



OPEN

Cerina: systematic circRNA functional annotation based on integrative analysis of ceRNA interactions

Jacob Cardenas^{1,2}, Uthra Balaji^{1,2} & Jinghua Gu¹✉

Circular RNAs, a family of covalently circularized RNAs with tissue-specific expression, were recently demonstrated to play important roles in mammalian biology. Regardless of extensive research to predict, quantify, and annotate circRNAs, our understanding of their functions is still in its infancy. In this study, we developed a novel computational tool: Competing Endogenous RNA for INtegrative Annotations (Cerina), to predict biological functions of circRNAs based on the competing endogenous RNA model. Pareto Frontier Analysis was employed to integrate ENCODE mRNA/miRNA data with predicted microRNA response elements to prioritize tissue-specific ceRNA interactions. Using data from several circRNA-disease databases, we demonstrated that Cerina significantly improved the functional relevance of the prioritized ceRNA interactions by several folds, in terms of precision and recall. Proof-of-concept studies on human cancers and cardiovascular diseases further showcased the efficacy of Cerina on predicting potential circRNA functions in human diseases.

Circular RNAs (circRNAs) are a family of RNAs that form circular structures by joining the 3' and 5' ends covalently. Although originally considered as by-products of “splicing noise”^{1,2}, researchers have recently discovered that circRNAs play important roles in human diseases, including cancers, neurological diseases, heart and vascular diseases, among many others. CircRNAs are highly stable and display tissue-specific expression patterns, making them promising candidates as disease biomarkers³⁻⁷.

Despite the rapid growth in cataloging new circRNAs, their biological functions in human diseases are yet largely unknown. Among many putative mechanisms, such as interaction with RNA binding proteins (RBP), alternative splicing competition, posttranscriptional gene regulation, and protein coding, one of the most well-studied circRNA functions is to act as the competing endogenous RNA (ceRNA) or miRNA “sponge”^{3,7-13}. In the ceRNA model^{14,15}, linear RNAs and circRNAs competitively interact with miRNAs through miRNA response elements (MREs) to leverage the amount of active miRNAs in a cell, which has been extensively demonstrated in a variety of diseases^{6,16,17}. One notable ceRNA examples is CDR1as, a brain-enriched circRNA that is found to function as miRNA sponge for miR-7^{18,19} in various human diseases, including colon cancer²⁰, gastric cancer²¹, esophageal cancer²², and myocardial infarction²³. A number of other ceRNA interactions have also been uncovered in cancer, such as circPVT1-miR125 in gastric cancer²⁴, circTCH-miR7/miR214 in lung cancer²⁵, circHIPK3-miR124 in liver cancer²⁶, and circTTBK2-miR217 in glioma²⁷. Other than cancer, the role of circRNA as microRNA sponge is also under heavy investigation in cardiovascular diseases^{28,29}. In addition to beforementioned CDR1as-miR7 interaction in myocardial infarction, new theories have emerged hypothesizing the involvement of various circRNAs in multiple cardiovascular diseases through sequestration of miRNAs, exemplified by CircRNA_081881-miR548³⁰ and MFCAR-miR652³¹ interactions in ischemia/reperfusion injury and myocardial infarction, sponging effect and therapeutic potential of circRNA_000203 and circRNA_010567 in cardiac fibrosis^{32,33}, and protective effect of circRNA HRCR against hypertrophy and heart failure by sequestering miR223³⁴. It was also reported that regulation of disordered vascular smooth muscle cell proliferation and migration through the circWDR77-miR124-FGF2 axis was vital in atherosclerosis pathogenesis³⁵. In neurological disease research, in addition to the prominent role of CDR1as-miR7 sponging events in Alzheimer's disease^{36,37}, hundreds of circRNA were recently identified from multiple high-throughput studies to investigate circRNA-miRNA-mRNA interactions related to AD pathogenesis^{38,39}.

¹Baylor Scott & White Research Institute, Dallas, TX 75204, USA. ²These authors contributed equally: Jacob Cardenas and Uthra Balaji. ✉email: Jinghua.Gu@BSWHealth.org

Recently, there has been a surge of interest in annotating circRNA functions. Databases such as CircInteractome⁴⁰, CircAtlas⁴¹, Circ2Traits⁴², CircNet⁴³, TSCD⁴⁴, CSCD⁴⁵, and Circbank⁴⁶, among others, have collected predicted MRE sites to bridge individual circRNAs to their potential functions through miRNAs. Some of these resources also report circRNA-interacting RBPs, due to their roles in circRNA formation, translation, targeted gene regulation, and transport⁴⁷. Additionally, Circ2Traits annotated circRNAs that harbored disease-related single nucleotide polymorphisms (SNPs) as putative evidence for circRNA-disease associations⁴². CircRNADb compiled detailed information regarding internal ribosomal entry site (IRES) and open reading frame (ORF) to implicate possible protein-coding potential of circRNAs⁴⁸. Recently, CircFunBase has reported a collection of more than 7,000 manually curated circRNAs on 15 different species, which is among the first to systematically summarize circRNA functions based on circRNA differential expression data⁴⁹.

Amid the existing efforts to annotate individual circRNAs and their potential functions, tools for systematic circRNA functional annotation and pathway analysis are still lacking. In this work, we developed a novel computational tool for circRNA functional analysis: Competing Endogenous RNA for INtegrative Annotations (Cerina). As the first statistical method for systematic circRNA functional analysis, Cerina has several major technical advances. Firstly, Cerina paired up circRNA, linear RNA, and miRNA expression data for 11 human organs from ENCODE⁵⁰ and jointly analyzed them for the first time, allowing comprehensive pan-tissue profiling of ceRNA expressions. Secondly, expression and binding data for ceRNAs are integrated and prioritized based on the principles of Pareto optimality, which further increased the biological relevance of predicted ceRNA interactions. Finally, a user-friendly, web-based interface is made available for users to query a circRNA and retrieve its interacting miRNAs, their significant target genes, and the enriched biological functions and pathways.

