

Review of Computational Prediction of Competing Endogenous RNA

Sherin K* and Achuthsankar S. Nair

Department of Computational Biology and Bioinformatics, University of Kerala, Karyavattom, Thiruvananthapuram 695581, India

Abstract

MicroRNAs (miRNAs) are the gene regulatory molecules that bind to microRNA response elements (MREs) in the 3' untranslated regions (UTRs) of mRNAs. The repressive activity of miRNA is counteracted by "miRNA sponges" or "competing endogenous RNAs (ceRNAs)" termed because of its competing nature of sequestering miRNA's effect. The ceRNAs with multiple MREs for a miRNA interact more, resulting in a regulatory gene network layer. The perturbation of ceRNA network causes various diseases including cancer. This discovery has increased the analysis of ceRNA interactions, networks and dynamics. Currently, a number of ceRNA-ceRNA interaction analysis and prediction tools are available online. This review focuses on computational prediction of mRNA-mRNA pairs acting as ceRNAs and its significance in RNA therapeutics. The computational prediction tools are compared with respect to the input data retrieved, features considered and the prediction method.

Keywords: ceRNA; ceRNA prediction; Feature extraction; Feature selection; ceRNA networks

Abbreviations: ceRNA: Competing Endogenous RNA; LSFS: Laplacian Score Based Feature Selection; MREs: MicroRNA Response Elements; UTRs: Untranslated Regions; ncRNA: Non Coding RNA; ceRNETs: ceRNA Networks; GOI: Gene of Interest

Introduction

In gene expression, the intermediary RNA plays much more important role than protein synthesis. RNA spans from very short RNAs to long non coding RNAs. It can self-replicate, storing information as DNA and can act as a catalyst paving way to the "RNA World Hypothesis" put forward by Gilbert [1]. Double stranded RNAs like short interfering RNAs (siRNA), short hairpin RNA (shRNA) and miRNAs can regulate gene expression and the process is termed RNA mediated interference or RNA interference (RNAi). In human genome, major portion of RNA comprises of non-coding RNAs (ncRNAs) while coding RNA are only ~2.3% [2]. Small non coding RNAs (sncRNAs) are less than 200 nucleotides. PIWI-interacting RNAs (piRNAs), miRNAs, small nucleolar RNA (snoRNA) and siRNAs are important sncRNAs [3]. In gene regulation either the upregulation or downregulation of the protein synthesis occurs. The gene regulation depends on the presence of transcription factors, transcription regulatory enzymes or double stranded RNAs. The miRNAs are ~22 nucleotides (nt) long molecule also known as gene silencers since it blocks the protein production [4]. The miRNA binding to mRNA leads to mRNA cleavage and translation repression [5]. Since its discovery in metazoans in 1993 [6] miRNAs are proved to have vital roles in biological processes such as cell division and death, immunity, cellular metabolism and cell movement [7]. MiRNA let-7 prevents the proliferation of cancer-initiating stem cells [8] allowing tumor suppressor miRNAs to be replaced in enhancing traditional cancer chemotherapy [9].

CeRNAs are transcripts which compete for miRNA binding, regulating each other's activity at post transcriptional level. The counter mechanism of miRNA repression activity by transcripts, described as miRNA-sponge activity was found in 2007 [10]. The hypothesis put forward by [11] established the existence of miRNA's competitive inhibitors known as CeRNAs. In case of miRNAs and other different types of RNAs there are conventional naming systems to name a particular sequence. The name gives us the detail such as the organism from which the sequence is derived, type of the sequence, position etc. The miRNAs are named as hsa-miR-19a where hsa denotes the

species, miR denotes it is a mature miRNA and 19 says that it was the 19th family that was named. In case of ceRNAs even though there is rapid growth in research, so far no standard nomenclature system is available for naming it. MiRNA targets are identified and validated by different low and high throughput expression technologies such as qRT-PCR, luciferase reporter assays and western blot techniques. If the miRNA target and the miRNAs bound are available, the identification of MREs can be done by using different expression profiling platforms like cross-linking and immunoprecipitation (CLIP), photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP), stable isotope labelling with amino acids in cell culture (SILAC) or translation profiling and cross-linking ligation and sequencing of hybrids (CLASH) which gives insights into ceRNAs activity [12-14]. As a preliminary and cost effective method the ceRNAs can be predicted computationally. In the recently evolved "RNA therapeutics" different types of RNAs are used as therapeutic agents and RNA pathways are intensively analysed to find new therapies. The concept of ceRNA has advanced rigorously since the identification of its role in cancer suppressing treatments [15]. It can be assumed that the increased complexity of molecular mechanism involving numerous mRNAs, lead ceRNA prediction tools to be centered more on a particular gene or RNA pathway. The complexity and the inaccuracy about the knowledge of miRNA mediated ceRNA interaction makes computational prediction of ceRNA inevitable. New prediction tools have been developed adding novel features and methods. In many cases ceRNA prediction tools analyse combination of input data such as sequence data, expression data or miRNA - mRNA interaction data and use both rule based and data driven method for the prediction. There are three reviews on ceRNA prediction methods, one focuses on ceRNA prediction from miRNA sponge interactions [16] and the other

***Corresponding author:** Sherin K, Department of Computational Biology and Bioinformatics, University of Kerala, Karyavattom, Thiruvananthapuram 695581, India, Tel: + 9895315145; E-mail: sherin.cb@gmail.com

Received February 13, 2019; **Accepted** March 05, 2019; **Published** March 12, 2019

Citation: Sherin K, Nair AS (2019) Review of Computational Prediction of Competing Endogenous RNA. J Proteomics Bioinform 12: 029-037. doi: [10.4172/0974-276X.1000493](https://doi.org/10.4172/0974-276X.1000493)

Copyright: © 2019 Sherin K, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

two from expression data [17,18]. In the former review, *computational approaches for ceRNA prediction from miRNA-ceRNA interaction data are divided into i) pair-wise correlation approach ii) partial association approach and iii) mathematical modelling approach.*

The publication report generated from Web of Science for the last 10 years for "competing + endogenous + RNA" OR "ceRNA" OR "miRNA + Sponge" on 12th February 2019 shows an increase in number of publications, showing the significance of ceRNA. Total number of publications for last 10 years is 1,446.

Figure 1 shows the number of publications made each year. This is a general review on ceRNA prediction focusing on protein coding RNAs acting as ceRNA from sequence, expression and interaction data. In this review, methods of implementation of ceRNA prediction tools are discussed.

