This is the peer reviewed author accepted manuscript (post print) version of a published work that appeared in final form in:

Le, Thuc Duy, Zhang, Junpeng, Liu, Lin, Li, Jiuyong 2017 'Computational methods for identifying miRNA sponge interactions', *Briefings in bioinformatics*, vol. 18, no. 4, pp. 577-590

This un-copyedited output may not exactly replicate the final published authoritative version for which the publisher owns copyright. It is not the copy of record. This output may be used for non-commercial purposes.

The final definitive published version (version of record) is available at:
https://doi.org/10.1093/bib/bbw042

Persistent link to the UniSA Research Outputs Repository:
https://researchoutputs.unisa.edu.au/11541.2/128647

General Rights
Copyright and moral rights for the publications made accessible in the UniSA Research Outputs Repository are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the persistent link identifying the publication in the UniSA Research Outputs Repository

If you believe that this document breaches copyright, please contact us and provide details. We will remove access to the work immediately and investigate your claim.

Computational methods for identifying miRNA sponge interactions

[1,*]Thuc Duy Le, [2]Junpeng Zhang, [1]Lin Liu and [1,*]Jiuyong Li

[1]School of Information Technology and Mathematical Sciences, University of South Australia, Mawson Lakes, SA 5095, Australia

[2]School of Engineering, Dali University, Dali, China

[*]Contact Author: Thuc.Le@unisa.edu.au and Jiuyong.Li@unisa.edu.au

Biographical note:

Thuc Duy Le is a research fellow at the University of South Australia (UniSA). He received his BSc (2002) and MSc (2006) in pure Mathematics from the University of Pedagogy, Ho Chi Minh City, Vietnam, and BSc (2010) in Computer Science from UniSA. He received his PhD degree in Computer Science (Bioinformatics) in 2014 at UniSA. His research interests are bioinformatics, data mining, and machine learning.

Junpeng Zhang is a teaching assistant at the School of Engineering, Dali University. He received his BSc (2009) in Bio-medical Engineering and MSc (2012) in Control Theory and Control Engineering from Kunming University of Science and Technology, Kunming City, China. His research interests include bioinformatics and data mining.

Lin Liu is a senior lecturer at the School of Information Technology and Mathematical Sciences, University of South Australia (UniSA). She received her bachelor and master degrees in Electronic Engineering from Xidian University, China in 1991 and 1994 respectively, and her PhD degree in computer systems engineering from UniSA in 2006. Her research interests include data mining and bioinformatics, as well as Petri nets and their applications to protocol verification and network security analysis.

Jiuyong Li is a professor at the School of Information Technology and Mathematical Sciences, University of South Australia. He received his PhD degree in computer science from the Griffith University, Australia (2002). His research interests are in the fields of data mining, privacy preserving and bioinformatics. His research has been supported by five prestigious Australian Research Council Discovery grants since 2005 and he has published more than 100 research papers.

Key points:

- There are coding and non-coding RNAs, known as competing endogenous RNAs (ceRNAs) or miRNA sponges, which compete with each other to interact with miRNAs.
- miRNA sponges play important roles in cancer progression.
- Computational methods have been shown to be effective tools to identify miRNA sponge interactions. Results of different computational methods are complementary.
- There is still a lack of ground truth for validating the computational predictions.
- There are great opportunities and challenges in designing reliable methods for inferring miRNA sponge regulatory mechanisms.

Keywords

miRNA, miRNA sponge interaction, ceRNA, ceRNA-ceRNA interaction, cancer, computational method.

Abstract

Recent studies on competing endogenous RNAs (ceRNAs) have shifted our understanding of how microRNAs (miRNAs) interact with their target RNAs. ceRNAs are RNAs that share miRNA recognition elements (MREs) and indirectly regulate each other via the titration mechanism, i.e. the increasing concentration of a ceRNA will decrease the number of miRNAs that are available for interacting with other targets. The cross-talks between ceRNAs, i.e. their interactions mediated by miRNAs, have been identified as the drivers in many disease conditions, including cancers. In recent years, some computational methods have emerged for identifying ceRNA-ceRNA interactions. However, there remain great challenges and opportunities for developing computational methods to provide new insights into ceRNA regulatory mechanisms.

In this paper, we review the publically available databases of ceRNA-ceRNA interactions and the computational methods for identifying ceRNA-ceRNA interactions (also known as miRNA sponge interactions). We also conduct a comparison study of the methods with a breast cancer dataset. Our aim is to provide a current snapshot of the advances of the computational methods in identifying miRNA sponge interactions and to discuss the remaining challenges.

INTRODUCTION

MicroRNAs (miRNAs) are small non-coding RNAs that regulate gene expression at the post-transcriptional level. They control a wide range of biological processes, including development, metabolism, differentiation, apoptosis and cellular signaling [1-5]. miRNAs are involved in many human diseases including several types of cancers [6-8]. Therefore,

understanding miRNA functions and their regulatory mechanisms is important for diagnostics and therapeutics.

Recent studies have shifted our understanding of how miRNAs interact with their target RNAs. As the common knowledge, miRNAs recognise their targets by base pairing to complementary sequences in the 3'-untranslated region (3' UTR), 5' UTR, and sometimes in the open reading frames of the targets. However, recent findings show that coding genes are not the only targets that miRNAs interact with. In fact, there is a pool of different RNAs competing with each other to attract miRNAs for interactions, thus acting as competing endogenous RNAs (ceRNAs). These ceRNAs are also called miRNA "decoys" or miRNA "sponges", which can be long non-coding RNAs (lncRNAs), pseudogenes, or circular RNAs (circRNAs) [9-12]. Through base pairing with miRNA binding sites, i.e. miRNA recognition/response elements (MREs), miRNA sponges interact with miRNAs and reduce the amount of miRNA transcripts in a cell and therefore release miRNA target genes from miRNA control [13]. Figure 1 illustrates the role of miRNA sponges in controlling the regulation of miRNA targets.

In recent years, more and more miRNA sponges have been identified by experiments and the roles of miRNA sponges in different conditions are being uncovered. Table 1 shows a part of the experimentally confirmed lncRNAs, pseudogenes, circRNAs, and coding RNAs acting as miRNA sponges in specific conditions.

The first type of miRNA sponge is lncRNA. lncRNAs are non-protein coding transcripts longer than 200 nucleotides, and they originate from thousands of loci across genomes. According to their proximity to protein-coding genes in the genome, lncRNAs are mainly classified into three classes: (1) intergenic lncRNA without any overlap with other loci, (2) long intronic lncRNA overlapping with the intronic region of any other loci, and (3) antisense

exonic lncRNA overlapping with exons of other loci on the opposite strand [29]. The first evidence of lncRNAs being miRNA sponges was demonstrated in muscle differentiation [9], where a muscle-specific lncRNA called linc-MD1 was identified to govern the time of muscle differentiation by acting as a ceRNA in mouse and human myoblasts. linc-MD1 sequestered miR-133 and miR-135 to regulate the expression of MAML1 and MEF2C, respectively. In human embryonic stem cell (ESC) self-renewal, linc-RoR has been shown to function as a miRNA sponge and plays an essential role in ESC [14]. linc-RoR prevents OCT4, SOX2, NANOG from miR-145 mediated suppression in self-renewing human ESC. linc-RoR also forms a feedback loop with OCT4, SOX2, NANOG and miR-145 to regulate ESC maintenance and differentiation. In addition to their crucial roles in tissue development, the importance of lncRNAs acting as miRNA sponges has also been implicated in various cancers, such as liver cancer [15] and thyroid cancer [16].

Pseudogenes have also been found to act as miRNA sponges. Pseudogenes are dysfunctional relatives of the genes that have lost their protein-coding ability due to various genetic disablements, including premature stop codes, frameshifts, insertions, or deletion [30]. Poliseno *et al.* [10] demonstrated that the pseudogene PTENP1 regulated the level of its cognate gene PTEN by competing with PTEN for shared miRNAs. They also showed that PTENP1 acted as a sponge of PTEN-targeting miRNAs to exert a growth-suppressive role and to be selectively lost in human cancers. In addition, it was found in [10] that the pseudogene KRAS1P 3’UTR overexpression resulted in increased KRAS mRNA abundance and accelerated cell growth. The KRAS and KRAS1P transcript levels are positively correlated in prostate cancer, and the KRAS1P locus is amplified in different human tumours, including neuroblastoma, retinoblastoma and hepatocellular carcinoma. Taken together, the functions of KRAS1P mirror the functions of its cognate gene KRAS by a miRNA sponge mechanism.

circRNAs are another type of non-coding ceRNAs, and they form a covalently closed continuous loop without 5' to 3' polarity or polyadenylated tail [31]. Many circular RNAs arise from otherwise protein-coding genes, but they have not been shown to code for proteins in the cell. Recent studies show that natural circRNAs can also function as efficient miRNA sponges. An example of a circRNA is CDR1as (also known as ciRS-7) [11, 12], which has been shown to be miR-7 sponge to affect the activity of miR-7 target genes. Another example is the testis-specific circRNA Sry, which has been validated as a miR-138 sponge [11].

Recent studies have shown that several protein-coding RNAs act as miRNA sponges [32]. The most extensively characterised coding RNA as miRNA sponge is the tumour suppressor gene PTEN. A combined computational and experimental approach [20] has identified and validated CNOT6L and VAPA as endogenous protein-coding transcripts that regulate PTEN, antagonise PI3K/AKT signalling, and possess growth and tumour suppressive properties. Notably, they have also shown that CNOT6L and VAPA display concordant expression patterns and copy number loss with PTEN in cancers.