Methods

Processing of sequencing data. 29 total RNA-Seq samples and 39 miRNA-Seq samples from 11 ENCODE tissues were analyzed in this study (Supplementary Table S1). Fastq files from replicate samples were merged before downstream analysis.

ENCODE total RNA-Seq data (linear RNA). Sequencing quality control of ENCODE total RNA-Seq data were performed by FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Adapter sequences were trimmed and low-quality reads (<20) were filtered using cutadapt⁵¹. Reads from total RNA sequencing were aligned to human genome (hg19/GRCh37) using hisat2⁵² and converted to BAM format using samtools⁵³. The featureCounts⁵⁴ tool was used to assign reads to each gene in the GENCODE⁵⁵ GTF file (https://www.genCODE.org/human/release_19.html). Counts from the total RNA-Seq pipeline were used to approximate linear RNA gene expression. After filtering out low-expressing genes (total read counts < 10), DESeq2⁵⁶ was used for count data normalization. A final layer of filtering was applied by removing genes with mean normalized counts less than 10. Counts per million (CPM) is used to report and visualize mRNA expression levels.

ENCODE total RNA-Seq data (circular RNA). We used back-splicing (BS) junction reads to approximate circRNA gene expression from ENCODE total RNA-Seq data. It was previously reported that individual circRNA detection method suffered from high false positives in BS junction prediction, yet combining results from two different algorithms (i.e., intersection) is a simple and effective remedy to significantly reduce false positives⁵⁷. In a recent review paper⁵⁸ comparing a dozen circRNA detection algorithms, CIRI2^{59,60} and CIRCexplorer⁶¹ demonstrated their best overall performance in terms of accuracy and efficiency based on the simulated and real RNaseR + data sets. Hence, we developed a framework that combines predictions from both CIRI2 and CIRCexplorer to improve the specificity of circRNA detection (Supplementary Methods). Spliced reads per billion mapped reads (SRPBM) is used to report and visualize circRNA expression levels.

ENCODE miRNA-Seq. The extra-cellular RNA processing toolkit (exceRpt)⁶² was used to process ENCODE miRNA sequencing data. Firstly, exceRpt filtered reads were mapped to UniVec vectors and ribosomal RNA sequences, followed by alignment of the remaining reads to the human genome (hg19) and then quantified for different types of RNAs, including miRNAs. The same normalization and filtering criteria were applied to process ENCODE miRNA data. CPM is used to report and visualize miRNA expression levels.

Prediction of miRNA binding sites on linear/circular RNA. Among a total of 33,461 confidently detected circRNA from ENCODE total RNA-Seq data, 30,282 had mature sequences (hg19) from circAtlas⁴¹ available for download at http://159.226.67.237/zhao/Data/circAtlas_supply/human_sequence_v1909.txt.zip. Mature sequences of novel circRNAs were estimated using a hierarchical framework described in the Supplementary Methods.

Perl script from TargetScan 7.1⁶³ was used to identify MREs based on circRNA mature sequences. For each circRNA, the number of MREs was normalized by the length of its mature splice sequence and defined as the circRNA-miRNA binding scores $S_{MRE}^{circ|mir}$.

miRNA-gene (i.e., linear RNA) binding data were download from TargetScan 7.2 (http://www.targetscan.org/vert_72/vert_72_data_download/Conserved_Family_Info.txt.zip, http://www.targetscan.org/vert_72/vert_72_data_download/Nonconserved_Family_Info.txt.zip) and used to form miRNA-gene scores, $S_{MRE}^{mir|gene}$. In addition to TargetScan, miRTarBase 7.0⁶⁴, a curated database of experimentally validated miRNA-target gene interactions ($S_{MTB}^{mir|gene}$) was used as another component of evidence. For each miRNA-gene pair, the number of MREs from

TargetScan and the number of publications from miRTarBase were integrated to obtain the final miRNA-gene binding score ($S_p^{\text{mir|gene}}$) based on the Pareto Frontier Analysis described below.

Pan-tissue co-expression analysis. Most existing databases predict circRNA-miRNA interactions solely relying on sequence-based algorithms, which completely ignore tissue-specific expression information of circRNAs^{41,44,65}. In order to form effective ceRNA networks, circRNA and miRNA both need to be expressed in the same tissue. Memczak et al. reported that CDR1as and miR-7 were both highly expressed in brain tissues, but not necessarily in other non-neuronal tissues¹⁸, which renders CDR1as a hallmark miR-7 sponge in neuronal tissues. Moreover, Guo et al. further argued that functional miRNA sponges require circRNAs to be expressed at consequential levels in the cell⁶⁶. Therefore, incorporating circRNA and miRNA expression into ceRNA network analysis can help filter out false positive interactions with low or no expression. To this end, we assigned circRNA-miRNA expression scores $S_{\text{exp}}^{\text{circ|mir}}$ to all interactions by the following methodology. First, for each tissue, circRNAs/miRNAs with no expression are excluded and the ones passed filtering were utilized to calculate the empirical cumulative distribution function (ECDF) on the mean normalized expression, giving rise to tissue-specific circRNA and miRNA ECDF scores, $S_{\text{exp}}^{\text{circ|tissue}}$ and $S_{\text{exp}}^{\text{mir|tissue}}$, that take values on (0, 1]. We then defined circRNA-miRNA score as $S_{\text{exp}}^{\text{circ|mir|t}} = \min_{t \in T} \left\{ S_{\text{exp}}^{\text{circ|t}}, S_{\text{exp}}^{\text{mir|t}} \right\}$, and the final circRNA-miRNA score across all tissues is given by $S_{\text{exp}}^{\text{circ|mir}} = \max_{t \in T} S_{\text{exp}}^{\text{circ|mir|t}}$, where T is the set of all tissues from ENCODE. Assigning scores in this manner ensures that a high $S_{\text{exp}}^{\text{circ|mir}}$ coincides with relatively high expression of the circRNA and miRNA in at least one tissue.

