

ceRNAshiny: An Interactive R/Shiny App for Identification and Analysis of ceRNA Regulation

Yueqiang Song^{1†}, Jia Li^{1†}, Yiming Mao^{2*} and Xi Zhang^{1,3*}

¹State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai, China, ²Department of Thoracic Surgery, Suzhou Kowloon Hospital, School of Medicine, Shanghai Jiao Tong University, Suzhou, China, ³Department of Rehabilitation, Huashan Hospital, Fudan University, Shanghai, China

The competing endogenous RNA (ceRNA) network is a newly discovered posttranscriptional regulation that controls both physiological and pathological progresses. Increasing research studies have been pivoted on this theory to explore the function of novel non-coding RNAs, pseudogenes, circular RNAs, and messenger RNAs. Although there are several R packages or computational tools to analyze ceRNA networks, an urgent need for easy-to-use computational tools still remains to identify ceRNA regulation. Besides, the conventional tools were mainly devoted to investigating ceRNAs in malignancies instead of those in neurodegenerative diseases. To fill this gap, we developed ceRNAshiny, an interactive R/Shiny application, which integrates widely used computational methods and databases to provide and visualize the construction and analysis of the ceRNA network, including differential gene analysis and functional annotation. In addition, demo data in ceRNAshiny could provide ceRNA network analyses about neurodegenerative diseases such as Parkinson's disease. Overall, ceRNAshiny is a user-friendly application that benefits all researchers, especially those who lack an established bioinformatic pipeline and are interested in studying ceRNA networks.

Keywords: IncRNA, miRNA, ceRNA, R, Shiny

INTRODUCTION

The vast majority of the human genome is non-coding sequences whose transcripts without proteincoding capacity are named non-coding RNAs (ncRNAs) (Alexander et al., 2010). ncRNAs are now regarded as core regulators involved in gene transcription, epigenetic regulation, and posttranscriptional regulation which exert their effects on the occurrence, development, and diagnosis of diverse diseases (Esteller 2011). A hypothesis of how ncRNAs work has been proposed and gradually confirmed, which was named the competing endogenous RNAs (ceRNAs) network (Salmena et al., 2011). According to this theory, long non-coding RNAs (lncRNAs), pseudogenes, and circular RNAs (circRNAs) act as microRNA (miRNA) sponges via miRNA response elements (MREs) or messenger RNA (mRNA) binding sites to control the availability of endogenous miRNAs for binding to their target mRNAs, which can form a ceRNA network to modulate mRNA expression and regulate protein levels. These complex networks may provide multiple clues for unraveling the pathogenesis in diseases (Tay et al., 2014). Notably, as a representative of various types of ceRNAs, lncRNA-associated ceRNA networks might be eligible candidates as promising therapeutic targets.

OPEN ACCESS

Edited by:

Junpeng Zhang, Dali University, China

Reviewed by:

Juan Xu, Harbin Medical University, China Meng Huang, University of Tsukuba, Japan

*Correspondence:

Yiming Mao 253992240@qq.com Xi Zhang zx931215@163.com [†]These authors have contributed equally to this work

Specialty section:

This article was submitted to RNA Networks and Biology, a section of the journal Frontiers in Molecular Biosciences

> Received: 29 January 2022 Accepted: 13 April 2022 Published: 13 May 2022

Citation:

Song Y, Li J, Mao Y and Zhang X (2022) ceRNAshiny: An Interactive R/ Shiny App for Identification and Analysis of ceRNA Regulation. Front. Mol. Biosci. 9:865408. doi: 10.3389/fmolb.2022.865408