MiRNA- Mediated RNA Regulation

MiRNAs are transcribed as long primary transcripts called primary microRNA (pri-miRNA) which are processed into precursor microRNA (pre-miRNA) and then cleaved into miRNA: miRNA* duplex by Dicer-like1 enzyme (DCL1) and subsequently exported into the cytoplasm. The single stranded mature miRNAs are incorporated into Argonaute (Ago) proteins to form RNA-induced silencing complex (RISC) [19]. MiRNA mediated gene regulation is caused by non mutual mechanisms targeting translation. It happens either by inhibition of initiation by repression of the competent ribosome assembly or miRISC inhibiting the formation of translation inhibition complexes or by blocking PolyA Binding Protein (PABP) binding on mRNA [20]. It also happens by the inhibition of elongation step in translation or by promoting mRNA deadenylation, degradation or mRNA sequestration [21]. Multitude of miRNA data shows that current miRNA research is not limited to functional aspects. Different diseases including lung cancer and liver cancer, viral diseases such as Hepatitis C and HIV-1, immune-related diseases and neurodegenerative diseases such as Parkinson's and Alzheimer's diseases are associated with cellular miRNAs. Cancer related miRNAs such as miR-15, miR-16 and let-7 are categorized into tumor suppressors and miR-21 and miR-155 are categorized as oncogenes based on the cancer cell proliferation

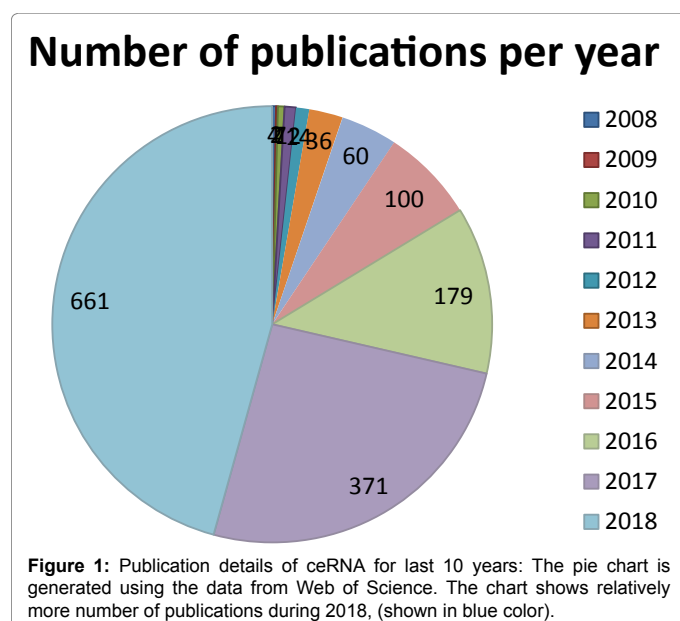
[7]. Studies have also proved that miRNA can be used as therapeutic targets for different human diseases including cancer [22]. To know the function of miRNA it is important to know the targets to which miRNA binds. The interaction between miRNA and the targets are different in plants and animals. In plants the sequences are characterized by extensive complementarity between miRNA and target and in animals the complementarity is imperfect making the target prediction more complicated [23]. Further, multiple genes are targeted by a single miRNA [24] and several miRNAs can target a single gene [25]. In spite of these complexities in miRNA target prediction, there are numerous tools available online implementing diverse algorithms and are reviewed by many researchers [26-28]. A Survey by Akhtar [29] had analyzed 129 miRNA-target prediction tools and around twenty tools are selected as precise and effective for supporting researchers in this field. DIANA-TarBase v8 [30], miRTarBase [31] and MiRecords [32] are some of the manually curated database hosting enormous records of validated interactions between miRNAs and target genes. The number of miRNA-gene interactions available in DIANA-TarBase V8 is 1080276 as on 9th August 2018, while in miRTarBase it is 422517. These numbers were only thousands a decade back and the increase in data shows the immense study in miRNA and its target predictions. Currently, there are about ten databases with experimentally validated miRNA targets [13].

CeRNA- Molecular Mechanism and Significance

The coding and non-coding transcripts competes each other to provide binding sites for miRNA. This RNA→miRNA logic has altered the concept of central dogma consisting of transcription and translation. The ceRNA concept replaces the conventional logic of miRNA binding, i.e. miRNA→RNA to RNA→miRNA→RNA interplay [33]. The microRNA response elements (MREs), was discovered first in plants [34]. CeRNA hypothesis refers to the MREs to which miRNA binds, as letters in the RNA language through which microRNA communicate with each other. The RNAs having same MREs compete with each other to provide binding for miRNAs regulating other transcripts expression level (Figure 2a). Effective binding is carried out by those RNAs that share multiple MREs (Figure 2b). CeRNAs in 3'UTRs act in trans also and regulate other transcripts expression level [33]. Coding as well as non-coding transcripts such as circular RNAs (circRNAs or ciRs), pseudogenes and lncRNAs can act as competing endogenous RNA [35,14]. Since the function of mRNAs acting as ceRNAs was proved [36], there were many more studies revealing the significance of ceRNA activity in cancer. The miRNA-mediated ceRNA interaction layers reveal mechanisms of pathogenesis and regulation of normal cell physiology. While it is difficult to study the mechanism of normal cells, tumorous cells provide insights into underlying function. Since ceRNAs are natural miRNA sponges they can be efficiently used in cancer treatment. A natural evidence is the over expression of CD44 3' -UTR inhibiting tumor formation [37]. The experimentally proved and computationally predicted mRNAs acting as ceRNAs and the genes that are free for protein synthesis are given in Table 1. In these ceRNAs, VCAN 3' UTR, ceRNA of VCAN [85], PTEN 3'UTR [36] and SERINC1, CNOT6L, and VAPA, ceRNAs of PTEN [38] show tumor suppressive properties.

ceRNA Networks

The interactions between coding transcripts and miRNAs forms a regulatory network of transcriptome layer and are termed as ceRNA network (ceRNETS). The effectiveness of ceRNETS is dependent on the relative concentration of ceRNA and the number of interacting



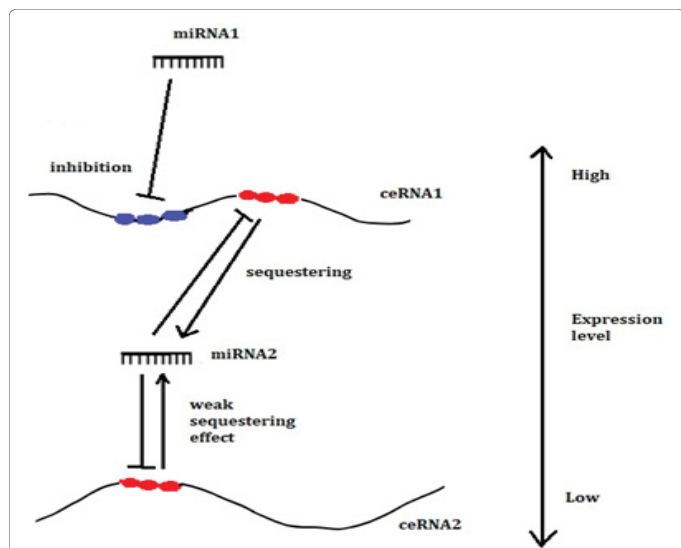


Figure 2a: ceRNA mechanism: mRNA1 and mRNA2 are ceRNAs that can bind to miRNA1. When mRNA1 expression level is high, miRNA2 is sequestered by mRNA1, causing decreased effect of miRNA2 on mRNA2. This in effect causes an increased mRNA2 level. The MREs are shown in red and blue ovals.