Integrative analysis of ceRNA interaction data using Pareto Frontier analysis. In order to improve the quality and functional relevance of the predicted ceRNA interactions, we integrated miRNA binding data and gene expression data using Pareto Frontier analysis (PFA). PFA is a technology to resolve the challenge of balancing among multiple competing objectives simultaneously to achieve an overall ranking optimality. In our case, it means to derive a combined interaction score between miRNAs and linear/circular RNAs based on various data types, such as expression data and predicted miRNA binding data. A key concept in PFA is the Pareto dominance. In the context of ranking 2-dimensional miRNA-circRNA interaction scores (expression score f_1 and binding score f_2), Pareto dominance is defined as: given two pairs of interactions x_1 and x_2 , x_2 is said to Pareto dominant x_1 if

$$f_n(x_1) \leq f_n(x_2), \quad \text{for all } n,$$

$$\text{and } f_n(x_1) < f_n(x_2), \quad \text{for at least one } n, \text{ where } n \in \{1,2\}.$$

The principle of Pareto dominance can be easily generalized for combing more than two scores. It is particularly suitable for combining asymmetric information without directly making comparisons across different data types, nor subjectively choosing a trade-off between different competing objectives. For circRNA-miRNA interactions, the length-normalized circRNA-miRNA binding score $S_{\text{MRE}}^{\text{circ|mir}}$ and circRNA-miRNA co-expression score $S_{\text{exp}}^{\text{circ|mir}}$ were combined using the PFA method to re-rank all pairs of circRNA-miRNA interactions. The new rank of each interaction pair was re-scaled by the total number of interaction pairs to obtain a final combined interaction score $S_p^{\text{circ|mir}}$ between 0 and 1, where 1 denotes the strongest interaction and 0 denotes no evidence for a given interaction. The new combined score $S_p^{\text{circ|mir}}$ is also referred to as the Pareto score of the circRNA-miRNA interaction. For circRNA-miRNA interactions that fall on the same Pareto front, their Pareto scores will be the same. Similarly, for a miRNA-gene interaction, previously described scores $S_{\text{MRE}}^{\text{mir|gene}}$ and $S_{\text{MTB}}^{\text{mir|gene}}$ were also combined using the Pareto Frontier method to calculate a new Pareto score $S_p^{\text{mir|gene}}$. More details regarding PFA are provided in the Supplementary Methods.

CircRNA functional enrichment analysis. Assigning functional annotations to an individual circRNA is based on the circRNA-miRNA-gene interaction framework we have built thus far, which consists of two steps: obtaining a list of significant genes and then testing functional enrichment of these genes. In the first step, given an individual circRNA c and a set of k miRNAs $M_k = [m_1, \dots, m_k] \subseteq M$, where $S_p^{c|m_i} > 0$ for all $i \leq k$ and M is the full set of miRNAs. We define the circRNA-miRNA Pareto score vector as $S_p^{c|M_k} = [S_p^{c|m_1}, S_p^{c|m_2}, \dots, S_p^{c|m_k}]$ and the miRNA-gene Pareto score matrix ($n \times k$).

$$S_p^{M_k|G} \begin{bmatrix} \left(S_p^{m_1|g_1} \dots S_p^{m_k|g_1} \right) \\ \vdots \\ \left(S_p^{m_1|g_n} \dots S_p^{m_k|g_n} \right) \end{bmatrix},$$

where G is the set of genes $[g_1, \dots, g_n]$. We define the circRNA-gene Pareto score vector $S_p^{c|G} = S_p^{c|M_k} \times \left(S_p^{M_k|G} \right)^T = [S_p^{c|g_1}, \dots, S_p^{c|g_n}]$ as the final statistic to measure the predicted association between a given circRNA and the set of all genes. Following a similar procedure as Bleazard et al. to adjust for observed

bias in miRNA functional enrichment analysis, we approximate the null distribution of $S_p^{c|G}$ ($S_{P,\text{null},n}^{c|G}$) by randomly drawing N (e.g., $N = 10,000$) $S_p^{M_k|G}$ matrices from $S_p^{M|G}$ and re-computing $S_p^{c|G}$ under each iteration⁶⁷, where $1 \leq n \leq N$. Then, the statistical significance of a gene g is given by p-value = $\frac{\sum_{n=1}^N \mathbb{1}(S_{P,\text{null},n}^{c|g} \geq S_p^{c|g}) + 1}{N+1}$ ⁶⁸, where $\mathbb{1}(x)$ is the indicator function that equals 1 when x is true, and 0 otherwise. Given a pre-specified statistical threshold (e.g., p-value ≤ 0.05), a list of genes surviving the threshold is identified. The second step, functional enrichment of significant genes, proceeds in standard fashion: apply Fisher's exact test to test for overrepresentation on sets of functional terms. All gene sets, including KEGG^{69–72} pathways and gene ontology (GO)^{72–74} terms were obtained from the R package pathfindR⁷².

Tools and software used. The Cerina tool is developed based on R⁷⁵ and R shiny⁷⁶, which also depends on several R packages (shinydashboard⁷⁷, shinyjs⁷⁸, shinycssloaders⁷⁹, shinyBS⁸⁰, DT⁸¹, tidyverse⁸², dendextend⁸³, visNetwork⁸⁴, heatmaply⁸⁵, Matrix⁸⁶, fastcluster⁸⁷, htmltools⁸⁸, reshape2⁸⁹, and igraph⁹⁰).

Additional R packages (ggplot2⁹¹, plotrix⁹², circlize⁹³) were used to produce the figures in this paper. Cytoscape⁹⁴ was used to create circRNA-miRNA-gene-function network for the prostate cancer case study. Commercial software Lucidchart (www.lucidchart.com) was used to assemble all final version of figures. All organ icons used in this paper were under Creative Commons licence (CC BY 3.0, <https://creativecommons.org/licenses/by/3.0/>), which were obtained from iconfinder (<https://www.iconfinder.com/>) through Lucidchart without any changes. Images “Anatomy, blood, coronary, heart, organ icon” (https://www.iconfinder.com/icons/4312967/anatomy_blood_coronary_heart_organ_icon), “Anatomy, bowel, digestion, intestine, small icon” (https://www.iconfinder.com/icons/4312981/anatomy_bowel_digestion_intestine_small_icon), “Abdomen, anatomy, cavity, diaphragm, organ icon” (https://www.iconfinder.com/icons/4312964/abdomen_anatomy_cavity_diaphragm_organ_icon), “Abdomen, digestion, gaster, organ, stomach icon” (https://www.iconfinder.com/icons/4312980/abdomen_digestion_gaster_organ_stomach_icon), and “Abdomen, anatomy, liver, metabolism, organ icon” (https://www.iconfinder.com/icons/4312973/abdomen_anatomy_liver_metabolism_organ_icon) by Eucalypt Studio; “Organs, uterus icon” (https://www.iconfinder.com/icons/1609656/organs_uterus_icon) by Design Sciences.