1

RNA Step by Step										
Input your expression file	Manual Heatmap plot Volcano plot Enrichment RNA classification Expression-based ceRNA prediction Sequence-based ceRNA prediction ceRNA prediction based on expression and se									
Example_array	Example									
& Download example	Here we mainly show the example of the expression profiling dataset GSE7621 from GPL370 array platform to identify ceRNA network 1. Input data: GSE76210 or GSE356666.zw 2. Easiby (lick the buttons from table panel to get corresponding results 3. NOTICE: Please check that your file types are the same as the ones you clicked on, e.g. all are Array types									
Input your sample information file Choose an example or your own data	Format of your files									
Example_array -	Array Data 1. Ubjad your file expression matrix, group list and annotation platform information									
★ Download example	2. Row names of the expression matrix are probe names 3. Column names are sample lists									
Input your gene ID information file Choose an example or your own data	Htigh-throughput RNA sequencing data 1. Upload your file: expression matrix and group list 2. Row names of the expression matrix are Ensembl numbers 3. Column names are sample lists									
Example_array -	Module									
▲ Download example	 Heatmap plot The module displays the based on expression and sequence The module displays the classification This module displays the results of differential genes Searchment This module displays the classification Searchment This module displays the classification The module allows users to module genes The module allows users to module genes The module allows users to restore dure eRNA networks, in which heRNA-miRNA-miRNA-miRNA mire module The module allows users to restore dure the ANA networks, in which heRNA-miRNA-miRNA-miRNA mire module The module allows users to restore dure eRNA networks, in which heRNA-miRNA-miRNA-miRNA mire module The module allows users to restore dure the eRNA networks, in which heRNA-miRNA-miRNA-miRNA module The module allows users to restore dure the eRNA networks, in which heRNA-miRNA-miRNA mire module The module allows users to restore dure the eRNA networks, in which heRNA-miRNA mire module The module all									

Due to the huge scale of ceRNA networks, the availability of computational methods has allowed theoretical construction of ceRNA networks which provides convincing evidence for further verification in vitro or in vivo (Yang et al., 2018). The development of different ceRNA-directed computational methods can be mainly categorized into two classes: 1) methods constructed by combining expression profiles and statistic indexes, such as the Pearson correlation coefficient (PCC) and mutual information (MI), sensitivity correlation (SI), multiple sensitivity correlation, conditional mutual information (CMI), intervention calculus when the DAG is absent (IDA), and liquid association (LA) (Lloyd 2000), and 2) mathematical models, such as the minimal model, stochastic model, mass-action model, coarse-grained model, and coarsegrained competition motif model (Le et al., 2017; Zhang et al., 2017, 2022). Additionally, a set of lncRNA-miRNA-mRNA pairs databases have been established (Li et al., 2014; Le et al., 2017), such as lnCeDB (Das et al., 2014), LncCeRBase (Pan et al., 2019), miRSponge (Wang et al., 2015), LncACTdb (Wang et al., 2022), ceRDB (Sarver Subramanian 2012), starBase (Li et al., 2014), HumanViCe (Ghosal et al., 2014), PceRBase (Yuan et al., 2017), Tarbase (Vergoulis et al., 2012), miRTarbase (Huang et al., 2020), miRecords (Xiao et al., 2009), miRWalk (Sticht et al., 2018), TargetScan (www.targetscan.org), miRanda (Betel et al., 2008), MicroCosm (Griffiths-Jones et al., 2008), PicTar (http:// www.pictar.org/), DIANA-microT (Vlachos et al., 2015), PITA (http://genie.weizmann.ac.il/pubs/mir07/mir07_data.html), and CLASH (Helwak Tollervey 2014). Some databases can function as not only a prediction tool to guide experiments but also a free hub which provides experimentally validated results.

Current computational methods are, respectively, biased in terms of the accuracy or sensitivity of prediction. The contents of different databases also vary with species, diseases, and tissues. At present, scholars have developed many R packages for constructing ceRNA networks, but all of them require proficiency in using R software (Li et al., 2018; Zhang et al., 2018, 2019, 2021; Wen et al., 2020). Meanwhile, the newly developed CeNet Omnibus, an R/Shiny-based application, is used to predict ceRNA network construction using different computational methods (Wen et al., 2021) with database information uploaded manually, which also analyzes the distribution of topological properties of networks. Then, ceRNAshiny, which we developed, has the advantage of being more convenient and reliable for students and researchers who majored in biology and medicine, especially people with limited experience in programing. Analysis processes in ceRNAshiny for sequencing data are currently recognized and meet basic requirements of users for sequencing data analysis. At the same time, ceRNAshiny can effectively classify large sequencing data by RNA types, which is important and convenient for distinguishing sequencing data containing various RNA types. In addition, ceRNAshiny provides predictions based on expression, sequence, and both. In terms of built-in databases, ceRNAshiny includes predictive and experimentally validated databases that, respectively, contain the human-sourced contents of multiple databases. Based on these two types of databases, users can obtain multiple ceRNA networks more easily, providing multiple options for subsequent data validation and wet experiments. Overall, ceRNAshiny could be a useful tool for people without enough time and a knowledge background to rapidly get results. With the support of the



R/Shiny framework, ccRNAshiny offers a web-based userfriendly interface for users to obtain the identification, analysis, and visualization of ccRNA regulation, such as differential gene analysis and functional annotation.