Given the important role of miRNA sponges, they have attracted growing attentions from researchers. As shown in Figure 2, the number of publications on miRNA sponge research has been increased gradually in the past few years. At the same time, the recent years have seen the emergence of computational methods in discovering ceRNA-ceRNA interactions (also known as miRNA sponge interactions or miRNA mediated interactions). Although the number of publications on computational methods is still small, as with other miRNA related research, computational methods will undoubtedly contribute to the research of miRNA sponge interactions, as an effective technique to generate hypotheses for further investigations. Therefore in this paper we present a review of these computational methods. Our aim is to provide a current snapshot of the computational methods in identifying miRNA

sponge interactions and to discuss the challenges, to help with future development of such methods.

We will firstly review the computational methods for identifying miRNA sponge interactions. Secondly, we review the publically available databases of miRNA sponge interactions. Apart from miRNA sponge interaction databases, we also list the commonly used miRNA-mRNA binding databases as the initial structure of computational methods to identify candidate miRNA sponge interactions. They contain both experimentally confirmed and predicted interactions between miRNAs and their targets, including lncRNAs, pseudogenes, cirRNAs and mRNAs. Finally, we conduct a case study with a breast cancer dataset to compare four well-known computational methods for identifying miRNA sponge interactions. We discuss some challenges of developing computational methods for identifying miRNA sponge interactions in the last section.

COMPUTATIONAL APPROACHES FOR IDENTIFYING miRNA SPONGE INTERACTIONS

In this section, we review the computational methods for identifying ceRNA-ceRNA interactions or miRNA sponge interactions. As shown in Table 2, we categorise the methods into i) pair-wise correlation approach, ii) partial association approach, and iii) mathematical modelling approach. In the following, we will review the methods arranged by their categories. Moreover, we also conduct a review on the publically available databases for validating and constructing miRNA sponge interactions.

Pair-wise correlation approach

The common principle of the methods in this category is that the expression levels of an interacting miRNA sponge interaction pairs are positively correlated or the correlation has a significant change in two different conditions. The positive correlation principle is based on the observation of the titration mechanism [42, 43] which states that a high concentration of a miRNA sponge will decrease the available miRNAs for interacting with other miRNA sponges. For example, suppose RNA_1 and RNA_2 share a miRNA and the miRNA up-regulates RNA_1. If the miRNA preferentially interacts with RNA_1 (the expression level of RNA_1 is thus increased), then RNA_2 is not suppressed (as there is less miRNA to interact with RNA_2). As a result, the expression level of RNA_2 is increased, and the expression levels of RNA_1 and RNA_2 show a positive correlation. As for the principle of significant change of the correlation between a miRNA sponge interaction pair in two different conditions, it is based on the dynamic and condition-specific properties of miRNA sponge regulation [39, 40].

This group of methods firstly search for all pairs of RNAs that share the miRNA recognition elements (MREs) of a set of miRNAs using putative target prediction databases. For a given RNA pair (RNA_1 and RNA_2), a statistical test, such as the hypergeometric test, is used to calculate the significance of sharing of miRNAs by these two RNAs. The significance *p*-value is computed as follows.

$$p = 1 - \sum_{i=0}^{x-1} \frac{\binom{K}{i}\binom{N-K}{O-i}}{\binom{N}{O}} \quad (1)$$

In the formula, N is the number of all miRNAs in the dataset of interest, K denotes the total number of miRNAs regulating RNA_1, O denotes the total number of miRNAs regulating RNA_2, and x is the number of miRNAs shared by RNA_1 and RNA_2. The correlation between each pair of miRNA sponge interaction which has significant sharing of miRNAs is then

calculated, and the pairs with significant positive correlations or significant difference of correlations in two different conditions are predicted to have miRNA sponge interactions.

Zhou *et al.* [33] combined matched breast cancer miRNA and mRNA expression data with miRNA-mRNA relationships to identify breast cancer specific miRNA sponge interactions. They firstly found the shared miRNAs for all mRNA-mRNA pairs using predicted miRNA target information and hypergeometric test. They then tested the significance of the correlations between the potential miRNA sponge interaction pairs sharing the same miRNAs using Pearson correlation coefficients. The miRNA sponge network is constructed using the identified significant miRNA sponge interaction pairs. The network was shown to be scale free. It was also found that in the network the hub nodes were closely associated with breast cancer and the dense clusters were related to cancer hallmarks. According to the results of survival analysis, the discriminative hub genes in the miRNA sponge network showed discriminative power and were regarded as good biomarker candidates in breast cancer.

Xu *et al.* [34] used the similar method as in Zhou *et al.* [33] to infer the mRNA related miRNA sponge interaction landscape across 20 major cancer types. They found that ~34.56% miRNA sponge interactions occurred only in one cancer, and only 3.78% miRNA sponge interactions were conserved in more than 10 cancers in the pan-cancer cohort. Although most miRNA sponge interactions are cancer-specific, the cancers within similar origin cell type share significant number of common miRNA sponge interactions. Functional analysis revealed conserved and rewired network miRNA sponge hubs in each cancer, and these miRNA sponge hubs formed tense miRNA sponge interactions to constitute conserved and cancer-specific modules.

Shao *et al.* [35] also proposed a pair-wise correlation based method to identify dysregulated miRNA sponge interactions in lung adenocarcinoma using expression profiles and miRNA-

target interactions. They divided the dysregulated miRNA sponge interactions into two types: gain and loss. They found that the correlations of gain miRNA sponge interactions are significantly positive in cancer condition, and in normal condition the correlations of loss miRNA sponge interactions are significantly positive. It was also found that gain miRNA sponge pairs have consistent expression in cancer, but it is not necessary the case for loss miRNA sponge pairs. Furthermore, gain and loss miRNA sponges involving in gain and loss miRNA sponge interaction networks as topological key nodes are implicated in the development of cancer. Additionally, several dyregulated miRNA sponge modules could be regarded as diagnostic biomarkers even in other three independent lung datasets. In the field of biomarkers for diagnosis, the study suggests the possibility of a new biological mechanism in cancer.

To study optimal conditions for miRNA sponge regulation, Chiu *et al.* [36] integrated expression profiles and putative miRNA-target interactions to identify optimal miRNA sponge interactions in TCGA datasets of GBM (glioblastoma multiforme), OV (ovarian serous cystadenocarcinoma), LUSC (lung squamous cell carcinoma), and LAML (large acute myeloid leukemia). They demonstrated that miRNA sponge regulation was positively associated with the number of common miRNAs and the target binding sites, and was dynamic across cancer types. Their study brings biological insights into the understanding of the dynamicity and essential roles of miRNA sponge regulation in cancer.

Partial association approach

Different from the pair-wise correlation methods, partial association methods take the miRNA expression levels into consideration when examining the miRNA sponge relationships. When the relationships between miRNA sponges are assumed linear, the partial correlation methods [13] measure the correlation between two miRNA sponges given the

presence of a miRNA. When no assumption is made on the type of relationships (i.e. linear or non-linear) between miRNA sponges, the conditional mutual information approaches [21, 37] can be used instead of the partial correlation.

Sumazin *et al.* [21] investigated the ability of coding and non-coding RNAs to act as miRNA sponges by analysing a large set of matched miRNA and gene expression profiles in glioblastoma. They identified an extensive miRNA-mediated network of RNA-RNA interactions, and uncovered miRNA-mediated crosstalk between different oncogenic pathways. According to distinct molecular mechanisms, they studied two types of miRNA-mediated modulators: sponge modulators and non-sponge modulators. Due to the prohibitive computational cost of evaluating a huge number of (RNA, miRNA, RNA) triplets, they used experimentally validated or predicted miRNA-target information to reduce the number of (RNA, miRNA, RNA) triplets to be evaluated. For sponge modulators, they only considered RNA-RNA pairs sharing a significant overlap of common miRNAs, and combined expression data with predicted miRNA-target regulatory interactions to identify (RNA, miRNA, RNA) triplets. For non-sponge modulators, they considered all RNA-RNA pairs including empty and non-statistically significant overlap of common miRNAs, and integrated expression data with experimentally validated miRNA-target information to infer (RNA, miRNA, RNA) triplets. The proposed method made use of the mutual information (MI) and conditional mutual information (CMI) and identified nearly 7,000 sponge modulators as miRNA sponges comprising more than 248,000 pairs of miRNA-mediated RNA-RNA interactions, and 148 non-sponge modulators participating in 169 miRNA-mediated RNA-RNA interactions with more than 100 genes. This extensive miRNA-mediated network provides new insights to the dysregulation of key mechanisms of pathogenesis as well as normal physiology mechanisms.

Paci and co-authors [13] proposed a method based on partial correlation to explore the role of lncRNAs as miRNA sponges. They applied the method to the matched miRNA and gene expression profiles of 72 breast invasive carcinoma (BRCA) individuals obtained from The Cancer Genome Atlas (TCGA). The samples of tumour and normal tissue were taken from the same patients. By computing for each (lncRNA, miRNA, RNA) triplet the difference between the Pearson correlation and partial correlation (defined as *sensitivity correlation*), they built the miRNA-mediated interaction (MMI) networks, in tumour and normal conditions (tumour-MMI-network and normal-MMI-network) respectively. The tumour-MMI-network appears to be completely different from the normal-MMI-network, implying that miRNA decoy mechanism may switch on or off its agents according to the physiological or pathological condition.

