al*. Genome-wide profiling of chromatin signatures reveals epigenetic regulation of MicroRNA genes in colorectal cancer. *Cancer Res* 2011;**71**:5646–58.
161. Migliore C, Martin V, Leoni VP, *et al*. MiR-1 downregulation cooperates with MACC1 in promoting MET overexpression in human colon cancer. *Clin Cancer Res* 2012;**18**:737–47.
162. Esposito F, Tornincasa M, Pallante P, *et al*. Down-regulation of the miR-25 and miR-30d contributes to the development of anaplastic thyroid carcinoma targeting the polycomb protein EZH2. *J Clin Endocrinol Metab* 2012;**97**:E710–18.
163. Razumilava N, Bronk SF, Smoot RL, *et al*. miR-25 targets TNF-related apoptosis inducing ligand (TRAIL) death receptor-4 and promotes apoptosis resistance in cholangiocarcinoma. *Hepatology* 2012;**55**:465–75.
164. Li Q, Zou C, Zou C, *et al*. MicroRNA-25 functions as a potential tumor suppressor in colon cancer by targeting Smad7. *Cancer Lett* 2013;**335**:168–74.
165. Wei W, Hu Z, Fu H, *et al*. MicroRNA-1 and microRNA-499 downregulate the expression of the ets1 proto-oncogene in HepG2 cells. *Oncol Rep* 2012;**28**:701–6.
166. Adachi T, Nakanishi M, Otsuka Y, *et al*. Plasma microRNA 499 as a biomarker of acute myocardial infarction. *Clin Chem* 2010;**56**:1183–5.
167. Ma L, Li N, He X, *et al*. miR-449b and miR-34c on inducing down-regulation of cell cycle-related proteins and cycle arrests in SKOV3-ipl cell, an ovarian cancer cell line [in Chinese]. *Beijing Da Xue Xue Bao* 2011;**43**:129–33.
168. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* 2011;**39**:D152–7.
169. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 2014;**42**:D68–73.
170. Girijadevi R, Sreedevi VC, Sreedharan JV, *et al*. IntmiR: a complete catalogue of intronic miRNAs of human and mouse. *Bioinformation* 2011;**5**:458–9.
171. Chou CH, Chang NW, Shrestha S, *et al*. miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic Acids Res* 2016;**44**:D239–47.
172. Li Y, Qiu C, Tu J, *et al*. HMDD v2. 0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res* 2014;**42**:D1070–4.
173. Jiang Q, Wang Y, Hao Y, *et al*. miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Rese* 2009;**37**:D98–104.
174. Cho S, Jun Y, Lee S, *et al*. miRGator v2. 0: an integrated system for functional investigation of microRNAs. *Nucleic Acids Rese* 2011;**39**:D158–62.
175. Cho S, Jang I, Jun Y, *et al*. MiRGator v3. 0: a microRNA portal for deep sequencing, expression profiling and mRNA targeting. *Nucleic Acids Res* 2013;**41**:D252–7.
176. Alexiou P, Vergoulis T, Gleditzsch M, *et al*. miRGen 2.0: a database of microRNA genomic information and regulation. *Nucleic Acids Res* 2010;**38**:D137–41.
177. Georgakilas G, Vlachos IS, Zagganas K, *et al*. DIANA-miRGen v3. 0: accurate characterization of microRNA promoters and their regulators. *Nucleic Acids Res* 2016;**44**:D190–5.
178. Barh D, Bhat D, Viero C. miReg: a resource for microRNA regulation. *J Integr Bioinform* 2010;**7**:55–62.
179. Xiao F, Zuo Z, Cai G, *et al*. miRecords: an integrated resource for microRNA–target interactions. *Nucleic Acids Res* 2009;**37**:D105–10.
180. Dweep H, Sticht C, Pandey P, *et al*. miRWalk–database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. *J Biomed Inform* 2011;**44**:839–47.
181. Dweep H, Gretz N, Sticht C. miRWalk database for miRNA–target interactions. *Methods Mol Biol* 2014;**1182**:289–305.

182. Vergoulis T, Vlachos IS, Alexiou P, *et al*. TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Rese* 2012;**40**:D222–9.
183. Vlachos IS, Paraskevopoulou MD, Karagkouni D, *et al*. DIANA-TarBase v7. 0: indexing more than half a million experimentally supported miRNA: mRNA interactions. *Nucleic Acids Res* 2015;**43**:D153–9.
184. Hsu SD, Chu CH, Tsou AP, *et al*. miRNAMap 2.0: genomic maps of microRNAs in metazoan genomes. *Nucleic Acids Res* 2008;**36**:D165–9.
185. Betel D, Wilson M, Gabow A, *et al*. The microRNA. org resource: targets and expression. *Nucleic Acids Res* 2008;**36**:D149–53.
186. Backes C, Kehl T, Stockel D, *et al*. miRPathDB: a new dictionary on microRNAs and target pathways. 2017;**45**:D90–6.