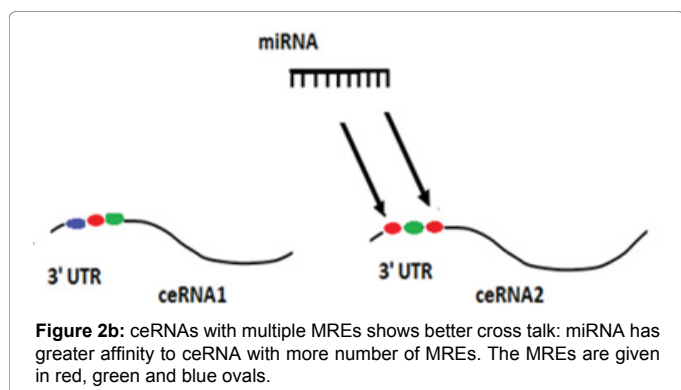


Figure 2b: ceRNAs with multiple MREs shows better cross talk: miRNA has greater affinity to ceRNA with more number of MREs. The MREs are given in red, green and blue ovals.

miRNAs in the cell. CeRNA expression levels will be reduced with lesser number of miRNAs. The miRNAs availability is dependent on the type and pathological condition of the tissue as well the sub cellular localization. Moreover the MREs capacity to bind is also crucial for the existence of ceRNETS [33]. Based on the position of ceRNAs, two types of networks exists. In direct linkage networks, there is direct link between two ceRNAs with shared MREs for miRNAs and in indirect linkage network two ceRNAs are connected through another ceRNA [39]. By considering the ceRNA-ceRNA cross talk, breast cancer specific ceRNETS described are proved to predict risk of metastasis in breast cancer patients [40]. The study on ceRNA interactions in cancer gives more insights into tumorigenesis and cancer therapy [41,42]. Biological network systems are either random networks or scale-free networks. Random networks have similar number of links between all nodes and scale-free network have more number of links connecting some nodes and these nodes act as hubs. PTEN, a tumor suppresser gene is predicted to be a potential target in replacement based therapeutic strategies since it is found to act as hub interacting with multiple genes [43]. Using miRNA-mRNA interaction tool miRTargetLink [44], a direct linkage ceRNA network where PTEN acting as hub is generated and is shown in Figure 3. The network is generated with genes *SERINC1*, *CNOT6L*, *VAPA*, *ABHD13*, *CCDC6*, *CTBP2*, *DCLK1*, *DKK1*, *HIAT1*, *HIF1A*, *KLF6*, *LRCH1*, *NRAS*, *RB1*, *TAF5*, *TNKS2* and *ZEB2* as input which are experimentally proved to be the ceRNAs of PTEN. The generated network includes *CTBP2*, *DKK1*, *HIF1A*, *KLF6*, *NRAS* and *ZEB2* with shared interactions and the genes with no shared interactions were excluded.

Computational Prediction of ceRNA

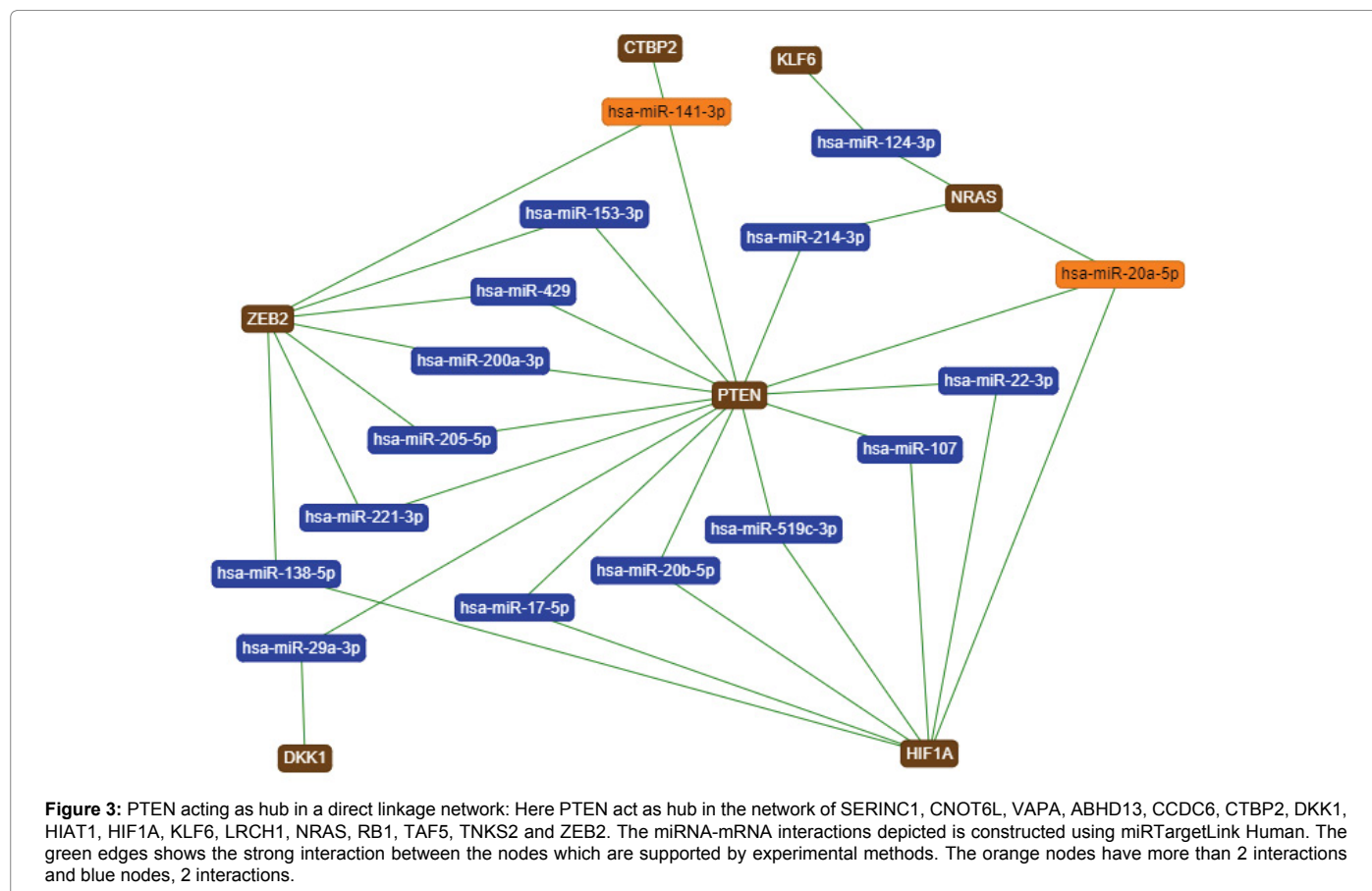
The computational prediction of ceRNA varies on the basis of whether the data analysed is sequence data, expression profile or miRNA-mRNA networks. The general procedure followed in ceRNA prediction from sequence data is discussed below.