IMPLEMENTATION

ceRNAshiny is an R-based Shiny application constructed using various R packages, including reshape2 (https://rdocumentation. org/packages/reshape2/versions/1.4.3), igraph (Csardi Nepusz

2006), edgeR (Robinson et al., 2010), DESeq2 (Michael I Love and Anders, 2014), limma (Brophy et al., 1987; Ritchie et al., 2015), glmnet (Simon et al., 2011; Engebretsen and Bohlin 2019), yulab. utils (https://CRAN.R-project.org/package=yulab.utils), ggplot2 (Wickham et a.l, 2016), rvcheck (https://github.com/ GuangchuangYu/rvcheck), shiny (https://github.com/rstudio/ shiny/issues), shinythemes (https://rstudio.github.io/ shinythemes/), DT (https://github.com/rstudio/DT), pheatmap (https://CRAN.R-project.org/package=pheatmap), ReactomeRA (Yu and He 2016), and clusterProfiler (Yu et al., 2012). The Shiny R platform was deployed on the webserver to host the web

		GeneSymbol (logFC 0	AveExp	r) t	P.Value	adj.P.Val 🛛	B	gene_id	type
Classification	1	ABHD6	0.469569711614769	8.379821701825	67 3.06307611959754	0.00511183380424444	0.277164537465978	-2.07706221488626	ENSG00000163686	mRNA
The file type:	2	ABI3BP	1.49183517946416	2.636694072755	23 2.44672375513913	0.0216280867370625	0.379936057110052	-3.27545988952168	ENSG00000154175	mRNA
array	3	ABLIM2	0.739773797587426	5.309470100542	48 2.31295795649241	0.0290572519947098	0.410102286730639	-3.51614746012885	ENSG00000163995	mRNA
Show the type of differential genes:	4	ACER1	1.46411835281664	2.912626280777	73 3.31256272247195	0.00276560532462046	0.229600215749645	-1.55991580655499	ENSG00000167769	mRNA
	5	ACSF3	1.11435993238987	2.695654735162	39 2.82545170294537	0.00904218265874418	0.313918875966241	-2.55435200621994	ENSG00000176715	mRNA
mRNA •	6	ACTN4	0.542486618033587	8.745603138719	16 2.5059251591847	0.018935726950348	0.370368343207339	-3.16643047526507	ENSG00000130402	mRNA
The pyalue of differential	7	ACY1	0.640687921592149	2.415533930777	35 2.26713801867351	0.0320952702541374	0.423097262243607	-3.59669131895543	ENSG00000243989	mRNA
ies:	8	ADA	0.849736467834814	6.121899968266	09 3.33549321626935	0.0026119853756127	0.223335033021876	-1.51169079407298	ENSG00000196839	mRNA
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 The fold change of	9	ADAMTS1	0.727157137400445	8.478952396514	19 2.06316078835232	0.0494080083035261	0.473311913250846	-3.94252380702295	ENSG00000154734	mRNA
	11	AHSA1	0.770502949192885	9.588584101312	28 3.71806379759093	0.000992790167395576	0.18142666751414	-0.693728265807345	ENSG00000100591	mRNA
	12	AKR1C1	0.988454693117847	4.809385687978	78 2.83052357142856	0.00893429533969107	0.313056998856569	-2.54434444658754	ENSG00000187134	mRNA
ferential genes:	13	AKR1C2	0.988752405057126	8.263454553658	52 3.13321345682393	0.00430739240344556	0.264844139436429	-1.93317699698892	ENSG00000151632	mRNA
1.5 5	14	AKR1C3	1.00236318002367	7.493536761624	74 3.09159978715248	0.00476868821280696	0.271970511044489	-2.01869859931664	ENSG00000196139	mRNA