Chui *et al.* [37] introduced an integrative framework called Cupid to simultaneously reconstruct miRNA-target and miRNA sponge interaction networks by integrating sequence-based evidence and functional clues derived from miRNA and RNA expression analysis. Cupid is a 3-step prediction method. Firstly, it applies Fisher's exact test [44] to estimate candidate miRNA sponge interaction pairs. Secondly, similar to [21], it also uses MI and CMI to identify miRNA sponge interactions. Finally, it uses Brown's method [45] to further infer significant miRNA sponge interactions. By using high-throughput data in breast cancer cell lines for validation, they reported about 500 validated miRNA-target interactions involved in breast cancer. To evaluate the miRNA targets which compete for miRNA regulation, they focused on five breast-cancer associated regulators, CCND1, ESR1, HIF1A, PDGFRA and NCOA3. As a result, eight out of 11 of the predicted miRNA sponge interactions showed significant up-regulation after transfection of the 3'UTR of the miRNAs shared by the interacting miRNA sponges. The results support miRNA sponge regulation, and are consistent with the results in [21].

Mathematical modelling approach

The minimal model [38] and stochastic model [39], although are different methods for modelling competing endogenous RNA networks consisting of miRNAs and mRNAs, reported consistent results that the binding free energies and repression mechanism are the key ingredients for the cross-talk among the miRNA sponges. In addition, their work found that statistically significant correlations among miRNA sponges can exist due to transcriptional fluctuations even when miRNA-mediated cross-talks are knocked out.

Ala *et al.* [40] presented a mass-action model (similar to the minimal model in [38]) to determine the optimal conditions for miRNA sponge activities in silico. The model describes the molecular transition mechanism for the miRNA sponge interactions mediated by a common set of miRNAs. They refined the generated network by removing all edges that were not found in the conserved co-expression networks built by Piro *et al.* [46]. The evaluation of the complexity of the miRNA sponge interaction networks revealed that transcription factor and miRNA sponge interaction networks were cross-regulated, and the miRNA sponge interaction networks could cause up-regulation or aberrant expression of transcription factor in cancer. Furthermore, the results suggest that alterations of one miRNA sponge have significant effects on the integrated transcriptional and miRNA sponge interaction networks.

Recently, Yuan *et al.* [41] established a coarse-grained model for a minimum miRNA sponge interaction system consisting of one miRNA and two miRNA sponges, and validated their computational results by using synthetic gene circuits in cultured human embryonic kidney 293 (HEK293) cells. They demonstrated that the range and the strength of miRNA sponge effect were affected by the abundance and the binding strength of miRNAs and miRNA sponges, as well as the degree of complementarities between miRNA and miRNA sponge. Furthermore, they found a nonreciprocal competing effect between partial and

complementary targets due to different miRNA loss rates in these two types of regulations. The integrated results of computational model and experimental validation shed light on a quantitative understanding of miRNA-mediated cross-regulation on miRNA sponges.

Resources for evaluating and constructing miRNA sponge interactions

As a resource for evaluating and comparing the results of different computational methods, in this section, we review the commonly used databases of miRNA sponge interactions. A summary of the databases is given in Table 3.

ceRDB [47] is one of the first miRNA sponge interaction databases. They use TargetScan [48] to examine the co-occurrence of miRNA response elements (binding sites) in mRNAs on a genome-wide basis. Given a specific mRNA, ceRDB predicts the miRNA sponges that compete with the mRNA for the binding sites of the same miRNAs. For convenience, ceRDB provides a web-based tool called CEFINDER to allow users to identify miRNA sponges for a given target mRNA based on the homology of miRNA binding sites present in the 3'UTR sequence. ceRDB is available at http://www.oncomir.umn.edu/cefinder/. Although the database provides putative miRNA sponge interactions, the interactions may not be very reliable since the database depends on TargetScan v5.2, which is outdated.

starBase [49] stores 108 datasets obtained from 37 studies, including 47 HITS-CLIP datasets, 51 PAR-CLIP datasets, 9 ICLIP datasets and 1 CLASH dataset. Recently, starBase is also regarded as a miRNA sponge interaction database. The experiment results stored in starBase provide physical biding information between miRNA-mRNA, miRNA-lncRNA, miRNA-circRNA, miRNA-pseudogene, and miRNA-sncRNA. The interactions can be used to verify computational predictions. starBase is available at http://starbase.sysu.edu.cn/.

As for lncRNAs, lnCeDB [50] contains human lncRNAs (obtained from GENCODE v19 [51]) which may behave as miRNA sponges. The miRNA-mRNA interactions in lnCeDB are predicted using TargetScan [48] and mapped to the predictions obtained based on the AGO PAR-CLIP data in starBase [49]. The miRNA-lncRNAs interactions in lnCeDB are either retrieved from miRcode [52], or predicated by lnCeDB's own algorithm. lnCeDB allows users to browse for lncRNA-mRNA pairs sharing the same miRNAs. At the same times, users can compare the expression of a lncRNA-mRNA pair in 22 human tissues to find out the possibility for the pair to be miRNA sponges. lnCeDB is available at http://gyanxet-beta.com/lncedb/.

Wang *et al*. [53] developed a lncRNA-associated competing triplets database called LncACTdb (http://www.bio-bigdata.net/LncACTdb/) to facilitate the study of lncRNA function as miRNA sponges. In LncACTdb, competition based lncRNA-miRNA-gene interactions are defined as lncACTs. The database contains 5119 functionally activated and over 530,000 computationally predicted lncACTs, which are obtained by integrating heterogeneous data from several in silico target prediction studies, Argonaute-CLIP experiments, and expression profiles from transcriptome sequencing.

Ghosal *et al.* [54] developed a comprehensive database called HumanViCe (http://gyanxet-beta.com/humanvice) to enable users to explore potential miRNA sponges in virus infected cells. HumanViCe contains a vast number of coding and non-coding RNAs acting as potential miRNA sponges in virus infected human cells, which potentially can be used in constructing miRNA sponge interaction networks in virus infected human cells.

Wang *et al*. [55] manually developed the first experimentally validated miRNA sponge interaction database called miRSponge. miRSponge (http://www.bio-bigdata.net/miRSponge/) collected 185 unique miRNA sponge interactions in 11 species by manually curating from

nearly 1000 articles. The types of miRNA sponges include lncRNAs, pseudogenes, circRNAs, coding RNAs, viral RNAs and artificial engineered sponges. The database will provide a good ground truth for validating predicted miRNA sponge interactions, and we will use this for evaluating different methods in our case study in the later section.

Apart from the miRNA sponge interaction databases, the miRNA-mRNA binding databases are also important resources used by computational methods to construct candidate miRNA sponge interactions since protein-coding RNA transcripts can act as miRNA sponges. miRNA-target binding databases can be divided into two types: those contain experimentally validated target bindings and those with predicted miRNA-mRNA interactions. Experimentally validated miRNA-mRNA binding databases mainly include miRTarBase [56], Tarbase [57], miRecords [58] and miRWalk [59]. In addition to starBase [49], the commonly used predicted miRNA-mRNA binding databases are TargetScan [48], miRanda [60], MicroCosm [61], PicTar [62], DIANA-microT [63], PITA [64] and CLASH [65]. Users can choose independent databases or combine multiple databases as initial structure of computational methods to identify candidate miRNA sponge interactions.

A COMPARISON STUDY WITH A BREAST CANCER DATASET

In this section, we conduct a case study to compare the relative performance of four representative methods mentioned in the above section using the breast cancer dataset from Paci *et al.* [13].

Dataset

The dataset contains the expression profiles of 72 tumors with matched normal samples for miRNAs, lncRNAs, and mRNAs. A miRNA, lncRNA or mRNA which has missing values in more than 50% of the samples is removed from the data set. All the remaining missing values

are imputed using the *k*-nearest neighbours (KNN) algorithm as implemented in the R-package *impute*. As a result, the dataset contains 453 miRNAs, 1084 lncRNAs and 11217 mRNAs in72 matched tumour and normal tissues.

We perform differential expression analysis on the expression profiles using the *limma* [66] package of Bioconductor to identify differentially expressed miRNAs, lncRNAs and mRNAs between tumour and normal tissues in the dataset. After the analysis, 278 miRNAs (adjusted *p*-value < 0.01, adjusted by Benjamini & Hochberg method), 404 lncRNAs (adjusted *p*-value <10E-04) and 5602 mRNAs (adjusted *p*-value <10E-04) are identified to be differentially expressed at significant level.

Computational methods for identifying miRNA sponge interactions

Since the computational methods in the third category (mathematical modelling approach) only quantitatively describe a minimum or small miRNA sponge interaction network, it is only applied to study a small number of miRNA sponge interactions. Therefore, in this study, we only compare the performance of four well-used computational methods in the first two categories (pair-wise correlation and partial association approach) for identifying miRNA sponge interactions. The methods are Positive Correlation [33, 34], Conditional Mutual Information [21], Partial Pearson Correlation, and Sensitivity Partial Pearson Correlation [13]. The four methods have two main steps, generating the candidate miRNA sponge interactions and evaluating the miRNA sponge interactions. We describe these steps in the following.

Generating candidate miRNA sponge interactions

The four computational methods have the same first step of generating the candidates of miRNA sponge interactions. They use putative miRNA target information and

hypergeometric test to identify candidate miRNA sponges which are expected to have a significant number of shared miRNAs.