Identification of miRNAs Targeting Gene of Interest (GOI)

According to individual research interest the GOI may vary. MiRNAs targeting the GOI are identified from resources like miRTarBase or MiRWalk [45] with validated miRNA targets.

mRNA free for protein synthesis	mRNAs acting as ceRNA	Reference
Experimentally verified		
<i>VCAN</i> , <i>Rb1</i> , <i>PTEN</i>	<i>VCAN</i> 3'UTR	Lee et al. [85]
<i>PTEN</i>	<i>PTEN</i> 3'UTR	Poliseno et al. [36]
<i>CD44</i> , <i>CDC42</i>	<i>CD44</i> 3'UTR	Jeyapalan et al. [37]
<i>PTEN</i>	<i>SERINC1</i> , <i>CNOT6L</i> , and <i>VAPA</i>	Tay et al. [38]
<i>PTEN</i>	<i>ZEB2</i>	Karreth et al. [35]
<i>PTEN</i>	<i>ABHD13</i> , <i>CCDC6</i> , <i>CTBP2</i> , <i>DCLK1</i> , <i>DKK1</i> , <i>HIAT1</i> , <i>HIF1A</i> , <i>KLF6</i> , <i>LRCH1</i> , <i>NRAS</i> , <i>RB1</i> , <i>TAF5</i> , and <i>TNKS2</i>	Sumazin et al. [62]
<i>CD44</i> 3'UTR	<i>FN1</i> , <i>Col1a1</i>	Rutnam & Yang [91]
<i>VCAN</i> 3'UTR	<i>VCAN</i> , <i>CD34</i> , <i>FN1</i>	Fang et al. [81]
<i>HMGA2</i>	<i>TGFBR3</i>	Kumar et al. [83]
<i>FOXO1</i>	E-cadherin	Yang et al. [93]
<i>AEG-1</i>	Snail, Vimentin	Liu et al. [47]
<i>C-Myc</i>	<i>PML/RARα</i>	Ding et al. [80]
Computationally predicted		
<i>LMNA</i>	<i>DICER1</i> , <i>CDKN1A</i> , <i>NFKB1</i> , <i>TP53</i> , <i>VEGFA</i> , <i>APC</i> , <i>BCL2</i> , <i>CD44</i> , <i>CDC25A</i> , <i>CDK6</i> , <i>EIF2C1</i> , <i>EIF2C2</i> , <i>HDAC9</i> , <i>IL1B</i> , <i>KRAS</i> , <i>MYC</i> , <i>RNASEN</i>	Arancio et al. [77]

Table 1: The mRNAs acting as ceRNAs.



Creating Training Dataset

Once the primary step of prediction of targets of miRNA's, targeting the GOI are done, next step is the creation of training set. Training set consists of features extracted from positive and negative set of miRNA and miRNA targets pairs. In most of the studies miRNAs are retrieved from miRNA database, miRBase [46] and miRNA targets from miRNA target database miRTarBase. In TargetMiner [47] systematically generated and biologically validated negative training set is proved to give better result than randomly generated artificial miRNA- mRNA negative training set. In this tool a set of negative examples were found using computational prediction tools among which potential negative examples were selected using expression profiling. These pairs were then confirmed as negative by biological validation.

Feature Extraction and Selection

In general the features are any measurable characteristic of the subject under analysis. The features that can be extracted for miRNA target prediction are generally classified into sequence features, structural features and positional features. Sequence features include base frequencies and compositions. Structural features include stems, loops, bulges and folding information for miRNA-target duplexes. Positional features are the matching status of miRNA-target interactions such as match, GU match or mismatch at a given position [48]. Unique features considered in the prediction of miRNA targets are seed match, conservation, free energy and site accessibility [49]. Features are extracted from region of interest from potential binding sequences. Potential binding sequences have the potential binding sites

i. e. the complementary sequences corresponding to the seed site of miRNA. MiRNA seed site is the first 2-8 nucleotides starting at the 5' end and counting towards the 3' end [50]. Majority of the tools look for Watson-Crick (WC) complementarity, adenosine (A) pairs with uracil (U) and guanine (G) pairs with cytosine (C) for seed match and these matching are found to reduce false-positive predictions [50,51].

Wobble base pair, a non-Watson-Crick pair model is also considered as a feature in miRNA target prediction. For finding the complementary sub-sequences, MTar [52] had adopted variation of Smith Waterman algorithm by applying a different scoring scheme. Distinction is made between Watson-Crick pair and Wobble base pair by giving a score of 5 for G:C and A:U and 1 for G:U pair. Mismatches are given a score of -3 and gap openings are given a higher penalty of -8 than the gap extension score of -2. In MBSTAR [47], wobble base pairing is considered in view of its significant properties such as ligand binding capacity and acceptable thermodynamic stability. Conservation which is the maintenance of a sequence across species is another common feature considered in target prediction. The conservation in miRNA seed region is higher than in the non-seed regions [50]. Thermodynamic stability also plays a key role in miRNA target prediction [53]. To measure the stability of miRNA:mRNA duplex, Free Energy (or Gibbs Free Energy) is calculated and change in Free Energy (ΔG) is used as an indicator to show the binding energy [28]. Other significant features are pair-wise binding structure features and UTR features such as length of 3' UTR, site density features and binding site score features [54]. The regions surrounding the complementary binding site known as flanking regions are also

significant in determining binding site accessibility [55]. In different tools different length of flanking regions are considered for feature extraction which varies from 10 to 30 nucleotides on either side of miRNA seed region. In each tool different combination of these features are incorporated. Once the features are extracted, they are converted into a feature vector for further analysis.

Feature selection methods, like filter, wrapper and embedded methods are used to select a subset of relevant features. Filters work independently of the predictor as a pre-processor step while wrappers score the variable according to their predictive power. Embedded methods select the subset of features during the training process depending on the learning technique [56]. The univariate filter technique ignores feature dependencies while multivariate filter, models feature dependencies. Wrapper feature selection methods, which are classified into deterministic and randomized methods, consider feature dependencies but has a drawback of overfitting compared with filter method. Both wrapper and embedded methods are specific to a given machine learning algorithm and embedded methods are computationally intensive [57]. Laplacian score based feature selection (LSFS), unsupervised discriminative feature selection (UDFS) and multiclass feature selection (MCFS) methods were used in miRNA target prediction tool, MBSTAR. The number of features extracted in MBSTAR was 371 comprising of sequence features like single, di, tri and quad-nucleotide frequencies and structural features like loops, bulge loops, hairpin loops and multibranch loop. These are extracted from the flanking region using Vienna RNA package. Since LSFS gave better result Laplacian score was used to select top 40 features. A good feature has small Laplacian score [58]. In MBSTAR, 40 features with high (1- Laplacian score) is selected for prediction. The method rRMR is used in SVMicro for feature selection where the initial number of 113 site and 30 UTR features were decreased to 21 optimal site features and 18 optimal UTR features.