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5	15	ALDH1A3	0.842216344256011	4.570911792410	51 2.11523956175051	0.0443351233504723	0.458834269406269	-3.85627958032375	ENSG00000184254	mRNA
	16	ALOX12	1.0985830171632	3.732606419139	56 2.9297675925555	0.00705318782723056	0.29936421676491	-2.34688315853457	ENSG00000108839	mRNA
	17	ANAPC16	0.604835999969438	9.096409714550	78 3.98341272918879	0.000501478606152734	0.163597553212122	-0.115558468860183	ENSG00000166295	mRNA
ssification	18	ANKLE1	1.32234598983117	2.169708103122	17 2.33894650147595	0.0274529274420811	0.405152676126066	-3.4700230162651	ENSG00000160117	mRNA
e file type:	19	ANKRD11	0.456382870885436	7.566245794276	14 2.08367250500794	0.0473513881787554	0.466721823038117	-3.90872955332507	ENSG00000167522	mRNA
ray 👻	20	ANKRD24	1.28023263051476	3.381946374243	61 2.1985567676245	0.037182946730369	0.439819579902954	-3.71533841264385	ENSG0000089847	mRNA
ow the type of differential	21	ANKRD37	0.719021844029707	7.617270929632	57 2.43449686643468	0.0222263956421298	0.382498304719315	-3.29779122091776	ENSG00000186352	mRNA
nes:	Showi	ng 1 to 20 of 677 en:	tries Pro	evious 1 2	3 4 5	34 Next				
All										
mRNA	Ma	nual Heatmap j	plot Volcano plot	Enrichment	RNA classification Ex	pression-based ceRNA predi	iction Sequence-base	ed ceRNA prediction	ceRNA predicition base	d on exp
miRNA	Shov	20 🗸 entrie	25		Search:					
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1		GeneSymbol	logFC	AveEx	pr o	P.Value	adj.P.Val	B 0	gene_id 0	type
The fold change of differential genes:	924	MIR186	-0.65387372057723	5 4.313411840	439 -2.308214592342	6 0.0293591336221321	0.410419209490758	-3.52453183141203	ENSG00000207721	miRNA
	925	MIR22	-0.58864382653479	1 5.87411443733	183 -2.1193207718407	0.0439579599060057	0.458354365194396	-3.84945982019699	ENSG00000283824	miRNA
	926	MIR3185	-0.85472591718495	2 3.53535204868	521 -2.233134709387	0.0345331871647891	0.431388729697762	-3.65580823694263	ENSG00000263602	miRNA
	927	MIR4458	0.46231495195694	4 5.71476832072	868 2.3206441885687	3 0.028574080170868	0.408891747081337	-3.50253881070099	ENSG00000284000	miRNA
	928	MIR4657	0.39616770295283	4 8.61824895773	268 3.328330727128	0.00265906063948109	0.223605061057345	-1.52676555811392	ENSG00000284261	miRNA
	929	MIR4723	0.30000753684347	7.6421465963	949 2.1969565320448	0.0373099025565974	0.440445662256012	-3.71807891478575	ENSG00000284532	miRNA
	930	MIR4745	0.5741192410899	8 7.83347451856	934 2.4392507978725	0.0219919785864673	0.3822197430705	-3.28911628763457	ENSG00000284361	miRNA
	931	MIR636	-0.46775718523303	8 8.18598065025	385 -2.6427962451114	8 0.013855960952708	0.344779743321392	-2.90889258589822	ENSG00000283805	miRNA
	933	MIR6836	-2.0899494294703	5 3.40827725142	931 -3.4911914957676	4 0.00176718529427487	0.214651735017124	-1.18161417479585	ENSG00000284047	miRNA
	934	MIR6845	-0.52813169621699	9.12533284523	022 -2.7203553262968	4 0.0115746077597422	0.328187341418849	-2.75978295740296	ENSG00000284142	miRNA
	Show	ing 1 to 10 of 10 entr	ries		Previou	s 1 Novt				