Results

Cerina overview: an integrative framework. Figure 1 gives the flow chart of Cerina, which consists of several streamlined modules: starting from linear/circRNA expression quantification, MRE prediction, ceRNA interaction integration, to circRNA functional annotation. Briefly, paired total RNA and miRNA sequencing data of 11 human organs from the ENCODE project were downloaded and processed to generate linear RNA (coding and non-coding), miRNA, and circular RNA expression profiles (Fig. 1a). To reduce false-positive circRNA predictions, we combined results from CIRI2 and CircExplorer, two methods that were previously validated to have the best overall performance on the simulated and real datasets. Based on the estimated mature sequences of circRNAs, TargetScan 7.2 was employed to predict putative MREs, which were further normalized by the length of each circRNA's mature splice sequence (Fig. 1b). Meanwhile, pan-tissue circRNA expression data were incorporated and tissue-specific ceRNA networks were also constructed (Fig. 1c). Following the Pareto dominance principle, all ceRNA interactions were ranked and grouped into a sequence of non-intersect sets called Pareto frontiers. These frontiers re-stratified all ceRNA interactions, integrating evidence from both gene expression and miRNA bindings. CeRNA interactions that fall onto the first Pareto frontier represent the circRNA-miRNA interactions with the highest confidence, either based on expression data or binding data, or both (Fig. 1d). Such procedure integratively re-prioritized a total of 1,540,275 ceRNA interactions between 33,455 circRNAs and 606 miRNAs detected in 11 ENCODE tissues. Finally, systems analysis was performed based on Pareto-ranked ceRNA interactions to identify top miRNAs, significant miRNA target genes, and enriched biological functions and pathways (Fig. 1e).

Pareto Frontier analysis improves accuracy and functional relevance of ceRNA interactions. We employed Pareto Frontier Analysis to integrate circRNA-miRNA binding data with their expression data, aiming to improve functional relevance of the predicted ceRNA interactions. Figure 2 gives 2531 circRNA-miRNA interaction pairs on the first 30 Pareto fronts with the top combination scores due to either strongest circRNA-miRNA binding potentials and/or highest co-expression from ENCODE. Nine circRNA-miRNA interactions, including eight unique circRNAs and five unique miRNAs, are located on the first Pareto front. Among those is the well-studied CDR1as-miR7 interaction, where 74 miR7 binding sites were predicted over the entire body of CDR1as. This circRNA-miRNA pair was also found to be co-expressed in several tissues, such as adrenal gland (circRNA SRPBM = 1680.4; miRNA CPM = 1032.4) and thyroid gland (circRNA SRPBM = 1843.5; miRNA CPM = 3503.0).

Besides the abovementioned interactions with both strong co-expression and binding scores, the Pareto method also highlights circRNA-miRNA pairs with unequal interaction strengths from two data sources. CircSPHKAP and miR-1-3p is one such example that ranks among the top Pareto fronts (front 19; $S_p^{\text{circ|mir}} = 0.9993254$) due to strong evidence from circRNA-miRNA co-expression ($S_{\text{exp}}^{\text{circ|mir}} = 0.9938584$) and relatively weaker binding potential (0.27 MREs/Kb). Further scrutinization revealed that circSPHKAP (chr2: 228,881,121–228,884,872) was exclusively expressed in heart tissues (circRNA SRPBM = 1055.9, rank: 62/9281 in hearts tissues) with no detectable back-splicing junction counts in the rest of ten tissues from ENCODE, which is consistent with findings from a recent study supporting the use of circSPHKAP as a biomarker for cardiomyocytes⁹⁵. On the other hand, miR-1-3p was among the highly expressed miRNA in heart tissues (CPM = 212,747), which was also known