CeRNA Prediction & Performance Analysis

In most tools miRNA targets are predicted based on some rules derived from experiments (Rule based) or by implementing machine learning algorithm (data driven). Tools like TargetScan and MiRanda use combination of features for target prediction. From these miRNA targets, targets with high scores or combination of scores along with predictive rules are considered to be likely to act as ceRNAs. Some of the Machine learning algorithms employed in ceRNA prediction are Bayesian Classification, Random Forest, Artificial Neural Networks (ANN), Hidden Markow Model (HMM) and Support Vector Machine (SVM). The ceRNAs are predicted based on scoring methods like 'confidence score' used in TraceRNA [59] or DT hybrid algorithm used in CERNIA [60]. For evaluating the performance of machine learning tasks, different performance measures are used. Performance measures for a binary classifier derived from the confusion matrix (Table 2) are given in Table 3 [61]. Area under the curve (AUC) on the reception operating characteristic (ROC) curve, Gain and Lift, Entropy, F-Score are other commonly used measures.

Data type	Classified as positive	Classified as negative
Positive	true positive (tp)	false negative (fn)
Negative	false positive (fp)	true negative (tn)

The matrix gives the performance of the problem. The data type denotes the actual positives and the negatives.

Table 2: Confusion matrix.

Performance measures	Calculation
Accuracy	$\frac{tp+tn}{tp+fn+fp+tn}$
Positive predicted value or precision	$\frac{tp}{(tp+fp)}$
Sensitivity (recall)	$\frac{tp}{(tp+fn)}$
Specificity	$\frac{tn}{(tn + fp)}$
Negative predicted value	$\frac{tn}{(fn+tn)}$

Different performance measures and calculations

Table 3: Performance Measures.

ceRNA Prediction Tools

Number of ceRNA prediction tools and ceRNA pathway analysis platforms are available online. For enhancing reliability of the prediction, ceRNA prediction tools incorporate results obtained from one or combination of miRNA target prediction tools. The features like seed matching, conservation and binding energy are used in common in ceRNA prediction. Selection of miRNA target prediction tools incorporated in ceRNA prediction are based on good performance and parameters considered for prediction. The details of the ceRNA prediction tools with input data type, features considered and resources and tools used for analysis are given in Table 4.

Transcriptome wide ceRNA discovery tool (TraceRNA) focuses on the prediction of ceRNAs in 3 genes, PTEN, ESR1/BRCA1. It utilises its local copy of validated miRNA:target pairs from miRTarBase and pre-calculated predictions from SVMicro, BCMicro and SiteTest (algorithm developed inspired by MuTaMe). First the user has to input the GOI, for which the miRNAs are known. The miRNAs targeting GOI are identified from the curated database. Next step is to predict the targets of these miRNAs. The average of the sequence-pairing scores, S of each miRNAs targeting GOI and the mRNA is calculated. From these scored targets the ceRNAs are predicted by calculating the probability (P) value by Fisher transformation. This P value is consolidated with the P value obtained by co-expression test using Borda counting method, which essentially sums ranks of scores [59]. NetceRNA is an extension of TraceRNA to find an optimized network representation. By the analysis of gene expression data in glioblastoma, over 248,000 microRNA (miR)-mediated interactions were identified with ~7,000 genes with miRNA sponge activity. Biochemical analyses in cell lines have confirmed that these interactions mediate crosstalk between canonical oncogenic pathways [62]. The study was done with the multivariate analysis method Hermes [62,63] which infers ceRNA interactions from expression profiles by using conditional mutual information.

The ceRNA prediction tool, Mutually Targeted MRE enrichment (MuTaMe) validated the existence of *SERINC1*, *VAPA* and *CNOT6L* as bona fide PTEN ceRNAs and established the significance of ceRNA mechanism in cancer [38]. This tool is also used in the in silico prediction of *ZEB2* mRNA as a PTEN ceRNA, and its involvement in tumor progression [35]. MuTaMe identifies targets that share MREs of same miRNAs. First MREs in mRNAs that are targeted by PTEN-targeting miRNAs are identified using miRNA target prediction algorithm, RNA22 [64]. RNA22 implements the pattern recognition algorithm, *Teiresias* [65]. For the predicted MREs, MuTaMe scores are given depending on the number of miRNAs it shares with the mRNA, number of MREs predicted in X for i-th miRNA and the width of the span they cover, density and distribution of the predicted MREs and the number of MREs predicted. If the mRNA is targeted by at least 7 of the 10 validated PTEN-targeting miRNAs and all predicted MREs occur in the candidate ceRNA's 3' UTR then also the mRNAs are considered as candidate ceRNA.

CeRNA prediction Tool	Features	Input Data	Databases
TraceRNA	Seed region complementarity, sequence conservation, binding free energy	sequence features	miRTarBase
MuTaME	pattern identification and folding energy	expression profile and sequence features	-
HERMES	-	expression profile	TCGA (The Cancer Genome Atlas)
CUPID	free energy, nucleotide complementarity, evolutionary conservation, seed match and accessibility energy	sequence features and miRNA-mRNA interaction data	TarBase (Papadopoulos et al.) [90], TRANSFAC (Matys et al.) [88]. miRecords
CERNIA	nucleotide complementarity, free energy and evolutionary conservation	expression profiles, sequence features and miRNA-mRNA interaction data	miRTarBase, starBase, miRecords, TCGA
CEFINDER	seed region and free energy	sequence features and miRNA-mRNA interaction data	TargetScan

Overview of different ceRNA prediction tools

Table 4: CeRNA prediction tools.