application of ceRNAshiny. The human-sourced lncRNA-miRNA-mRNA pairing information included in ceRNAshiny was obtained from various databases (John et al., 2004; Lewis et al., 2005), which could fall into two categories: predicted databases, like starBase 3.0 (starBase 3.0 was also named as ENCORI; Li et al., 2014), miRWalk (Sticht et al., 2018), TargetScan (www.targetscan.org), and miRanda (Betel et al., 2008), and experimentally validated databases, like Tarbase (Vergoulis et al., 2012) and miRTarbase (Huang et al., 2020). To show the functionality and usability of ceRNAshiny, we used the publicly available array datasets (GSE7621) that were generated from the substantia nigra from the postmortem human brain of Parkinson's disease (PD) patients and control ones (Lesnick et al., 2007). Furthermore, GSE136666, which contains transcriptomic results of human substantia nigra and putamen samples from PD patients and age-matched controls, was chosen as template data of high-throughput RNA sequencing data (Xicoy et al., 2020). The app can be accessed here: https:// cerna.shinyapps.io/cerna_shiny/. The source code and

related documents can be obtained through GitHub: https://github.com/yqsongGitHub/ceRNA_shiny.

DATA INPUT

The ceRNAshiny app is more suitable for users who are interested in ceRNA but are limited by programing. For an individual user, the only step is to input the expression matrix, group list, and annotation platform information (optional) that are consistent with the format of template data (could be downloaded). Both array data and high-throughput RNA sequencing data can be acceptable and processible as the input expression matrix. If the input expression matrix is array data, it is necessary to synchronously provide annotation platform information, with probe names as row names of the matrix and sample lists as column names. In case the input expression matrix comprised high-throughput RNA sequencing data, Ensembl numbers should be provided as row names and sample lists as column names.



DATA PROCESSING AND ANALYSIS

For data processing, the following parameters are performed, including missing value imputation, log2 transformation, background adjustment, and quantile normalization. Cluster analysis results are presented in heat maps to allow users to remove low-quality samples. Depending on the category of input data, differential gene analysis, annotation analysis, and enrichment analysis can be performed using the corresponding R packages.

NETWORK CONSTRUCTION

Based on differentially expressed genes (DEGs), PCC, sensitivity partial Pearson correlation (SPPC), and the partial Pearson correlation (PC) algorithm are supposed to be employed for predicting potential lncRNA-miRNA-mRNA pairs, which can identify the maximum number of miRNA sponge interactions (Zhang et al., 2019). In addition, we downloaded and compared the relevant information in starBase 3.0 (Li et al., 2014), miRWalk (Sticht et al., 2018), TargetScan (www.targetscan. org), and miRanda (Betel et al., 2008) databases and aggregated miRNA-lncRNA pairs (63,556 pairs) and miRNA-mRNA pairs (1,441,765 pairs) recorded in these databases as the predicted database. For the experimentally validated database, we aggregated the contents of the Tarbase database (Vergoulis et al., 2012) and miRTarbase database (Huang et al., 2020), which finally contained miRNA-lncRNA pairs (1,506 pairs) and miRNA-mRNA pairs (652,703 pairs). Thus, we predicted ceRNA networks using three approaches. In the first approach, we analyzed statistical relationships of genes based on gene expression and different algorithms (PCC, SPPC, and PC) to identify potential lncRNA-miRNA-mRNA pairs. In the second approach, we compared uploaded data with the predicted and the experimentally verified database, database respectively, to get the potential lncRNA-miRNA-mRNA pairs based on the sequence. Then in the last approach, we intersected the results of the previous two steps to obtain lncRNA-miRNA-mRNA pairs that satisfied both requirements. Then the intersection of the aforementioned results would be output as more credible clues for subsequent experiments. Finally, enrichment analysis is going to be performed on the aforementioned ceRNA networks for biological functions and pathways.



RESULTS

Case Study: GSE7621 Expression Data of Substantia Nigra from a Postmortem Human Brain of Parkinson's Disease

For demonstration, we used the expression profiling dataset GSE7621 from the GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array platform (54317 probes per sample) to identify ceRNA networks in 25 cases with PD (Lesnick et al., 2007). In addition, the dataset GSE13666 from GPL11154 Illumina HiSeq 2000 (*Homo sapiens*) was chosen as the template data of high-throughput RNA sequencing data (Xicoy et al., 2020). Due to the similar process, the analysis process of GSE7621 is selected as the example.

After uploading the expression profiling data into the input table panel of ceRNAshiny (**Figure 1**A), data can be imported conveniently for further analyses. Through easily clicking the buttons from the table panel, users can perform the following editing: "Heatmap plot," "Volcano plot," "Enrichment," "RNA classification," "Expression-based ceRNA prediction," "Sequence-based ceRNA prediction," and "ceRNA prediction based on expression and sequence" (**Figure 1A**). The analyses indicated by these buttons are based on differential gene analysis. Sliders of the p value and fold change value are set up for users to adjust the p value and fold change value independently and to gain corresponding plots (volcano plot, heatmap plot, enrichment plots, etc.) and results (Figure 2). Moreover, ceRNAshiny can help users effectively lncRNAs, miRNAs, mRNAs distinguish and among multitudinous genes in the expression profiles (Figure 3A). It is worth noting that the correct type of uploaded data should be selected in the panel to avoid errors during these analyses (Figure 3B).

These three buttons in the interface, "Expression-based ceRNA prediction," "Sequence-based ceRNA prediction," and "ceRNA prediction based on expression and sequence," are designed for identifying ceRNA networks. Users can select the "Expression-based ceRNA prediction" module to construct and, respectively, download potential lncRNA-miRNA-mRNA pairs, whose gene expression conformed to statistical relationships based on different algorithms (PCC, SPPC, and PC) (Figures 4A-C). In addition, the "Sequence-based ceRNA prediction" module allows users to match genes in uploaded data with predicted and expression validated database resources to find potential lncRNA-miRNA pairs, based on sequence



(Figures 5A, B). Finally, through the "ceRNA prediction based on expression and sequence" module, users can create other ceRNA networks, in which lncRNA-miRNA-mRNA pairs simultaneously conform to both the PCC algorithm and sequence prediction (Figures 6A, B). The analysis is based on default parameters, which can be adjusted in accordance to requirements.