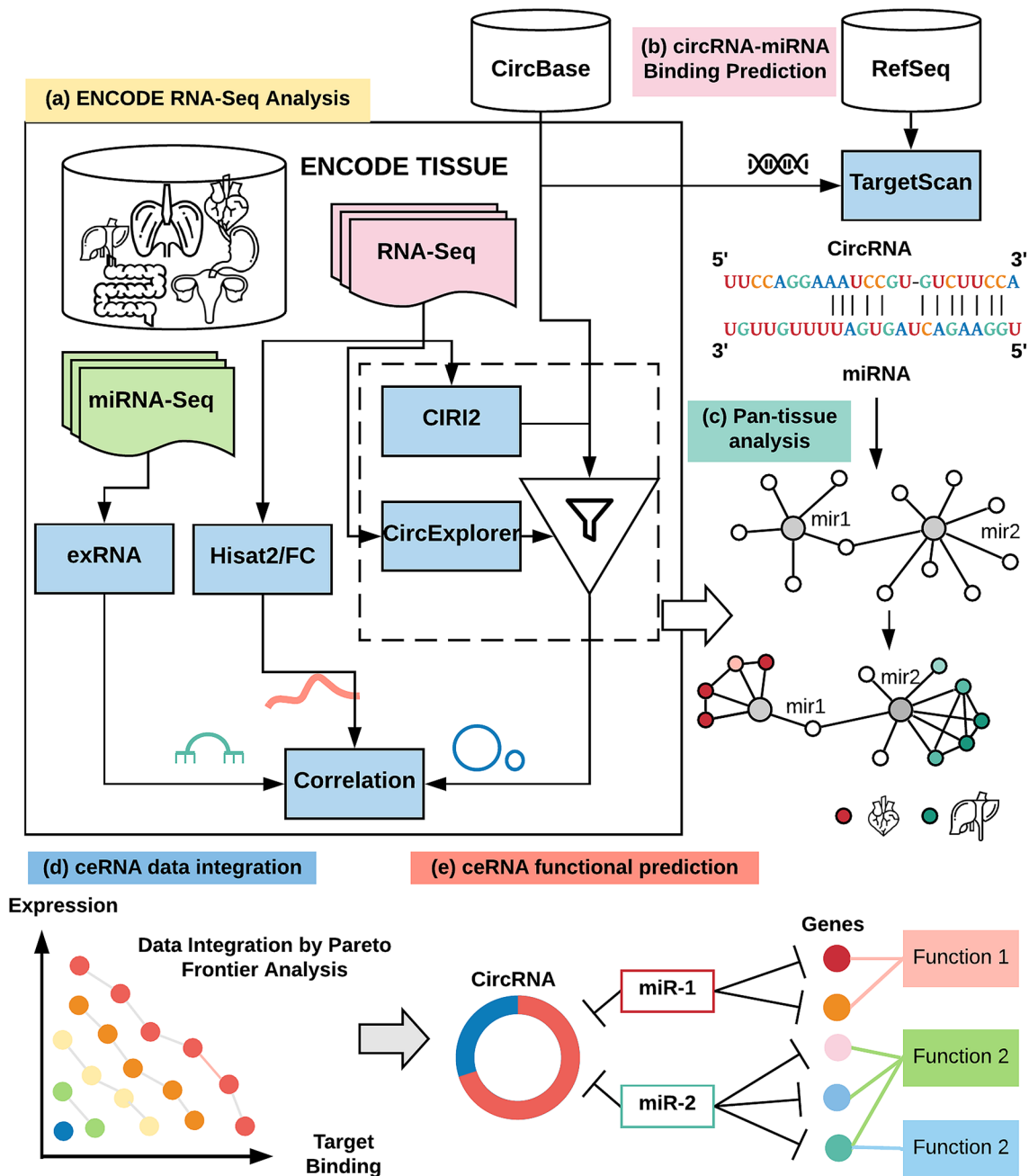


Figure 1. Flow chart of Cerina. (a) Analysis workflow for ENCODE tissue RNA-Seq/miRNA-Seq data. (b) Prediction of circRNA-miRNA bindings using TargetScan. (c) Pan-tissue analysis of ceRNA interactions. Incorporation of ENCODE gene expression data allows construct of tissue-specific circRNA-miRNA interaction networks. (d) Integrative analysis of ceRNA interactions. Pareto frontiers were calculated by integrating co-expression data with TargetScan-predict MRE data. (e) Based on the circRNA-miRNA-mRNA (gene)-function axis, circRNA functional prediction was performed by permuting the connections between a given circRNA and its interacting miRNAs/mRNAs. See section “Tools and software used” for license and attribution regarding the organ images used in the figure.

to be directly involved or implicated in various heart and cardiovascular diseases, including hypertrophic cardiomyopathy, coronary artery disease, myocardial infarction, heart failure and stroke⁹⁶. Interestingly, circSPHKAP was reported to have dynamic expression changes in human induced pluripotent stem cell derived cardiomyocytes during cardiac development⁹⁷, which further underscores its potential functional role in cardiac tissues.

To systematically evaluate the advantage of integrating circRNA-miRNA co-expression with MRE data over conventional MRE-based approaches, we performed functional relevance analysis of the top ranked circRNAs. We used circRNAs from CircFunBase⁴⁹ that were previously reported to be differentially expressed in one or more disease studies as one of the references. Figure 3a upper panel gives the percentages of overlap between circRNAs from the top- n ($1 \leq n \leq 3000$) circRNA-miRNA interactions and those from CircFunBase based on