In our study, we firstly use the putative target binding information in starBase v2.0 [49] to generate putative miRNA-mRNA and miRNA-lncRNA relationships among the differentially expressed miRNAs, lncRNAs and mRNAs in the dataset. We set the minimum number of shared miRNAs to two. Then the hypergeometric test is conducted to evaluate the significance of the shared miRNAs for each possible gene pairs (lncRNA-lncRNA, lncRNA-mRNA, and mRNA-mRNA). Those gene pairs with p-value < 0.05 are selected as candidate miRNA sponge interaction pairs.

Evaluating miRNA sponge interactions

In this step, the methods use different strategies to select significant miRNA sponge interactions.

The first method, Positive Correlation (PC) [33, 34], is based on the positive correlation between the expression levels of a pair of miRNA sponges. Applying the method, we evaluate the candidate miRNA sponge interactions by computing the Pearson correlation coefficients of a pair of miRNA sponges using their expression data. The positively correlated miRNA sponge interaction pairs with p-value<0.05 are regarded as significant miRNA sponge interaction pairs.

The second method, Conditional Mutual Information (CMI), is based on conditional mutual information to estimate the effect of a miRNA sponge on its competing miRNA sponges [21]. The method evaluates candidate miRNA sponge interaction pairs by estimating the statistical significance of the difference between the conditional mutual information between the expression levels of a miRNA and its target (conditional on the expression level of the miRNA sponge) and the (unconditional) mutual information of the miRNA and target

expression levels. Specifically, given a miRNA sponge (*C*), a miRNA (*M*), and a target of the miRNA (*T*), the method estimate the statistical significance of $\Delta = I[M;T|C] - I[M;T]$ from matched miRNA and gene expression profiles using a variety of CMI estimators. The miRNA sponge interaction pairs with *p*-value<0.05 are regarded as significant miRNA sponge interaction pairs.

The third method, Partial Pearson Correlation (PPC), is similar to the second method [21] but it uses PPC rather than CMI to estimate the effect of a miRNA sponge on its competing miRNA sponges. The method evaluates candidate miRNA sponge interaction pairs by estimating the statistical significance of the difference between the *PPC* of the miRNA expression level and the target expression level (conditional on the expression level of miRNA sponge) and the Pearson correlation (*PC*) of the expression levels of the miRNA and target. Specifically, given a miRNA sponge (*C*), a miRNA (*M*), and a regulated target (*T*), The statistical significance of the difference, $\Delta = PPC[M;T|C] - PC[M;T]$ can be estimated from matched miRNA and gene expression profiles using a variety of *PPC* estimators. The miRNA sponge interaction pairs with the statistical significance *p*-value<0.05 are regarded as significant miRNA sponge interaction pairs.

The last method, Sensitivity Partial Pearson Correlation (SPPC), is based on the PPC to estimate the contributed effect (sensitivity) of the miRNA in a miRNA sponge-miRNA-target triplet [13]. Applying the method, we evaluate candidate miRNA sponge interaction pairs by estimating the sensitivity correlation (*S*) between the partial Pearson correlation (*PPC*) of miRNA and target expression levels (conditional on the expression level of miRNA sponge) and the Pearson correlation (*PC*) of miRNA and target expression levels. Specifically, given a miRNA sponge (*C*), a miRNA (*M*), and a regulated target (*T*), the sensitivity correlation

between *PPC[T;C|M]* and *PC[T;C]* is estimated. The miRNA sponge interaction pairs with $S>0.3$ are regarded as significant miRNA sponge interaction pairs.

Comparison results

The numbers of significant miRNA sponge interactions identified by the four methods are very different. In total, the four methods identified 1274708, 509504, 702029, and 99662 interactions respectively. Moreover, as shown in Figure 3, the numbers of various types of miRNA sponge interactions (lncRNA-lncRNA, lncRNA-mRNA and mRNA-mRNA) identified by the four computational methods are also very different. There are a significant number of miRNA sponge interactions that are inferred by a single method only, and only a small fraction of miRNA sponge interactions are shared by the four methods. This implies that different methods may have their own merits, but they may also have a high rate of false discoveries. The first potential reason of high false discoveries is that some miRNA sponges are false or indirect targets of miRNAs, and they are also considered to identify miRNA sponge interactions. Another potential reason is that the correlations of some identified miRNA sponge interactions are sensitive to the number of samples, and they randomly show high correlations. Nevertheless, it is noted that the common miRNA sponge interactions predicted by all the four methods may be regarded as high-confidence miRNA sponge interactions.

Since the study of miRNA sponge interaction networks is still at the exploratory stage and the ground truth for validating computational predictions is limited. We combine the experimentally validated miRNA sponge interactions in Table 1 with the newly emerged database, miRSponge [55], and compare the ability of the methods in recovering these confirmed interactions. The last column in Table 4 shows experimentally confirmed interactions identified by each method. As we can see from Table 4, the PC method identified

the largest number of confirmed interactions, with four interactions related to PTEN. A possible reason is that the PC method is based on positive correlations and it has identified the largest number of interactions (1274708). Interestingly, the CMI method identified three confirmed interactions, but the PPC method could not find any confirmed interactions. As PPC assumes linear relationships between variables, the results suggest that the relationships between variables may not be linear.

We conducted the enrichment analysis to explore the number and percentage of the miRNA sponge interactions associated with breast cancer identified by each method. We define breast cancer related miRNA sponge interactions as those in which the two interacting miRNA sponges are breast cancer related genes. To obtain a list of breast cancer related miRNA sponges, we create a list of genes and lncRNAs associated with breast cancer from a wide range of sources of experimentally validated databases. The list of breast cancer genes is collected from COSMIC v74 [67], GAD [68], OMIM [69], BCGD [70] and G2SBC [71]. The breast cancer lncRNAs are obtained from LncRNADisease [72] and Lnc2Cancer [73].

As shown in Table 4, the numbers of breast cancer related miRNA sponge interactions discovered by the four methods respectively are not strictly correlated with the total numbers of the interactions identified by the four methods respectively. PC identifies the largest number of breast cancer related miRNA sponge interactions but the lowest percentage of breast cancer related miRNA sponge interactions. SPPC is the best method in terms of the percentage of discovered miRNA sponge interactions associated with breast cancer.

The results of the comparison study indicate that the methods should be used complementarily.

DISCUSSIONS AND CONCLUSION

Identifying miRNA sponge interactions with computational methods is an emerging research area. miRNA sponge interactions have been found to be involved in many diseases including cancer. Their cross-talk is changing our understanding and approaches to elucidating disease mechanisms. In this paper, we have reviewed the recent computational approaches for identifying miRNA sponge interactions.

As the research into miRNA sponge interactions has just emerged, we do not have the ground truth to validate the computational predictions. Our evaluations in the case study are simply the enrichment analysis, and they are not sufficient to conclude about the performance of the computational methods. Comprehensive databases for experimentally confirmed miRNA sponge interactions are hence in demand.

Our case study of the methods with the breast cancer dataset shows that the methods generate different sets of prediction results, suggesting that the methods may have their own merits but may also involve a high rate of false discoveries, and they should be used complementarily such as using an ensemble method [74].

Currently the relationships between miRNA sponge interactions and diseases are still unclear. Although the roles and functions of miRNAs in disease development have been well studied, insights into the roles of miRNA sponges in diseases are needed. Similar to miRNA research, the computational methods would help generate hypotheses for wet-lab experiments to elucidate the roles of miRNA sponges in a particular disease.

One of the challenges for computational methods of studying miRNA sponge interactions is that the number of candidate miRNA sponges in a dataset is large. Different from miRNA target research which only considers a set of possible targets for the miRNAs, we need to

consider all types of miRNA sponges (lncRNAs, pseudogenes, cirRNAs, coding genes, etc.) competing for the same miRNAs when studying miRNA sponge interactions. The reason is that all types of miRNA sponges could become potential competitors of target genes for the miRNAs. Therefore, a good computational method should include different types of miRNA sponges in the model. On the other hand, this leads to the high computation complexity of the model. A possible solution is that we filter non-competitive miRNA sponges, e.g. ones with low expression level, before inferring their regulatory relationships. Another alternative solution is to study the miRNA sponge interaction for a specific disease by pre-selecting the miRNA sponges associated with the disease based on public databases.

Moreover, differential gene expression analysis that has been used frequently in miRNA research may not be suitable for identifying miRNA sponge interactions. The reason is that the RNAs that are not differentially expressed may contribute to a number of MREs that will affect the amount of the miRNAs available for bindings. Therefore, only using differentially expressed RNAs would affect the accuracy of the computational methods, but using all RNAs would significantly increase the computational complexity of the methods. Given this dilemma, it is necessary to find a trade-off between the accuracy and the computational complexity of the computational methods.

Another challenge of miRNA sponge computational methods is that the number of samples in matched expression profiles is usually small, especially when we include multiple types of RNAs. Small sample size does not support the statistical significance of the findings. Similar to miRNA research, we may use bootstrapping or permutation methods to address this problem. Furthermore, existing methods use the association/correlation between miRNA sponges to infer their regulatory relationships. The correlation may not imply the causal

relationship between the miRNA sponges. Therefore, it is necessary to develop causal discovery based methods to identify miRNA sponge interactions.

In summary, the research into miRNA sponges is emerging and it is an interesting and important area. However, there are both opportunities and challenges for designing computational methods of exploring miRNA sponge functions.