Platforms/ Resources	Data sources	Purpose
StarBase (Li et al.) [86]	TargetScan, PITA, PicTar, RNA22 and miRanda	predict the function of ncRNAs (miRNAs, lncRNAs, pseudogenes) and protein-coding genes from the miRNA-mediated regulatory networks.
CeRDB (Sarver & Subramanian) [68]	Targetscan	Predict ceRNAs
NetceRNA (Flores et al.) [17]	miRTarBase	finds an optimized ceRNA network representation
PceRBase (Yuan et al.) [94]	Phytozome10 (Goodstein et al.) [82], TAIR10 (Lamesch et al.) [84] and MSU RGCP7 (Ouyang et al.) [89]	Potential significant ceRNA target-target and target-mimic pairs for 26 plant species
Pan-ceRNADB (Xu et al.) [42]	TargetScan, PITA, PicTar, RNA22 and miRanda	mRNA related ceRNA-ceRNA cross talk, for biomedical scientists
Bosia et al. [78]	miRNA - mRNA interaction	analyze the equilibrium and out-of-equilibrium properties of mRNA-miRNA interaction in ceRNA networks
Ala et al. [76]	TargetScan	determine the optimal conditions for ceRNA activity in-silico
Shao et al. [92]	expression profiles in cancer and normal tissues	diagnostic biomarkers
Chiu et al. [67]	Cancer Genome Atlas (TCGA)	quantified the optimal conditions for ceRNA regulation

Overview of ceRNA resources with source of data and purpose

Table 5: Different ceRNA resources.

CUPID integrates scores from Miranda [66] TargetScan [50] and PITA [55]. Information regarding miRNA and putative targets and the likelihoods of each predictive feature are integrated and the predictions are done using a support vector machine (SVM) algorithm. The tool then checks whether the predicted targets act as ceRNAs. In an evaluation of seven miRNA target prediction tools [26], it is found that TargetScan has highest sensitivity and PicTar [51] highest specificity. Since TraceRNA incorporates results from both TargetScan and PicTar we can infer that TraceRNA also have high sensitivity and specificity. Since CUPID uses results from TargetScan it can have high sensitivity. Another tool CeRNA prediction algorithm (CERNIA) predicts ceRNAs by applying DT-Hybrid recommendation algorithm. For each pair, MREs and hybridization energy are found using miRanda, MuTaME scores and DT-Hybrid recommendation score. Along with these scores the correlations between gene expression values for a specific tissue type was added to form a vector of seven scores. Then by applying SVM a subset of the gene pairs are predicted as putative ceRNAs [60]. CEFINDER predicts ceRNA from conserved human miRNA-mRNA interactions derived from TargetScan, by converting the interactions into a matrix of '1's and '0's. The presence and absence of predicted conserved miRNA-mRNA interaction is denoted by '1' and '0' respectively in the matrix. A shuffled matrix is also generated and interaction score is obtained from both the real and shuffled matrix and are used to sort the predicted results [67,68]. In a recent study to find the sequence features which is responsible for the ceRNA activity, number and spatial distribution of binding sites of genes in the PTEN network were used. By probabilistic approach and implementing hyper-geometric test TNRC6B was predicted as a ceRNA of PTEN

[69]. Online resources and network analysis platforms available to analyze ceRNAs are given in Table 5.

Conclusion

The computational prediction of ceRNAs from miRNA targets has improved over the last decade and many new tools have emerged showing better prediction capability. Most of the tools has implemented TargetScan, miRanda or PicTar. The recently developed miRNA target prediction tool MBSTAR claim to outperform TargetScan, miRanda, MirTarget2 and SVMicrO with an area under curve (AUC) of 0.71 at the target level and highest F-Score (harmonic mean of positive predictive value and sensitivity) of 0.337 in the binding level prediction. Another miRNA target prediction tool, TarPmiR [70] is developed by adding seven new features. TarPmiR shows good performance than miRanda, different versions of TargetScan [71,72] and miRmap [73]. In another study by Wang [74] full spectrum of sequence features are integrated for the prediction of clinically relevant functional microRNA-mRNA interactions. This tool also outperforms other existing tools. ceRNA prediction can be improved by incorporating novel features used in these tools. Moreover studies reveal that combinations of target prediction tools have different level of performance. In a study by [75] the union of Target Scan and MiRanda-mirSVR showed good performance in terms of specificity and precision, and the union of TargetScan, MiRand-mirSVR and RNA22 offered remarkable sensitivity. Hence there is a scope for experimenting with different combination of tools in enhancing ceRNA prediction accuracy. Moreover, in [18], five miRNA target prediction tools TargetScan, miRanda, PicTar, PITA and RNA22 were used to analyze ceRNA-ceRNA interactions and better

performance was obtained for an ensemble of minimum of four tools. When different methods was analyzed, integrated expression profile methods such as Hypergeometric test combined with coexpression based prediction-HyperC, and Conditional Mutual Information (CMI) showed better performance than Significant Correlation (SC), Ratio based and Hypergeometric test. The limitations such as small number of samples and lack of any standards for validating the methods are reported in the analysis. Numerous miRNAs are validated as PTEN regulators and the expression of PTEN is altered in a wide spectrum of human cancers. Hence currently major research in ceRNA are concentrated on a few human's transcripts especially PTEN. The studies can be carried on to other genes which are specific to diseases. As ceRNAs play key role in carcinogenesis and other diseases and its interactions have impacts on molecular pathways, ceRNA analysis may reveal underlying mechanisms of diseases. The computationally predicted ceRNAs can be further investigated and validated by biochemical methods. CeRNA-ceRNA network analysis, the emerging area of research can provide insights into normal and abnormal cell mechanism and in turn can assist in prediction of potential targets in therapeutics [76-94].

Acknowledgment

Authors thank Prof. P. R. Sudhakaran (Emeritus Professor, Dept. of Computational Biology and Bioinformatics, University of Kerala) for his critical comments in preparing this review article.