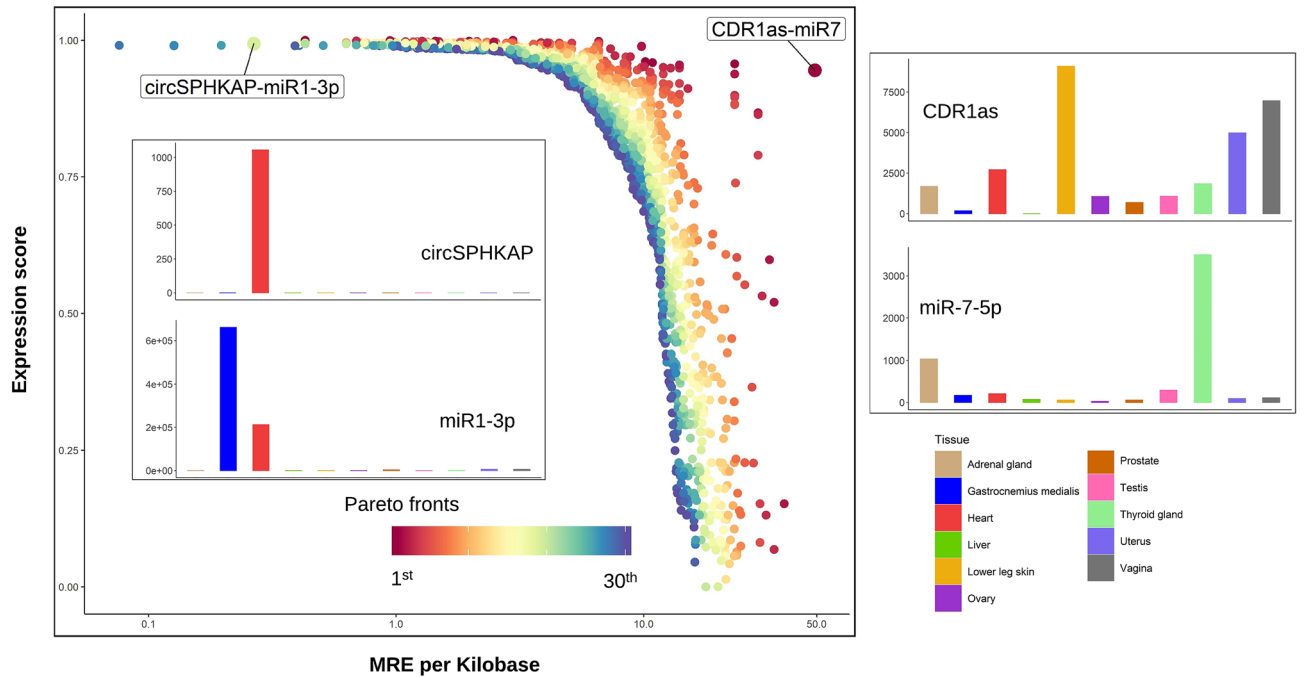


Figure 2. Pareto Frontier Analysis re-prioritize ceRNA interactions by integrating gene expression with MRE data. 2531 circRNA-miRNA interactions on the top 30 Pareto fronts. Tissue-specific expression for selected ceRNA interactions: CDR1as and miR-7, and CircSPHKAP and miR-1. R package ggplot2 v3.3.2 (<https://cran.r-project.org/web/packages/ggplot2/index.html>) was used to create the figure.

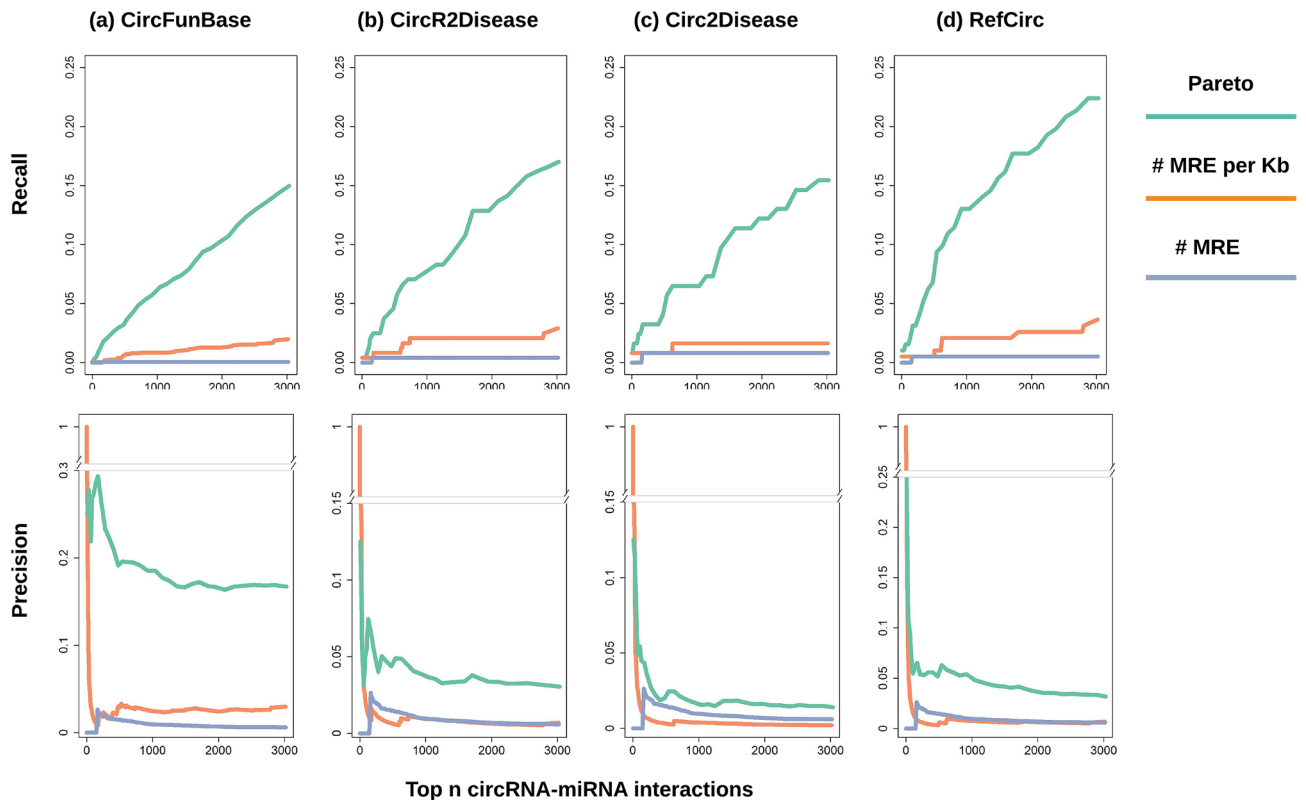


Figure 3. Evaluating the performance of different circRNA-miRNA interaction ranking methods using data from several circRNA function/disease databases. Upper panel: recall (sensitivity) curve. Lower panel: precision. R package plotrix v3.7-8 (<https://cran.r-project.org/web/packages/plotrix/index.html>) was used to create gap plot.