Funding

This work has been partially supported by the Australian Research Council Discovery Projects (DP130104090, DP140103617) and the Applied Basic Research Foundation of Science and Technology of Yunnan Province (No: 2013FD038).

References

1. Ambros V. microRNAs: tiny regulators with great potential. Cell 2001; 107:823-6.
2. Ambros V. MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. Cell 2003; 113:673-6.
3. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 2004; 116:281-97.
4. Bushati N, Cohen SM. microRNA functions. Annual Review of Cell and Developmental Biology 2007; 23:175-205.
5. Du T, Zamore PD. Beginning to understand microRNA function. Cell Research 2007; 17:661-3.
6. Esquela-Kerscher A, Slack FJ, Oncomirs—microRNAs with a role in cancer. Nature Reviews Cancer 2006; 6:259-269.
7. Cheng Q, Yi B, Wang A, et al. Exploring and exploiting the fundamental role of microRNAs in tumor pathogenesis. OncoTargets and Therapy 2013; 6:1675-1684.
8. Lin S, Gregory RI. MicroRNA biogenesis pathways in cancer. Nature Reviews Cancer 2015; 15:321-333.
9. Cesana M, Cacchiarelli D, Legnini I, et al. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell 2011; 147:358-369.
10. Poliseno L, Salmena L, Zhang J, et al. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 2010; 465:1033-1038.
11. Hansen TB, Jensen TI, Clausen BH, et al. Natural RNA circles function as efficient microRNA sponges. Nature 2013; 495:384-388.
12. Memczak S, Jens M, Elefsinioti A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 2013; 495:333-338.
13. Paci P, Colombo T, Farina L. Computational analysis identifies a sponge interaction network between long non-coding RNAs and messenger RNAs in human breast cancer. BMC Systems Biology 2014; 8:83.
14. Wang Y, Xu Z, Jiang J, et al. Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. Developmental Cell 2013; 25:69-80.

15. Wang J, Liu X, Wu H, et al. CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. Nucleic Acids Research 2010; 38:5366-83.
16. Fan M, Li X, Jiang W, et al. A long non-coding RNA, PTCSC3, as a tumor suppressor and a target of miRNAs in thyroid cancer cells. Experimental and Therapeutic Medicine 2013; 5:1143-1146.
17. Kallen AN, Zhou XB, Xu J, et al. The imprinted H19 lncRNA antagonizes let-7 microRNAs. Molecular Cell2013; 52:101-12.
18. Franco-Zorrilla JM, Valli A, Todesco M, et al. Target mimicry provides a new mechanism for regulation of microRNA activity. Nature Genetics 2007; 39:1033-7.
19. Marques AC, Tan J, Lee S, et al. Evidence for conserved post-transcriptional roles of unitary pseudogenes and for frequent bifunctionality of mRNAs. Genome Biology 2012; 13:R102.
20. Tay Y, Kats L, Salmena L, et al. Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. Cell 2011; 147:344-57.
21. Sumazin P, Yang X, Chiu HS, et al. An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. Cell 2011; 147:370-81.
22. Karreth FA, Tay Y, Perna D, et al. In vivo identification of tumor- suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. Cell 2011; 147:382-95.
23. Lee DY, Jeyapalan Z, Fang L, et al. Expression of versican 3'-untranslated region modulates endogenous microRNA functions. PLoS One 2010; 5:e13599.
24. Jeyapalan Z, Deng Z, Shatseva T, et al. Expression of CD44 3'-untranslated region regulates endogenous microRNA functions in tumorigenesis and angiogenesis. Nucleic Acids Research 2011; 39:3026-41.
25. Rutnam ZJ, Yang BB. The non-coding 3' UTR of CD44 induces metastasis by regulating extracellular matrix functions. Journal of Cell Science 2012; 125:2075-85.
26. Fang L, Du WW, Yang X, et al. Versican 3'-untranslated region (3'-UTR) functions as a ceRNA in inducing the development of hepatocellular carcinoma by regulating miRNA activity. FASEB Journal 2013; 27:907-19.
27. Lee DY, Shatseva T, Jeyapalan Z, et al. A 3'-untranslated region (3'UTR) induces organ adhesion by regulating miR-199a* functions. PLoS One 2009; 4:e4527.
28. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. Nature 2014; 505:344-52.
29. Huang Y, Liu N, Wang JP, et al. Regulatory long noncoding RNA and its functions. Journal of Physiology and Biochemistry 2012; 68:611-8.
30. D'Errico I, Gadaleta G, Saccone C. Pseudogenes in metazoa: origin and features. Briefings in Functional Genomics 2004; 3:157-67.
31. Chen LL, Yang L. Regulation of circRNA biogenesis.RNA Biology 2015; 12:381-8.
32. Seitz H. Redefining microRNA targets. Current Biology 2009; 19:870-3.
33. Zhou X, Liu J, Wang W, Construction and investigation of breast-cancer-specific ceRNA network based on the mRNA and miRNA expression data. IET Systems Biology 2014; 8:96-103.
34. Xu J, Li Y, Lu J, et al. The mRNA related ceRNA-ceRNA landscape and significance across 20 major cancer types. Nucleic Acids Research 2015; 43:8169-82.
35. Shao T, Wu A, Chen J, et al. Identification of module biomarkers from the dysregulated ceRNA-ceRNA interaction network in lung adenocarcinoma. Molecular Biosystems 2015; 11:3048-58.
36. Chiu YC, Hsiao TH, Chen Y, et al. Parameter optimization for constructing competing endogenous RNA regulatory network in glioblastoma multiforme and other cancers. BMC Genomics 2015; 16:S1.
37. Chiu HS, Llobet-Navas D, Yang X, et al. Cupid: simultaneous reconstruction of microRNA-target and ceRNA networks. Genome Research 2015; 25:257-67.
38. Figliuzzi M, Marinari E, De Martino A, MicroRNAs as a selective channel of communication between competing RNAs: a steady-state theory. Biophysical Journal 2013; 104:1203-13.
39. Bosia C, Pagnani A, Zecchina R, Modelling competing endogenous RNA networks. PLoS One 2013; 8:e66609.

40. Ala U, Karreth FA, Bosia C, et al. Integrated transcriptional and competitive endogenous RNA networks are cross-regulated in permissive molecular environments. Proceedings of National Academy of Science of USA 2013; 110:7154-9.
41. Yuan Y, Liu B, Xie P, et al. Model-guided quantitative analysis of microRNA-mediated regulation on competing endogenous RNAs using a synthetic gene circuit. Proceedings of National Academy of Science of USA 2015; 112:3158-63.
42. Elf J, Paulsson J, Berg OG, et al. Near-critical phenomena in intracellular metabolite pools. Biophysical Journal 2003; 84:154-70.
43. Buchler NE, Louis M. Molecular titration and ultrasensitivity in regulatory networks. Journal of Molecular Biology 2008; 384:1106-19.
44. Alexa A, Rahnenfuhrer J, Lengauer T, Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics 2006; 22: 1600-1607.
45. Brown MB, A method for combining non-independent, one-sided tests of significance. Biometrics 1975; 31: 987–992.
46. Piro RM, Ala U, Molineris I, et al. An atlas of tissue-specific conserved coexpression for functional annotation and disease gene prediction. European Journal of Human Genetics 2011; 19:1173-80.
47. Sarver AL, Subramanian S. Competing endogenous RNA database. Bioinformation 2012; 8:731-3.
48. 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.
49. Li JH, Liu S, Zhou H, et al. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Research 2014; 42:D92-7.
50. Das S, Ghosal S, Sen R, et al. lnCeDB: database of human long noncoding RNA acting as competing endogenous RNA. PLoS One 2014; 9:e98965.
51. Harrow J, Frankish A, Gonzalez JM, et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Research 2012; 22:1760-74.
52. Jeggari A, Marks DS, Larsson E. miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. Bioinformatics 2012; 28:2062-3.
53. Wang P, Ning S, Zhang Y, et al. Identification of lncRNA-associated competing triplets reveals global patterns and prognostic markers for cancer. Nucleic Acids Research 2015; 43:3478-89.
54. Ghosal S, Das S, Sen R, et al. HumanViCe: host ceRNA network in virus infected cells in human. Frontiers in Genetics 2014; 5:249.
55. Wang P, Zhi H, Zhang Y, et al. miRSponge: a manually curated database for experimentally supported miRNA sponges and ceRNAs. Database: the journal of biological databases and curation. 2015; doi: 10.1093/database/bav098.
56. Chou CH, Chang NW, Shrestha S, et al. miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. Nucleic Acids Research 2015; doi: 10.1093/nar/gkv1258.
57. 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 Research 2015; 43:D153-9.
58. Xiao F, Zuo Z, Cai G, et al. miRecords: an integrated resource for microRNA-target interactions. Nucleic Acids Research 2009; 37:D105-10.
59. Dweep H, Gretz N. miRWalk2.0: a comprehensive atlas of microRNA-target interactions. Nature Methods 2015; 12:697.
60. John B, Enright AJ, Aravin A, et al. Human MicroRNA targets. PLoS Biology 2004; 2:e363.
61. Griffiths-Jones S, Saini HK, van Dongen S, et al. miRBase: tools for microRNA genomics. Nucleic Acids Research 2008; 36:D154-8.
62. Krek A, Grün D, Poy MN, et al. Combinatorial microRNA target predictions. Nature Genetics 2005; 37:495-500.