References

- Gilbert W (1986) Origin of life: The RNA world. *Nature* 319: 618.
- Dinger ME, Pang KC, Mercer TR, Mattick JS (2008) Differentiating protein-coding and noncoding RNA: challenges and ambiguities. *PLoS Comput Biol* 4: e1000176.
- Gomes AQ, Nolasco S, Soares H (2013) Non-coding RNAs: multi-tasking molecules in the cell. *Int J Mol Sci* 14: 16010-16039.
- Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116: 281-297.
- Pillai RS (2005) MicroRNA function: multiple mechanisms for a tiny RNA?. *RNA* 11: 1753-1761.
- Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75: 843-854.
- Li Y, Kowdley KV (2012) MicroRNAs in common human diseases. *Genomics Proteomics Bioinformatics* 10: 246-253.
- Johnson CD, Esquela-Kerscher A, Stefani G (2007) The *let-7* microRNA represses cell proliferation pathways in human cells. *Cancer Res* 67: 7713-7722.
- Broderick JA, Zamore PD (2011) MicroRNA therapeutics. *Gene Ther* 18: 1104-1110.
- Ebert MS, Neilson JR, Sharp PA (2007) MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nature Methods* 4: 721-726.
- Seitz H (2009) Redefining microRNA targets. *Curr Biol* 19: 870-873.
- Thomson DW, Bracken CP, Goodall GJ (2011) Experimental strategies for microRNA target identification. *Nucleic Acids Res* 39: 6845-6853.
- Singh NK (2017) miRNAs target databases: developmental methods and target identification techniques with functional annotations. *Cell Mol Life Sci* 74: 2239-2261.
- Kartha RV, Subramanian S (2014) Competing endogenous RNAs (ceRNAs): new entrants to the intricacies of gene regulation. *Front Genet* 5: 8.
- de Giorgio A, Krell J, Harding V, Stebbing J, Castellano L, et al. (2013) Emerging roles of competing endogenous RNAs in cancer: insights from the regulation of PTEN. *Mol Cell Biol* 33: 3976-3982.
- Le TD, Zhang J, Liu L, Li J (2016) Computational methods for identifying miRNA sponge interactions. *Brief Bioinform* 18: 577-590.
- Flores M, Huang Y, Chen Y (2013) NetceRNA: An algorithm for construction of phenotype-specific regulation networks via competing endogenous RNAs. 2013 IEEE International Workshop on Genomic Signal Processing and Statistics Pp: 24-27.
- Li Y, Jin X, Wang Z, Li L, Chen H, et al. (2017) Systematic review of computational methods for identifying miRNA-mediated RNA-RNA crosstalk. *Brief Bioinform*.
- Wahid F, Shehzad A, Khan T, Kim YY (2010) MicroRNAs: synthesis, mechanism, function, and recent clinical trials. *BBA - Molecular Cell Research* 1803: 1231-1243.
- Zhao S, Liu MF (2009) Mechanisms of microRNA-mediated gene regulation. *Science in China Series C. Life Sci* 52: 1111-1116.
- Vasudevan S, Steitz JA (2007) AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. *Cell* 128: 1105-18.
- Christopher AF, Kaur RP, Kaur G, Kaur A, Gupta V, et al. (2016) MicroRNA therapeutics: Discovering novel targets and developing specific therapy. *Perspect Clin Res* 7: 68.
- Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, et al. (2002) Prediction of plant microRNA targets. *Cell* 110: 513-520.
- Medina PP, Slack FJ (2008) Micromnas and cancer: An overview. *Cell Cycle* 7: 2485-2492.
- Ebert MS, Sharp PA (2012) Roles for microRNAs in conferring robustness to biological processes. *Cell* 149: 515-524.
- Fan X, Kurgan L (2014) Comprehensive overview and assessment of computational prediction of microRNA targets in animals. *Brief Bioinform* 16: 780-794.
- Min H, Yoon S (2010) Got target?: computational methods for microRNA target prediction and their extension. *Exp Mol Med* 42: 233-244.
- Yue D, Liu H, Huang Y (2009) Survey of computational algorithms for microRNA target prediction. *Curr Genom* 10: 478-492.
- Akhtar MM, Micolucci L, Islam MS, Olivieri F, Procopio AD, et al. (2015) Bioinformatic tools for microRNA dissection. *Nucleic Acids Res* 44: 24-44.
- Karagkouni D, Paraskevopoulou MD, Chatzopoulos S, Vlachos IS, Tastsoglou S, et al. (2017) DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA-gene interactions. *Nucleic Acids Res* 46: D239-D245.
- Hsu SD, Lin FM, Wu WY (2010) miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res* 39: D163-D169.
- Xiao F, Zuo Z, Cai G, Kang S, Gao X, et al. (2009) miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res* 37: D105-D110.
- Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP, et al. (2011) A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?. *Cell* 146: 353-358.
- Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, et al. (2007) Target mimicry provides a new mechanism for regulation of microRNA activity. *Nature genetics* 39: 1033-1037.
- Karreth FA, Tay Y, Perna D, Ala U, Tan SM, et al. (2011) *In vivo* identification of tumor-suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. *Cell* 147: 382-395.
- Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, et al. (2010) A coding-independent function of gene and pseudogene mRNAs regulates tumor biology. *Nature* 465: 1033-1038.
- Jeyapalan Z, Deng Z, Shatseva T, Fang L, He C, et al. (2010) Expression of CD44 3'-untranslated region regulates endogenous microRNA functions in tumorigenesis and angiogenesis. *Nucleic Acids Res* 39: 3026-3041.
- Tay Y, Kats L, Salmena L, Weiss D, Tan SM, et al. (2011) Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell* 147: 344-357.
- Sen R, Ghosal S, Das S, Balti S, Chakrabarti J, et al. (2014) Competing endogenous RNA: The key to Posttranscriptional Regulation. *Sci World J* 2014: 1-6.
- Zhou X, Liu J, Wang W (2014) Construction and investigation of breast-cancer-specific ceRNA network based on the mRNA and miRNA expression data. *IET Syst Biol* 8: 96-103.