187. Li J, Wu Z, Cheng F, *et al*. Computational prediction of microRNA networks incorporating environmental toxicity and disease etiology. *Sci Rep* 2014;**4**:5576.
188. Bhattacharya A, Cui Y. SomamiR 2.0: a database of cancer somatic mutations altering microRNA-ceRNA interactions. *Nucleic Acids Res* 2016;**44**:D1005–10.
189. Xie B, Ding Q, Han H, *et al*. miRCancer: a microRNA–cancer association database constructed by text mining on literature. *Bioinformatics* 2013;**29**:638–44.
190. Ulfenborg B, Jurcevic S, Lindlöf A, *et al*. miREC: a database of miRNAs involved in the development of endometrial cancer. *BMC Res Notes* 2015;**8**:104.
191. Yang Z, Ren F, Liu C, *et al*. dbDEMC: a database of differentially expressed miRNAs in human cancers. *BMC Genomics* 2010;**11**:S5.
192. Wang D, Gu J, Wang T, *et al*. OncomiRDB: a database for the experimentally verified oncogenic and tumor-suppressive microRNAs. *Bioinformatics* 2014;**30**:2237–8.
193. Khurana R, Verma VK, Rawoof A, *et al*. OncomiRdbB: a comprehensive database of microRNAs and their targets in breast cancer. *BMC Bioinformatics* 2014;**15**:15.
194. Zhang L, Ai H, Chen W, *et al*. CarcinoPred-EL: novel models for predicting the carcinogenicity of chemicals using molecular fingerprints and ensemble learning methods. *Sci Rep* 2017;**7**:2118.
195. Jiang Q, Hao Y, Wang G, *et al*. Prioritization of disease microRNAs through a human phenome-microRNAome network. *BMC Syst Biol* 2010;**4(Suppl 1)**:S2.
196. Chen X, Yan CC, Zhang X, *et al*. WBSMDA: within and between score for MiRNA-disease association prediction. *Sci Rep* 2016;**6**:21106.
197. Xuan P, Han K, Guo Y, *et al*. Prediction of potential disease-associated microRNAs based on random walk. *Bioinformatics* 2015;**31**:1805–15.
198. Sun D, Li A, Feng H, *et al*. NTSMDA: prediction of miRNA-disease associations by integrating network topological similarity. *Mol Biosyst* 2016;**12**:2224–32.
199. Chen X, Clarence YC, Zhang X, *et al*. HGIMDA: heterogeneous graph inference for miRNA-disease association prediction. *Oncotarget* 2016;**7**(40):65257–69.
200. Ding P, Luo J, Xiao Q, *et al*. A path-based measurement for human miRNA functional similarities using miRNA-disease associations. *Sci Rep* 2016;**6**:32533.
201. You ZH, Huang ZA. PBMDA: a novel and effective path-based computational model for miRNA-disease association prediction. 2017;**13**:e1005455.
202. Xu J, Li CX, Lv JY, *et al*. Prioritizing candidate disease miRNAs by topological features in the miRNA target–dysregulated network: case study of prostate cancer. *Mol Cancer Ther* 2011;**10**:1857–66.
203. Xuan P, Han K, Guo M, *et al*. Prediction of microRNAs associated with human diseases based on weighted k most similar neighbors. *PLoS One* 2013;**8**:e70204.
204. Chen X, Yan GY. Semi-supervised learning for potential human microRNA-disease associations inference. *Sci Rep* 2014;**4**:5501.
205. Pasquier C, Gardes J. Prediction of miRNA-disease associations with a vector space model. *Sci Rep* 2016;**6**:27036.
206. Li JQ, Rong ZH, Chen X, *et al*. MCMDA: matrix completion for MiRNA-disease association prediction. *Oncotarget* 2017;**8**:21187.
207. Chen X, Wu QF, Yan GY. RKNNMDA: ranking-based KNN for MiRNA-disease association prediction. *RNA Biol* 2017;**14**:952–62.
208. Luo J, Xiao Q, Liang C, *et al*. Predicting MicroRNA-disease associations using Kronecker regularized least squares based on heterogeneous omics data. *IEEE Access* 2017;**5**:2503–13.
209. Mørk S, Pletscher-Frankild S, Caro AP, *et al*. Protein-driven inference of miRNA–disease associations. *Bioinformatics* 2014;**30**:392–7.
210. Shi H, Xu J, Zhang G, *et al*. Walking the interactome to identify human miRNA-disease associations through the functional link between miRNA targets and disease genes. *BMC Syst Biol* 2013;**7**:101.
211. Lan W, Wang J, Li M, *et al*. Predicting microRNA-disease associations by integrating multiple biological information. In: IEEE International Conference on Bioinformatics and Biomedicine. IEEE, 2015, 183–8.
212. Weinberg MS, Wood MJ. Short non-coding RNA biology and neurodegenerative disorders: novel disease targets and therapeutics. *Hum Mol Genet* 2009;**18**:R27–39.
213. Chen X. miREFRWR: a novel disease-related microRNA-environmental factor interactions prediction method. *Mol Biosyst* 2016;**12**:624–33.