63. Paraskevopoulou MD, Georgakilas G, Kostoulas N, et al. DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. Nucleic Acids Research 2013; 41:W169-73.
64. Kertesz M, Iovino N, Unnerstall U, et al. The role of site accessibility in microRNA target recognition. Nature Genetics 2007; 39:1278-84.
65. Helwak A, Kudla G, Dudnakova T, et al. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell 2013; 153:654-65.
66. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research 2015; 43:e47.
67. Futreal PA, Coin L, Marshall M, et al. A census of human cancer genes. Nature Reviews Cancer 2004; 4: 177-183.
68. Becker KG, Barnes KC, Bright TJ, et al. The genetic association database. Nature Genetics 2004; 36:431-2.
69. Hamosh A, Scott AF, Amberger JS, et al. Online Mendelian Inheritance in Man(OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Research 2005; 33:D514-7.
70. Baasiri RA, Glasser SR, Steffen DL, et al. The breast cancer gene database: a collaborative information resource. Oncogene 1999; 18:7958-65.
71. Mosca E, Alfieri R, Merelli I, et al. A multilevel data integration resource for breast cancer study. BMC Systems Biology 2010; 4:76.
72. Chen G, Wang Z, Wang D, et al. LncRNADisease: a database for long-non-coding RNA-associated diseases. Nucleic Acids Research 2013; 41:D983-6
73. Ning S, Zhang J, Wang P, et al. Lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers. Nucleic Acids Research 2016; 44:D980-5.
74. Marbach D, Costello JC, Küffner R, et al. Wisdom of crowds for robust gene network inference. Nature Methods 2012; 9:796-804.

Figure legend

Figure 1. miRNA sponge mechanism. The four types of miRNA sponges are shown (in red), each containing binding sites (red short rectangles) for a miRNA of interest (red semi-circle). (A) The miRNA sponges have lost the competition, the target mRNA (in purple) will be post-transcriptionally repressed by the miRNA and degraded. (B) The miRNA sponges have won the competition, the miRNA is freed from regulating the target mRNA, and the target mRNA will be released of repression and translated into protein (purple structure).

Figure 2. The number of ceRNA or miRNA sponge related publications in the past 8 years. The number of related publications is from PubMed library with the keyword ‘ceRNA’ or ‘miRNA sponge’.

Figure 3. Overlaps and differences between miRNA sponge interactions identified by PC, CMI, PPC, and SPPC. (A) lncRNA-lncRNA interactions. (B) lncRNA-mRNA interactions. (C) mRNA-mRNA interactions.

Tables

Table 1: The experimentally confirmed miRNA sponges. Most of them are extracted from [28]. In the Table, ":" stands for "competing with".

miRNA sponges species	Involved miRNA sponges	Shared miRNAs	Diseases or processes	References
LncRNA	linc-MD1 : MAML1	miR-133	Governs the time of muscle differentiation	[9]
	linc-MD1 : MEF2C	miR-135	Governs the time of muscle differentiation	[9]
	linc-RoR : (NANOG , OCT4, SOX2)	miR-145	ES cell pluripotency and self-renewal	[14]
	HULC : PRKACB	miR-372	Hepatocellular carcinoma	[15]
	PTCSC3 : targets of miR-574-5p	miR-574-5p	Thyroid cancer	[16]
	H19 : targets of let-7 family	let-7 family	Causes precocious muscle differentiation	[17]
	IPS1 : PHO2	miR-399	Controls phosphate starvation responses and homeostasis	[18]
Pseudogene	PTENP1 : PTEN	miR-17, miR-19b, miR-20a, miR-21,miR-26 and miR-214 family	Controls downstream PI3K signalling and cell proliferation in several cancers	[10]
	FOXO3B : FOXO3	miR-182	Prostate cancer	[10]
	KRASIP : KRAS	miR-143 and let-7 family	Prostate cancer	[10]
	Pbcas4 : BCAS4	miR-185	Post-transcriptional role of its protein-coding gene ancestor	[19]
CircRNA	CDR1as/ciRS-7 : targets of miR-7	miR-7	Impairs midbrain development in a manner analogous to miR-7 knockdown	[11, 12]
	Sry : targets of miR-138	miR-138	The testis-determining gene interacts with miR-138 to function as miR-138 sponge	[11]
Coding RNAs	CNOT6L : PTEN	miR-17, miR-19a, miR-19b, miR-20a, miR-20b and miR-106b	Prostate cancer	[20]
	VAPA : PTEN	miR-17, miR-19a, miR-20a, miR-20b, miR-26b, miR-106a and miR-106b	Prostate cancer	[20]
	(ABHD13, CCDC6, CTBP2, DCLK1, DKK1, HIAT1, HIF1A, KLF6, LRCH1, NRAS, RB1, TAF5 and TNKS2) : PTEN	32 common miRNAs	Glioblastoma	[21]
	ZEB2 : PTEN	miR-25, miR-181, miR-200b, miR-	Melanoma	[22]

		92		
	VCAN : (Rb1 and PTEN)	miR-144, miR-136 and miR-199a-3p	Breast cancer	[23]
	CD44 : CDC42	miR-216a, miR-330 and miR-608	Breast cancer and many other cancers	[24]
	CD44 : FN1	miR-491, miR-512-3p and miR-671	Induces metastasis by regulating extracellular matrix functions	[25]
	CD44 : Col1α1	miR-328	Induces metastasis by regulating extracellular matrix functions	[25]
	CD34 : VCAN	miR-133a, miR-144 and miR-431	Hepatocellular carcinoma development	[26]
	FN1 : VCAN	miR-133a, miR-199a-3p and miR-431	Cell proliferation	[26, 27]

Table 2: summary of computational approaches for identifying miRNA sponge interaction networks.

Methods	Type of methods	Brief description	Data sources	Notes	Software tool
Zhou *et al.*[33]	Pair-wise correlation	Explore the breast-cancer-specific miRNA sponge interaction network in matched miRNA and gene expression data. • Extract significant miRNA sponge interaction pairs by checking whether they share the similar set of miRNAs using hypergeometric cumulative distribution test. The positively correlated RNAs are then predicted as miRNA sponges. • Survival analysis shows that the hub genes in the miRNA sponge interaction network are candidate biomarkers in breast cancer.	• Matched miRNA and gene expression data. • Putative miRNA-target interactions.	• Suitable for inferring cancer-specific miRNA sponge interaction networks in heterogeneous data, including expression data and miRNA-target binding information.	Upon request
Xu *et al.* [34]	Pair-wise correlation	Construct the cancer related miRNA sponge interaction networks for 20 cancer types. • Extract significant miRNA sponge interaction pairs by checking whether they share the same miRNAs using hypergeometric cumulative distribution test. The positively correlated RNAs are then predicted as miRNA sponges. • Identify common interactions and conserved hubs in the 20 cancer types, as well as the specific interactions for each cancer.	• Ago CLIP-supported miRNA-target interactions, cancer-associated miRNAs, and genes related with cancer hallmarks. • Genome-wide protein coding expression profiles across 20 major cancers.	• Provide a systematic analysis of the mRNA related miRNA sponge interaction landscape across 20 major cancer types.	Upon request
Shao *et al.* [35]	Pair-wise correlation	Identify dysregulated miRNA sponge interactions by integrating expression profiles and putative miRNA-target interactions. • Extract significant miRNA sponge interaction pairs by checking whether they share the similar set of miRNAs using hypergeometric cumulative distribution test. The positively correlated RNAs are then predicted as miRNA sponges. • Identify dysregulated miRNA sponge	• Matched miRNA and gene expression data. • Putative miRNA-target interactions, list of lung cancer genes, lncRNAs and miRNAs.	• Provide a method to identify modules of biomarkers from the dysregulated miRNA sponge interaction network.	Upon request

		interactions by checking whether the correlation coefficients are significantly different in cancer and normal samples.			
Chui *et al.* [36]	Pair-wise correlation	Construct the optimal miRNA sponge regulatory network by combining expression profiles and putative miRNA-target interactions. • Calculate pairwise correlation coefficients of putative miRNA sponge interaction pairs from gene expression data using Pearson. • Partition the pairwise correlation coefficients into different groups based on the states of each essential factor of putative miRNA sponge interaction pairs, and identify optimal miRNA sponge interaction pairs using the two-sample Kolmogorov-Smirnov test.	• Matched miRNA and gene expression data. • Putative miRNA-target interactions.	• Provide a method to identify the optimal miRNA sponge regulatory network.	Upon request
Paci *et al.*[13]	Partial association	Explore the role of lncRNAs as miRNA sponges by sequestering mRNAs from miRNA repression or degradation in matched miRNA and gene expression data. • Compute for each (lncRNA, miRNA, mRNA) triplet the difference between the Pearson correlation of (lncRNA, mRNA) and partial correlation (lncRNA, mRNA|miRNA) to examine the contribution of the miRNA into the relationship between lncRNA and mRNA. • Compare the miRNA sponge interaction network in normal and tumor samples	• Matched miRNA and gene expression data.	• A good method for inferring miRNA sponge interaction networks in multiple biological conditions.	Upon request
Sumazin *et al.*[21]	Partial association	Investigate the ability of coding and non-coding RNAs to act as miRNA sponges in a large set of matched miRNA and gene expression profiles. • Consider RNA-RNA pairs sharing a significant overlap of common miRNAs, and combine expression data with predicted miRNA-target regulatory interactions to identify the	• Matched miRNA and gene expression data. • Predicted miRNA-target interactions.	• Suitable for inferring cancer-specific miRNA sponge interaction networks in heterogeneous data, including expression data and miRNA-target binding information.	http://wiki.c2b2.columbia.edu/califanolab/index.php/Software/HERMES