CONCLUSION AND OUTLOOK

ceRNAshiny is designed with a clear purpose to identify and visualize ceRNA networks for users with limited experience in programing. ceRNAshiny was developed for customizable generation of volcano plots, heatmap plots, ceRNA graphs, and other results using input expression datasets. Unlike conventional tools, ceRNAshiny not only identified the ceRNA networks using computational methods but also matched lncRNA-miRNA-mRNA pairs from multiple human source databases, eliminating the need for users to use online databases. With the basic R environment and an internet connection, this userfriendly Shiny application will be automatically set up and intuitively applied to visually review the reported results. We provide a downloadable source code and an offline version for researchers who are good at programming. Since the Shiny package was designed to build interactive web applications, it is straightforward to deploy ceRNAshiny on servers to provide an online service so that it can be used by researchers from a variety of backgrounds with ranging interests.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

YM and XZ designed the study. XZ collected and analyzed the mentioned datasets. YS, JL, and XZ performed bioinformatical analysis. YS and XZ wrote the manuscript. YM and XZ reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

- Alexander, R. P., Fang, G., Rozowsky, J., Snyder, M., and Gerstein, M. B. (2010). Annotating Non-coding Regions of the Genome. *Nat. Rev. Genet.* 11, 559–571. doi:10.1038/nrg2814
- Betel, D., Wilson, M., Gabow, A., Marks, D. S., and Sander, C. (2008). The microRNA.Org Resource: Targets and Expression. *Nucleic Acids Res.* 36, D149–D153. doi:10.1093/nar/gkm995
- Brophy, A. L. (1987). Efficient Estimation of Probabilities in Thet Distribution. Behav. Res. Methods Instr. Comput. 19, 462–466. doi:10.3758/bf03205616
- Csardi, G., and Nepusz, T. (2006). The Igraph Software Package for Complex Network Research. *InterJournal, Complex Syst.*, 1695. 1. https://igraph.org.
- Das, S., Ghosal, S., Sen, R., and Chakrabarti, J. (2014). lnCeDB: Database of Human Long Noncoding RNA Acting as Competing Endogenous RNA. *PloS one* 9 (6). e98965. doi:10.1371/journal.pone.0098965
- Engebretsen, S., and Bohlin, J. (2019). Statistical Predictions with Glmnet. *Clin. Epigenet* 11, 123. doi:10.1186/s13148-019-0730-1
- Esteller, M. (2011). Non-coding RNAs in Human Disease. Nat. Rev. Genet. 12, 861–874. doi:10.1038/nrg3074
- Ghosal, S., Das, S., Sen, R., and Chakrabarti, J. (2014). HumanViCe: Host ceRNA Network in Virus Infected Cells in Human. *Front. Genet.* 5, 249. doi:10.3389/ fgene.2014.00249
- Griffiths-Jones, S., Saini, H. K., van Dongen, S., and Enright, A. J. (2008). miRBase: Tools for microRNA Genomics. *Nucleic Acids Res.* 36 (Database issue), D154–D158. doi:10.1093/nar/gkm952
- Helwak, A., and Tollervey, D. (2014). Mapping the miRNA Interactome by Cross-Linking Ligation and Sequencing of Hybrids (CLASH). *Nat. Protoc.* 9 (3), 711–728. doi:10.1038/nprot.2014.043
- Huang, H.-Y., Lin, Y.-C. -D., Li, J., Huang, K.-Y., Shrestha, S., Hong, H.-C., Tang, Y., Chen, Y. G., Jin, C. N., Yu, Y., Xu, J. T., Li, Y. M., Cai, X. X., Zhou, Z. Y., Chen, X. H., Pei, Y. Y., Hu, L., Su, J. J., Cui, S. D., Wang, F., Xie, Y. Y., Ding, S. Y., Luo, M. F., Chou, C. H., Chang, N. W., Chen, K. W., Cheng, Y. H., Wan, X. H., Hsu, W. L., Lee, T. Y., Wei, F. X., and Huang, H. D. (2020). miR.TarBase 2020: Updates to the Experimentally Validated microRNA-Target Interaction Database. *Nucleic Acids Res.* 48 (D1), D148–D154. doi:10.1093/nar/gkz896
- John, B., Enright, A. J., Aravin, A., Tuschl, T., Sander, C., and Marks, D. S. (2004). Human MicroRNA Targets. *Plos Biol.* 2, e363. doi:10.1371/journal.pbio. 0020363
- Le, T. D., Zhang, J., Liu, L., and Li, J. (2017). Computational Methods for Identifying miRNA Sponge Interactions. *Brief. Bioinform.* 18, bbw042–590. doi:10.1093/bib/bbw042
- Lesnick, T. G., Papapetropoulos, S., Mash, D. C., Ffrench-Mullen, J., Shehadeh, L., de Andrade, M., Henley, J. R., Rocca, W. A., Ahlskog, J. E., and Maraganore, D. M. (2007). A Genomic Pathway Approach to a Complex Disease: Axon Guidance and Parkinson Disease. *Plos Genet.* 3, e98. doi:10.1371/journal. pgen.0030098
- Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes Are microRNA Targets. *Cell* 120, 15–20. doi:10.1016/j.cell.2004.12.035
- Li, J.-H., Liu, S., Zhou, H., Qu, L.-H., and Yang, J.-H. (2014). starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and Protein-RNA Interaction Networks from Large-Scale CLIP-Seq Data. *Nucl. Acids Res.* 42, D92–D97. doi:10.1093/nar/gkt1248
- Li, R., Qu, H., Wang, S., Wei, J., Zhang, L., Ma, R., Lu, J., Zhu, J., Zhong, W. D., and Jia, Z. (2018). GDCRNATools: an R/Bioconductor Package for Integrative Analysis of lncRNA, miRNA and mRNA Data in GDC. *Bioinformatics* 34, 2515–2517. doi:10.1093/bioinformatics/bty124