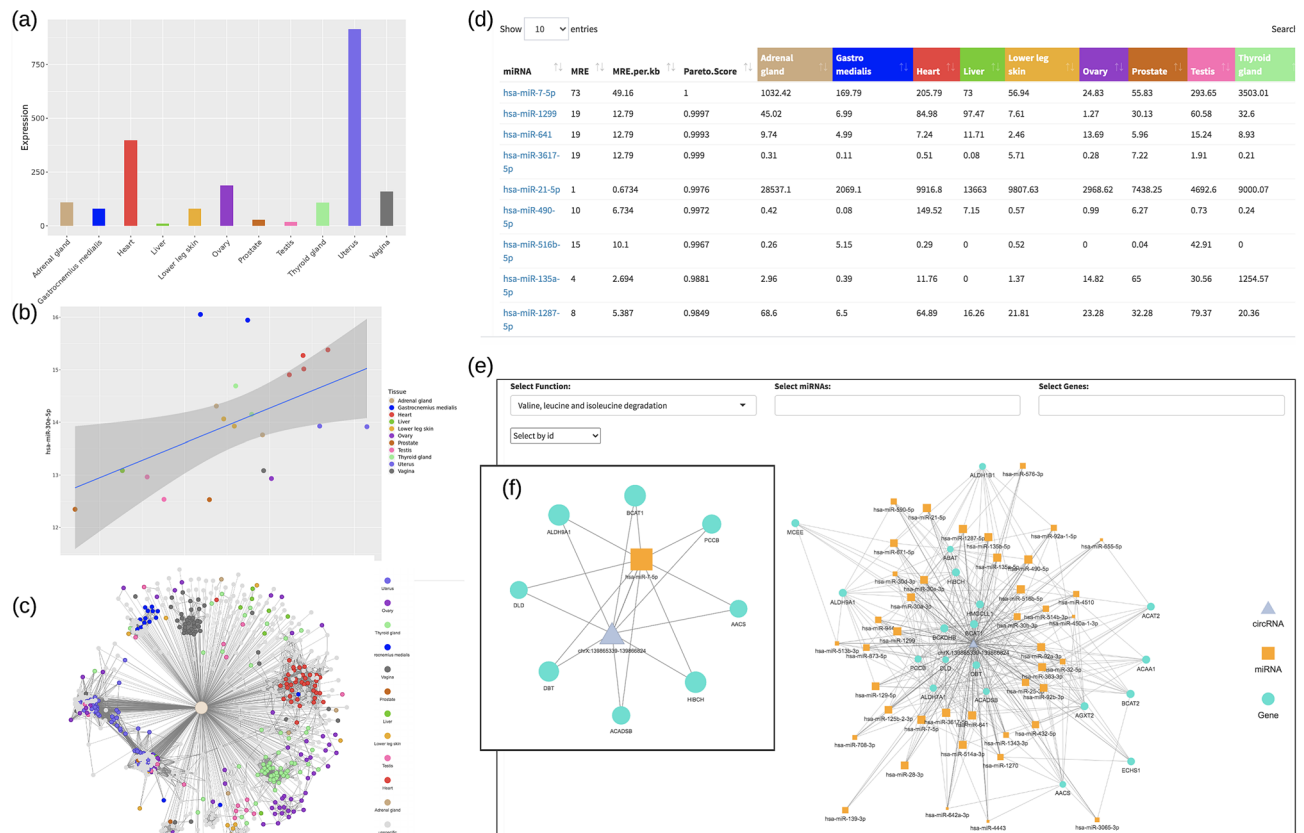


Figure 4. Web interface of Cerina. **(a,b)** In the data exploration section, users can query an mRNA, miRNA, or circRNA to examine its expression across different tissues. Users can also perform correlation analysis for ceRNA interactions. **(c)** MiRNA-circRNA interaction page allows users to visualize and report tissue-specific miRNA-circRNA interactions as a network. **(d)** CircRNA-miRNA exploration tab reports all miRNAs interacting with a selected circRNA as a downloadable table displaying MRE, Pareto score, and miRNA expression. **(e)** Functional analysis tab reports significantly enriched pathways and gene ontology terms that can be visualized as a network displaying the selected circRNA with interacting miRNAs and genes. **(f)** Users can choose a subset of miRNAs/genes/functions to display in the network. R packages ggplot2 v3.3.2 (<https://cran.r-project.org/web/packages/ggplot2/index.html>), visNetwork v2.0.9 (<https://cran.r-project.org/web/packages/visNetwork/index.html>), and DT v0.15 (<https://cran.r-project.org/web/packages/DT/index.html>), were used to create the figure.

three different ranking methods: the Pareto method, total number of MREs (nMRE), length-normalized number of MRE (i.e., number of MREs per kilo bases: nMRE/Kb). Apparently, Pareto integration of co-expression data with binding data significantly improved the recall of known circRNAs from CircFunBase: ~15% from the Pareto method compared to 2% from the length-normalized MRE method. It is worth noting that using the total number of MREs yielded very low recall compared to its length-normalized counterpart. Moreover, the precision of the Pareto method was also significantly increased (Fig. 3a lower panel). Additionally, we applied similar analysis on three more circRNA databases, including CircR2Disease⁹⁸, Circ2Disease⁹⁹, and RefCirc (<http://www.ncvcr.org/RefCirc/index.php>), which contained annotated disease-associated circRNAs from independent research groups. Figure 3b–d show that circRNAs prioritized by the Pareto method was consistently more enriched in known disease associations by several folds, in terms of both precision and recall, which provided strong evidence that incorporation of co-expression data significantly increased functional relevance of prioritized ceRNA interactions.

Moreover, we further validated our Pareto-ranked miRNA-gene interactions on three additional miRNA target databases: miRDB^{100,101}, miRTAR¹⁰², and miRWalk¹⁰³. We considered the top 3000 miRNA-gene interactions ranked by Pareto, TargetScan, and miRTarBase and found that Pareto improved the performance, in terms of precision and recall, on all three databases (Supplementary Fig. S3). This demonstrates Pareto's utility in combining multiple pieces of information, namely TargetScan and miRTarBase, to improve overall performance.

Cerina shiny server interface. We developed a user-friendly R Shiny web application of Cerina for researchers to visualize ceRNA interactions and perform circRNA functional enrichment analysis. The tool consists of three main components: data exploration, miRNA-circRNA network visualization, and functional enrichment analysis.

In the data exploration section (Fig. 4a,b), users can query individual circRNAs, miRNAs, and linear RNAs to view their expression profiles across the 11 ENCODE tissues. Additionally, correlation analysis of any queried pair can be visualized via scatterplots. The miRNA-circRNA network page allows users to query an individual miRNA to visualize a network of its interacting circRNAs (Fig. 4c). A downloadable table listing all circRNAs plotted in the network is also provided. The table shows detailed information such as parental gene, tissue specific expression, number of MREs, number of MREs per kilobase, and the Pareto score. Finally, in the functional enrichment component, users can enter an individual circRNA to visualize a network and download a table including all interacting miRNAs (Fig. 4d). Here, users have the option to run the permutation test using either all interacting miRNAs or a subset of “top” miRNAs to identify target genes. After running the permutation test, users can proceed to functional enrichment analysis of significant genes (e.g., $p\text{-value} \leq 0.05$) on KEGG^{69–72} pathways or gene ontology (GO)^{72–74} terms. The enrichment results of a circRNA are output as a downloadable table that can be categorized by either the functional term (i.e., KEGG or GO term) or the binding miRNA, with graphic visualization also made available (Fig. 4e). Cerina allows users to choose a subset of miRNA/genes/functions to display in the graph (Fig. 4f). A detailed tutorial is accessible at the Cerina website.