		mediated RNA-miRNA-RNA triplets. • Estimate statistical significance of information divergence between the mutual information (MI) and conditional mutual information (CMI) to identify miRNA sponge modulators.			
Chui *et al.*[37]	Partial association	Integrate sequence-based evidence and functional clues from RNA and miRNA expression analysis to simultaneously reconstruct miRNA-target and miRNA sponge interaction network. • Evaluate candidate miRNA binding sites in 3' UTR. • Integrate information about selected miRNA binding sites, their multiplicity, and the statistical dependency between miRNA and gene expression. • Assessing whether inferred targets compete for miRNAs to infer miRNA sponge interaction networks.	• 3'UTR sequence of mRNA and miRNA sequence. • Matched miRNA and gene expression data, and protein expression data.	• Allow users to simultaneously identify miRNA-target interactions and miRNA-mediated sponge interactions.	http://sourceforge.net/projects/cupidtool/
Figliuzzi *et al.*[38]	Mathematical modelling	Study a minimal, rate-equation-based model to describe the cross-talk between miRNA sponges arising from completion at steady state through a systematic analytical characterization of the sensitivity to small changes in the transcription rates.	• Simulation data	• Provide a clear quantitative understanding of miRNA-mediated sponge effect from mathematical models.	Upon request
Bosia *et al.*[39]	Mathematical modelling	Propose a stochastic model to analyse the equilibrium and out-of- equilibrium properties of miRNA sponge interaction network between miRNAs and mRNAs.	• Simulation data	• Provide a clear qualitative and quantitative understanding of miRNA-mediated sponge effect from mathematical models.	Upon request
Ala *et al.*[40]	Mathematical modelling	Present a mass-action model to determine the optimal conditions for miRNA sponge activity in silico. • Define a miRNA sponge interaction network linking two miRNA sponges if one miRNA sponge is within the top 40 genes in terms of the distance	• Putative miRNA-target interactions. • Conserved coexpression networks of RNA-RNA interactions.	• Allows for miRNA sponge identification and qualitative miRNA-mediated sponge interactions predictions.	Upon request

		from the other miRNA sponge and vice versa. • Refine miRNA sponge interaction network by removing all edges not existed in the conserved coexpression networks.			
Yuan *et al.* [41]	Mathematical modelling	Establish a coarse-grained model for a minimum miRNA sponge interaction network to quantitatively analyse the behaviour of miRNA sponge regulation. • Validate the computational results by using synthetic gene circuits in HEK293 cells.	• Simulation data	• Provide a clear quantitative understanding of miRNA-mediated sponge effect from mathematical models and synthetic gene circuits.	Upon request

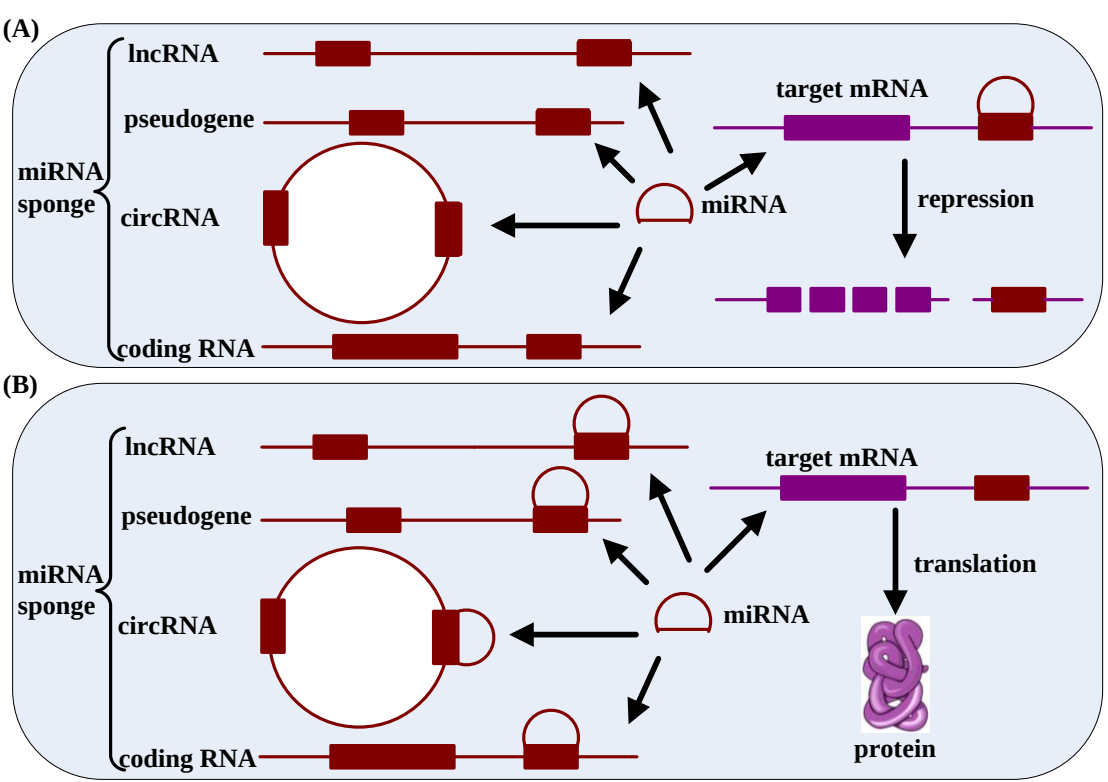
Table 3: summary of databases for miRNA sponge interactions.

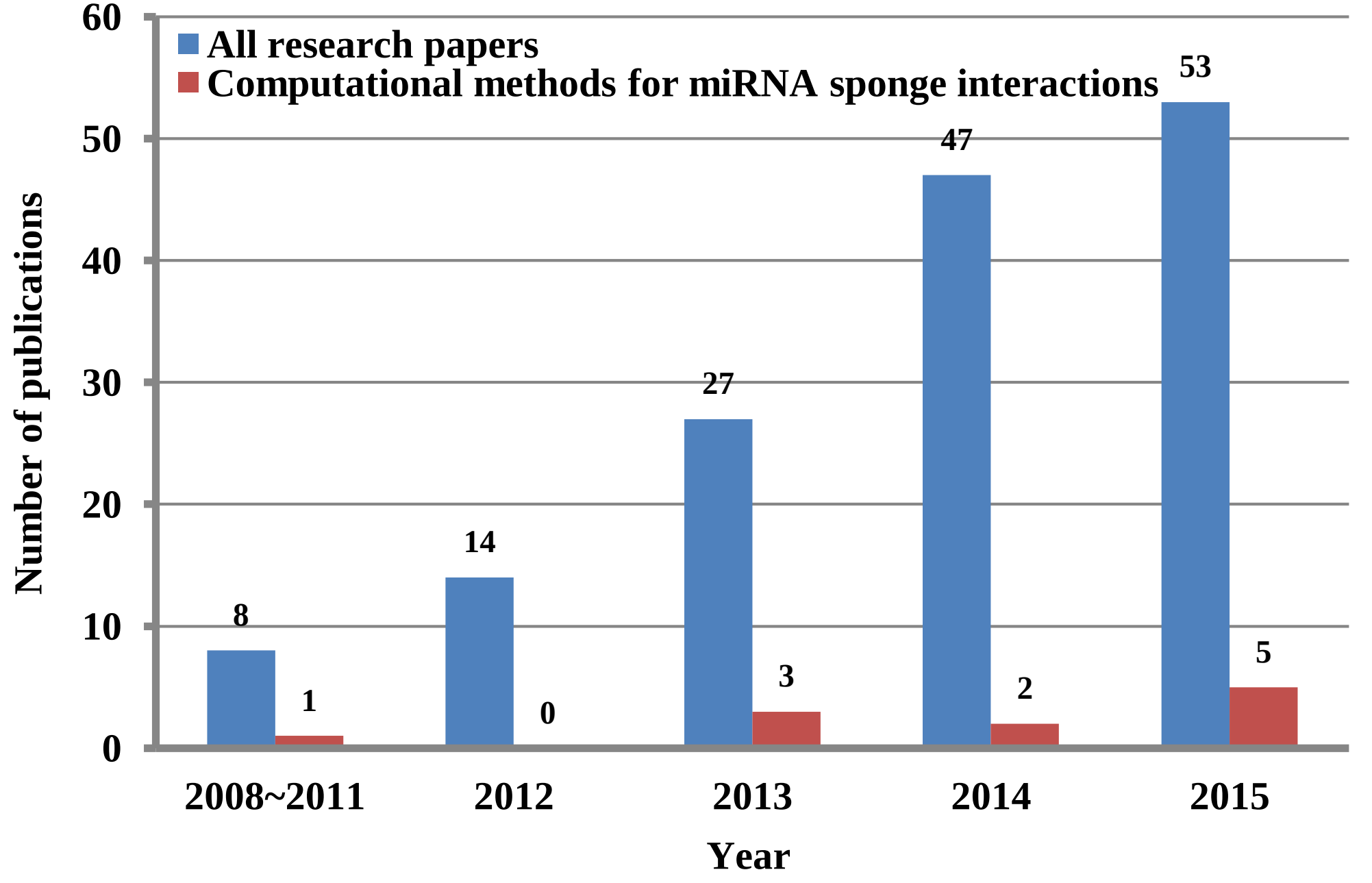
Databases	Type of databases	Brief description	Data source	Notes	Link
ceRDB	Computational prediction	• Predict miRNA sponges for specific mRNAs targeted by miRNAs by examining the co-occurrence of miRNA response elements in the mRNAs on a genome-wide basis.	• 3'UTR sequence of mRNA and miRNA sequence.	• The first miRNA sponge interactions database. • Allow users to indentify miRNA sponge for a given mRNA target based on the homology of miRNA binding sites present in the 3'UTR sequence.	http://www.oncomir.umn.edu/cefinder/
starBase	Computational prediction	Use hypergeometric test to evaluate whether each miRNA sponge pair share significant common miRNAs. • Utilise HITS-CLIP, PAR-CLIP, iCLIP and experiment results of 108 datasets from 37 studies and tumour samples (14 cancer types, >6000 samples) to explore miRNA sponge interactions.	• Large-scale CLIP-Seq (HITS-CLIP, PAR-CLIP, iCLIP, CLASH) data and tumor samples (14 cancer types, >6000 samples). • Predicted miRNA-target interactions	• Allow user to browse the predicted miRNA sponge pairs or regulatory networks by integrating the interactions from potential miRNA targets (miRanda/mirSVR) overlapping with CLIP-Seq data.	http://starbase.sysu.edu.cn/
lnCeDB	Computational prediction	Provide a database of human lncRNAs that can potentially act as miRNA sponges. • Calculating miRNA sponge score from the ratio of the number of shared MREs between the miRNA sponge pair with the total number of MREs of the individual candidate gene. • Determining *p*-value for each miRNA sponge pair by hypergeometric test using the number of shared miRNAs between the miRNA sponge pair against the number of miRNAs interacting with the individual RNAs.	• Putative miRNA-target interactions, miRNA-lncRNA interactions and mapped AGO clipped regions of the targets. • Tissue specific miRNA, lncRNA and mRNA expression data.	• Not only browse for common miRNAs of lncRNA-mRNA pairs, but also compare the expression of the pair in 22 human tissues to estimate the chances of the pair for actually being miRNA sponges.	http://gyanxet-beta.com/lncedb/
LncACTdb	Computational prediction	Provide a database of human lncRNA-associated competing triplets.	• Computationally predicted or	• Allow user to browse, search, and download lncRNA	http://www.bio-bigdata.net/LncACTdb/