ACKNOWLEDGMENTS

We appreciate the support from Donglai Chen (Department of Thoracic Surgery, Zhongshan Hospital, Fudan University, Shanghai 200032, China), and we would like to thank the GEO database for providing high-quality biological data.

- Lloyd, A. (2000). Computational Methods in Molecular Biology. Brief. Bioinform. 1, 315–316. doi:10.1093/bib/1.3.315
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8
- Pian, C., Zhang, G., Tu, T., Ma, X., and Li, F. 2019. LncCeRBase: A Database of Experimentally Validated Human Competing Endogenous Long Non-coding RNAs, *Database (Oxford)* 2019. baz090. doi:10.1093/database/baz090
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., and Smyth, G. K. (2015). Limma powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* 43, e47. doi:10. 1093/nar/gkv007
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* 26, 139–140. doi:10.1093/bioinformatics/btp616
- Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. (2011). A ceRNA Hypothesis: the Rosetta Stone of a Hidden RNA Language? *Cell* 146, 353–358. doi:10.1016/j.cell.2011.07.014
- Sarver, A. L., and Subramanian, S. (2012). Competing Endogenous RNA Database. Bioinformation 8 (15), 731–733. doi:10.6026/97320630008731
- Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2011). "Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent." J. Stat. Softw., 39(5), 1–13. doi:10.18637/jss.v039.i05
- Sticht, C., De La Torre, C., Parveen, A., and Gretz, N. (2018). miRWalk: An Online Resource for Prediction of microRNA Binding Sites. *PLoS One* 13. e0206239. doi:10.1371/journal.pone.0206239
- Tay, Y., Rinn, J., and Pandolfi, P. P. (2014). The Multilayered Complexity of ceRNA Crosstalk and Competition. *Nature* 505, 344–352. doi:10.1038/ nature12986
- Vergoulis, T., Vlachos, I. S., Alexiou, P., Georgakilas, G., Maragkakis, M., Reczko, M., Gerangelos, S., Koziris, N., Dalamagas, T., and Hatzigeorgiou, A. G. (2012). TarBase 6.0: Capturing the Exponential Growth of miRNA Targets with Experimental Support. *Nucleic Acids Res.* 40 (Database issue), D222–D229. doi:10.1093/nar/gkr1161
- Vlachos, I. S., Zagganas, K., Paraskevopoulou, M. D., Georgakilas, G., Karagkouni, D., Vergoulis, T., Dalamagas, T., and Hatzigeorgiou, A. G. (2015). DIANAmiRPath v3.0: Deciphering microRNA Function with Experimental Support. *Nucleic Acids Res.* 43 (W1), W460–W466. doi:10.1093/nar/gkv403
- Wang, P., Zhi, H., Zhang, Y., Liu, Y., Zhang, J., Gao, Y., Guo, M., Ning, S., and Li, X. (2015). miRSponge: a Manually Curated Database for Experimentally Supported miRNA Sponges and ceRNAs. *Database* 2015, bav098. doi:10. 1093/database/bav098
- Wang, P., Guo, Q., Qi, Y., Hao, Y., Gao, Y., Zhi, H., Zhang, Y., Sun, Y., Zhang, Y., Xin, M., Zhang, Y., Ning, S., and Li, X. (2022). LncACTdb 3.0: an Updated Database of Experimentally Supported ceRNA Interactions and Personalized Networks Contributing to Precision Medicine. *Nucleic Acids Res.* 50 (D1), D183–D189. doi:10.1093/nar/gkab1092
- Wen, X., Gao, L., and Hu, Y. (2020). LAceModule: Identification of Competing Endogenous RNA Modules by Integrating Dynamic Correlation. *Front. Genet.* 11, 235. doi:10.3389/fgene.2020.00235
- Wen, X., Gao, L., Song, T., and Jiang, C. (2021). CeNet Omnibus: an R/Shiny Application to the Construction and Analysis of Competing Endogenous RNA Network. *BMC Bioinformatics* 22, 75. doi:10.1186/s12859-021-04012-y
- Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, Available at: https://ggplot2.tidyverse.org.
- Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X., and Li, T. (2009). miRecords: an Integrated Resource for microRNA-Target Interactions. *Nucleic Acids Res.* 37, D105–D110. doi:10.1093/nar/gkn851