41. Cheng DL, Xiang YY, Ji LJ, Lu XJ (2015) Competing endogenous RNA interplay in cancer: mechanism, methodology, and perspectives. *Tumor Biol* 36: 479-488.
42. Xu J, Li Y, Lu J, Pan T, Ding N, et al. (2015) The mRNA related ceRNA-ceRNA landscape and significance across 20 major cancer types. *Nucleic Acids Res* 43: 8169-8182.
43. Poliseno L, Pandolfi PP (2015) PTEN ceRNA networks in human cancer. *Methods* 77: 41-50.
44. Hamberg M, Backes C, Fehlmann T, Hart M, Meder B, et al. (2016) MiRTargetLink -miRNAs, genes and interaction networks. *Int J Mol Sci* 17: 564.
45. Dweep H, Sticht C, Pandey P, Gretz N (2011) miRWalk-database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. *J Biomed Inform* 44: 839-847.
46. Kozomara A, Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 42: D68-D73.
47. Bandyopadhyay S, Mitra R (2009) TargetMiner: microRNA target prediction with systematic identification of tissue-specific negative examples. *Bioinformatics* 25: 2625-2631.
48. Xiao J, Li Y, Wang K, Wen Z, Li M, et al. (2009) In silico method for systematic analysis of feature importance in microRNA-mRNA interactions. *BMC Bioinformatics* 10: 427.
49. Peterson SM, Thompson JA, Ufkin ML, Sathyanarayana P, Liaw L, et al. (2014) Common features of microRNA target prediction tools. *Front Genet* 5: 23.
50. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB, et al. (2003) Prediction of mammalian microRNA targets. *Cell* 115: 787-798.
51. Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, et al. (2005) Combinatorial microRNA target predictions. *Nature Genetics* 37: 495-500.
52. Chandra V, Girijadevi R, Nair AS, Pillai SS, Pillai RM, et al. (2010) MTar: a computational microRNA target prediction architecture for human transcriptome. *BMC bioinformatics* 11.
53. Long D, Lee R, Williams P, Chan CY, Ambros V, et al. (2007) Potent effect of target structure on microRNA function. *Nat Struct Mol Biol* 14: 287-294.
54. Liu H, Yue D, Chen Y, Gao SJ, Huang Y, et al. (2010) Improving performance of mammalian microRNA target prediction. *BMC Bioinformatics* 11: 476.
55. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E, et al. (2007) The role of site accessibility in microRNA target recognition. *Nature Genet* 39: 1278-1284.
56. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *J Mach Learn Res* 3: 1157-1182.
57. Saeyns Y, Inza I, Larrañaga P (2007) A review of feature selection techniques in bioinformatics. *Bioinformatics* 23: 2507-2517.
58. He X, Cai D, Niyogi P (2006) Laplacian score for feature selection. In *Advances in Neural Information Processing Systems* Pp: 507-514.
59. Flores M, Huang Y (2012) A new algorithm for predicting competing endogenous mas. *Proceedings 2012 IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS)* Pp: 118-121.
60. Sardina DS, Alaimo S, Ferro A, Pulvirenti A, Giugno R (2016) A novel computational method for inferring competing endogenous interactions. *Brief Bioinform* 18: 1071-1081.
61. Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. *Inf Process Manag* 45: 427-437.
62. Sumazin P, Yang X, Chiu HS, Chung WJ, Iyer A, et al. (2011) An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell* 147: 370-381.
63. Rigoutsos I, Floratos A (1998) Combinatorial pattern discovery in biological sequences: The TEIRESIAS algorithm. *Bioinformatics* 14: 55-67.
64. Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, et al. (2006) A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* 126: 1203-1217.
65. Rigoutsos I, Floratos A (1998) Combinatorial pattern discovery in biological sequences: The TEIRESIAS algorithm. *Bioinformatics* 14: 55-67.
66. Enright AJ, John B, Gaul U, Tuschl T, Sander C, et al. (2003) MicroRNA targets in *Drosophila*. *Genome biology* 5: R1.
67. Chiu HS, Llobet-Navas D, Yang X, Chung WJ, Iyer A, et al. (2015) Cupid: simultaneous reconstruction of microRNA-target and ceRNA networks. *Genome Res* 25: 257-267.
68. Sarver AL, Subramanian S (2012) Competing endogenous RNA database. *Bioinformatics* 8: 731-733.
69. Zarringhalam K, Tay Y, Kulkarni P, Bester AC, Pandolfi PP, et al. (2017) Identification of competing endogenous RNAs of the tumor suppressor gene PTEN: A probabilistic approach. *Sci Rep* 7: 7755.
70. Ding J, Li X, Hu H (2016) TarPmiR: a new approach for microRNA target site prediction. *Bioinformatics* 32: 2768-2775.
71. Friedman RC, Farh KKH, Burge CB, Bartel DP (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 19: 92-105.
72. Agarwal V, Bell GW, Nam JW, Bartel DP (2015) Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 4: e05005.
73. Vejnar CE, Zdobnov EM (2012) MiRmap: comprehensive prediction of microRNA target repression strength. *Nucleic Acids Res* 40: 11673-11683.
74. Wang Z, Xu W, Liu Y (2015) Integrating full spectrum of sequence features into predicting functional microRNA-mRNA interactions. *Bioinformatics* 31: 3529-3536.
75. Oliveira AC, Bovolenta LA, Nachtigall PG, Herkenhoff ME, Lemke N, et al. (2017) Combining results from distinct microRNA target prediction tools enhances the performance of analyses. *Front Genet* 8.
76. Ala U, Karreth FA, Bosia C, Pagnani A, Taulli R, et al. (2013) Integrated transcriptional and competitive endogenous RNA networks are cross-regulated in permissive molecular environments. *Proc Natl Acad Sci U S A* 110: 7154-7159.
77. Arancio W, Giordano C, Pizzolanti G (2013) A ceRNA analysis on LMNA gene focusing on the Hutchinson-Gilford progeria syndrome. *J Clinical Bioinformatics* 3: 2.
78. Bosia C, Pagnani A, Zecchina R (2013) Modelling competing endogenous RNA networks. *PLoS One* 8: e66609.
79. Chiu YC, Hsiao TH, Chen Y, Chuang EY (2015) Parameter optimization for constructing competing endogenous RNA regulatory network in glioblastoma multiforme and other cancers. *BMC Genomics* 16: S1.
80. Ding Y, Wang ZC, Zheng Y, Hu Z, Li Y, et al. (2016) C-Myc functions as a competing endogenous RNA in acute promyelocytic leukemia. *Oncotarget* 7: 56422-56430.
81. Fang L, Du WW, Yang X, Chen K, Ghanekar A, et al. (2013) Versican 3'-untranslated region (3'-UTR) functions as a ceRNA in inducing the development of hepatocellular carcinoma by regulating miRNA activity. *FASEB J* 27: 907-919.
82. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, et al. (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* 40: D1178-D1186.
83. Kumar MS, Armenteros-Monterroso E, East P, Chakravorty P, Matthews N, et al. (2014) HMG2A functions as a competing endogenous RNA to promote lung cancer progression. *Nature* 505: 212-217.
84. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, et al., (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res* 40: D1202-D1210.
85. Lee DY, Jeyapalan Z, Fang L, Yang J, Zhang Y, et al. (2010) Expression of versican 3'-untranslated region modulates endogenous microRNA functions. *PLoS ONE* 5: e13599.
86. Li JH, Liu S, Zhou H, Qu LH, Yang JH, et al. (2013) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Research* 42: D92-D97.
87. Liu K, Guo L, Guo Y, Zhou B, Li T, et al. (2015) AEG-1 3'-untranslated region functions as a ceRNA in inducing epithelial-mesenchymal transition of human non-small cell lung cancer by regulating miR-30a activity. *Eur J Cell Biol* 94: 22-31.
88. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, et al. (2006) TRANSFAC® and its module TRANSCOMP®: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* 34: D108-D110.

89. Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, et al. (2007) The TIGR rice genome annotation resource: improvements and new features. *Nucleic Acids Res* 35: D883-D887.
90. Papadopoulos GL, Reczko M, Simossis VA, Sethupathy P, Hatzigeorgiou AG, et al. (2008) The database of experimentally supported targets: a functional update of TarBase. *Nucleic Acids Res* 37: D155-D158.
91. Rutnam ZJ, Yang BB (2012) The non-coding 3' UTR of CD44 induces metastasis by regulating extracellular matrix functions. *J Cell Sci* 125: 2075-2085.
92. Shao T, Wu A, Chen J, Chen H, Lu J, et al. (2015) Identification of module biomarkers from the dysregulated ceRNA-ceRNA interaction network in lung adenocarcinoma. *Mol Biosyst* 11: 3048-3058.
93. Yang J, Li T, Gao C, Lv X, Liu K, et al. (2014) FOXO1 3' UTR functions as a ceRNA in repressing the metastases of breast cancer cells via regulating miRNA activity. *FEBS Letters* 588: 3218-3224.
94. Yuan C, Meng X, Li X, Illing N, Ingle RA, et al. (2017) PceRBase: a database of plant competing endogenous RNA. *Nucleic Acids Res* 45: D1009-D1014.