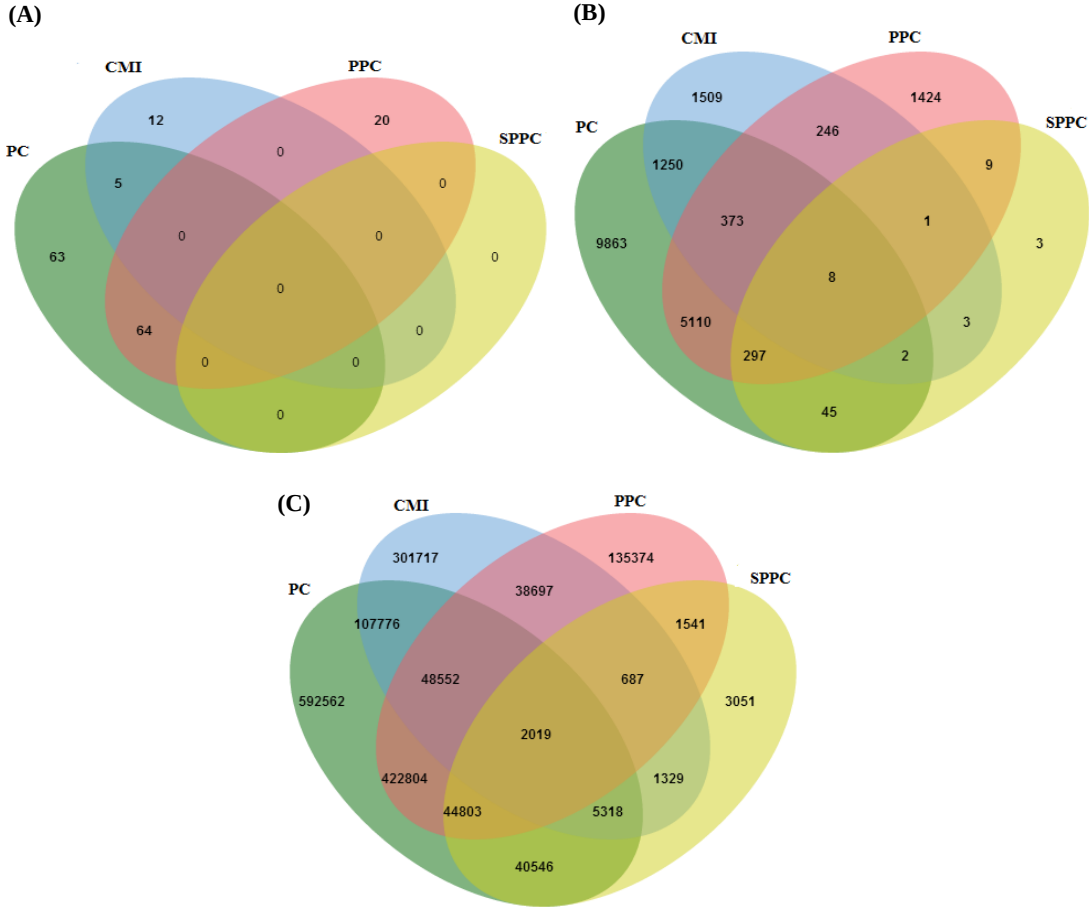
		• Predicting lncRNA related miRNA sponge interactions by integrating data from several in silico target prediction studies, Argonaute-CLIP experiments, and expression profiles from transcriptome sequencing.	experimentally supported miRNA regulatory interactions • Expression profiles of lncRNA, miRNA and gene from transcriptome sequencing	related miRNA sponge interactions. • Provide three tools called LncACTFun, LncACTGet and LncACTView, to open an lncACTs function location map within Gene Ontology-directed Acyclic Graph and KEGG pathways, generatelncACTs for user-customized expression profiles, and browse genome-wide lncACTs.	
HumanViCe	Computational prediction	Develop a comprehensive database to provide the potential miRNA sponge networks in virus infected human cells. • Conduct Gene ontology (GO) and pathway enrichment analysis of the genes comprising the miRNA sponge networks in the virus infected cells. • Identify a vast number of non-coding transcripts playing as potential miRNA sponges to the immune response associated genes.	• Human and viral mature miRNA sequence data, miRNA sequences from 11 virus species known to infect human, and human lncRNAs and circRNA sequence data.	• Users can browse for viral miRNA targets on host protein-coding and non-coding genes by choosing a particular virus, view details of the targets by choosing a particular viral miRNA, and check the potential miRNA sponges for a chosen transcript. • Users can download the results of pathway enrichment analysis for the genes comprising the miRNA sponge networks for each of the 10 viruses after searching by a particular virus name.	http://gyanxet-beta.com/humanvice
miRSponge	Experimental confirmation	The first experimentally validated miRNA sponge interactions database, and miRNA sponges include lncRNAs, pseudogenes, circRNAs, coding RNAs, viral RNAs and artificial engineered sponges.	• 599 sponge-miRNA interactions and 463 miRNA sponge associations in 11 species, following manual curating from nearly 1000 articles.	• Provide a user-friendly interface for convenient browsing, retrieval and downloading, and offer a submission page for newly validated miRNA-sponge and miRNA sponge association data.	http://www.bio-bigdata.net/miRSponge/

Table 4: comparison of four methods in identifying experimentally confirmed and breast cancer related miRNA sponge interactions.

Methods	**# interactions**	**# breast cancer interactions**	**Percentage of breast cancer interactions**	**Validated miRNA sponge interactions**
PC	1274708	46716	3.66%	CNOT6L: PTEN, CCDC6: PTEN, LRCH1: PTEN, ZEB2: PTEN, AFF1: PTEN, RUNX1:PTEN, SERINC1:PTEN
CMI	509504	20082	3.94%	CNOT6L: PTEN, TNKS2: PTEN, ZEB2: PTEN, AFF1: PTEN, RUNX1:PTEN
PPC	702029	26803	3.82%	/
SPPC	99662	4223	4.24%	CNOT6L: PTEN, ZEB2: PTEN


(A)
lncRNA
pseudogene
miRNA sponge
circRNA
coding RNA
miRNA
target mRNA
repression
(B)
lncRNA
pseudogene
miRNA sponge
circRNA
coding RNA
miRNA
target mRNA
translation
protein


All research papers
Computational methods for miRNA sponge interactions
Number of publications
60
50
40
30
20
10
0
8
1
14
0
27
3
47
2
53
5
2008~2011
2012
2013
2014
2015
Year


(A)
CMI
PPC
PC
SPPC
12
20
0
5
0
0
0
63
0
0
0
64
0
0
0
0
(B)
CMI
PPC
PC
SPPC
1509
1424
246
1250
9
373
1
9863
3
8
5110
3
297
2
45
(C)
CMI
PPC
PC
SPPC
301717
135374
38697
107776
1541
48552
687
592562
3051
2019
422804
1329
44803
5318
40546