- Xicoy, H., Brouwers, J. F., Wieringa, B., and Martens, G. J. M. (2020). Explorative Combined Lipid and Transcriptomic Profiling of Substantia Nigra and Putamen in Parkinson's Disease. *Cells* 9, 1966. doi:10.3390/cells9091966
- Yang, X.-Z., Cheng, T.-T., He, Q.-J., Lei, Z.-Y., Chi, J., Tang, Z., Liao, Q. X., Zhang, H., Zeng, L. S., and Cui, S. Z. (2018). LINC01133 as ceRNA Inhibits Gastric Cancer Progression by Sponging miR-106a-3p to Regulate APC Expression and the Wnt/ β-Catenin Pathway. *Mol. Cancer* 17, 126. doi:10.1186/s12943-018-0874-1
- Yu, G., and He, Q.-Y. (2016). ReactomePA: an R/Bioconductor Package for Reactome Pathway Analysis and Visualization. *Mol. Biosyst.* 12, 477–479. doi:10.1039/c5mb00663e
- Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118
- Yuan, C., Meng, X., Li, X., Illing, N., Ingle, R. A., Wang, J., and Chen, M. (2017). PceRBase: a Database of Plant Competing Endogenous RNA. *Nucleic Acids Res.* 45 (D1), D1009–D1014. doi:10.1093/nar/gkw916
- Zhang, J., Le, T. D., Liu, L., and Li, J. (2017). Identifying miRNA Sponge Modules Using Biclustering and Regulatory Scores. BMC Bioinformatics 18, 44. doi:10. 1186/s12859-017-1467-5
- Zhang, J., Liu, L., Li, J., and Le, T. D. (2018). LncmiRSRN: Identification and Analysis of Long Non-coding RNA Related miRNA Sponge Regulatory Network in Human Cancer. *Bioinformatics* 34, 4232–4240. doi:10.1093/ bioinformatics/bty525
- Zhang, J., Liu, L., Xu, T., Xie, Y., Zhao, C., Li, J., et al. (2019). miRspongeR: an R/Bioconductor Package for the Identification and Analysis of miRNA Sponge

Interaction Networks and Modules. BMC Bioinformatics 20, 235. doi:10.1186/s12859-019-2861-y

- Zhang, M., Jin, X., Li, J., Tian, Y., Wang, Q., Li, X., Xu, J., Li, Y., and Li, X. (2021). CeRNASeek: an R Package for Identification and Analysis of ceRNA Regulation. *Brief. Bioinform.* 22, bbaa048. doi:10.1093/bib/bbaa048
- Zhang, J., Liu, L., Xu, T., Zhang, W., Li, J., Rao, N., and Le, T. D. (2022). Time to Infer miRNA Sponge Modules. WIREs RNA 13. e1686. doi:10.1002/wrna.1686

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Song, Li, Mao and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.