Case studies on differentially expressed circRNAs. *ceRNAs in cancers: integrative analysis with TCGA data.* We first performed Cerina analysis on gastric cancer¹⁰⁴ and prostate cancer⁶⁵ datasets to demonstrate its utility to identify potential roles of circRNAs in functional sequestration of miRNAs. Enrichment analysis validated that Cerina-predicted miRNAs and their target genes were strongly associated with gastric and prostate cancer annotated by Human MicroRNA Disease Database (HMDDv3.2)⁹⁶ and KEGG (Supplementary Table S2). To further explore the potential roles of circRNAs as competing endogenous RNAs in tumorigenesis, we further analyzed miRNA and mRNA expression data from The Cancer Genome Atlas^{105,106}, allowing construction of ceRNA networks with expression changes of circRNA/mRNA that were inversely correlated with those of the miRNAs (Supplementary Methods).

In the gastric cancer dataset, circARHGGEF12 (hsa_circ_0002089; chr11: 120,347,369–120,348,235), a circRNA that was down-regulated in cancer, was significantly enriched in both miRNA and pathway enrichment analysis (Supplementary Tables S2, S3). Two precursor miRNAs, hsa-mir-134 and hsa-mir-590, with predicted MREs on circARHGGEF12 were significantly up-regulated in TCGA gastric cancer dataset. Differential gene expression analysis of TCGA mRNA-Seq data identified 12 genes from the KEGG gastric cancer pathway that were down-regulated in cancer tissue, eight of which had predicted interactions in Cerina with the two up-regulated miRNAs. SMAD4, hub of TGF β signaling and a tumor suppressor for gastrointestinal carcinogenesis¹⁰⁷ was among the down-regulated target genes of circARHGGEF12, suggesting a potential tumorigenesis effect caused by un-sequester of oncogenic miR-134 and miR-590^{108,109}. Catenin Alpha (CTNNA1 and CTNNA2) expression was also down-regulated in cancerous tissues, which was consistent to well-reported tumor-suppressor functions of CTNNA1 and CTNNA2 in various cancers^{110–112}. Interestingly, CTNNA2 was previously predicted to be part of lincRNA-mediated miR-590-3p sponge network (<http://cis.hku.hk/GastricCancerMAP/index.php>), unveiling a novel role of circARHGGEF12 in gastric carcinogenesis and its involvement in complicated wiring of ceRNA interactions harboring miR-590.

In the prostate cancer study comparing localized primary prostate adenocarcinoma and matched normal tissues, five differentially expressed circRNAs: circHIPK3 (hsa_circ_0000284; chr11: 33,307,958–33,309,057), circN4BP2L2 (hsa_circ_0000471; chr13: 33,091,993–33,101,669), circUNC13B (hsa_circ_0008518; chr9: 35,295,692–35,313,986), circZCCHC6 (hsa_circ_0001869; chr9: 88,920,106–88,924,932), and circSENP6 (hsa_circ_0001614; chr6: 76,412,360–76,412,788), had significant enrichment in miRNAs related to “carcinoma, prostate” and KEGG “prostate cancer” pathway (Supplementary Tables S2, S4). CircHIPK3, in particular, was one of the most abundantly expressed circRNA with an average log expression (SRPBM) of 12.1, compared to the median average log expression of 2.4 among all detected circRNAs in prostate tissues. Dysregulation of circHIPK3 was frequently reported in multiple cancers¹¹³. Interestingly, both up- and down-regulation of circHIPK3 were identified in tumor tissues, indicating a dual role of circHIPK3 in cancer to regulate tumor progression through sponging different miRNAs¹¹⁴. As a side note, the host gene of circHIPK3 was also significantly down-regulated in the prostate cancer tissues, which was further confirmed by an independent TCGA RNA-Seq data. Integrative analysis of TCGA RNA-Seq/miRNA-Seq data revealed that two Cerina-predicted miRNAs mir-10b and mir-375 that can be sequestered by circHIPK3 were up-regulated in TCGA prostate cancer samples. Oncogenic functions of mir-10b has been well-documented in various cancers, including oral cancer¹¹⁵, head and neck cancer¹¹⁶, hepatocellular carcinoma¹¹⁷, breast cancer¹¹⁸, and colon cancer¹¹⁹. Top mir-10b targets (Pareto score $S_p^{\text{mir}|\text{gene}} \geq 0.95$) that were also down-regulated in tumors included HOXD10, KLF4, and PTEN, all of which had well-known tumor-suppressor functions and reduced expression in prostate cancer tissues^{120–122}. It is worth noting that depletion of mir-10b restored PTEN expression in breast cancer, which led to decreased cancer stem cell renewal through inhibition of AKT¹¹⁸. Mir-375, another highly expressed miRNA that can be sponged by circHIPK3, was well described as a tumor suppressor in many cancers, yet its expression was found to be up-regulated in breast and prostate cancers¹²³. Consistently, our analysis of TCGA prostate miRNA-Seq data showed significant up-regulation of mir-375 ($\log_2\text{FC} = 1.874$, adjusted $p\text{-value} = 4.85\text{E}-43$). Top down-regulated mir-375 target genes ($S_p^{\text{mir}|\text{gene}} \geq 0.95$) included tumor suppressors such as ZFP36L2, CDKN2B, PRKCA, KLF4, and EXT1, suggesting possible protumorigenic activity of circHIPK3 implemented through ceRNA interaction with mir-375. Figure 5a gives the circHIPK3-centered ceRNA interaction network, connecting mir-10 and mir-375 to significant dysregulated KEGG pathways, including PI3K-AKT signaling and p53 signaling pathways.

Reduced circTTN expression correlates with down-regulation of immune response in end-stage dilated cardiomyopathy. In the dilated cardiomyopathy (DCM) dataset¹²⁴, circTTN (hsa_circ_0141774; chr2: 179,542,851–

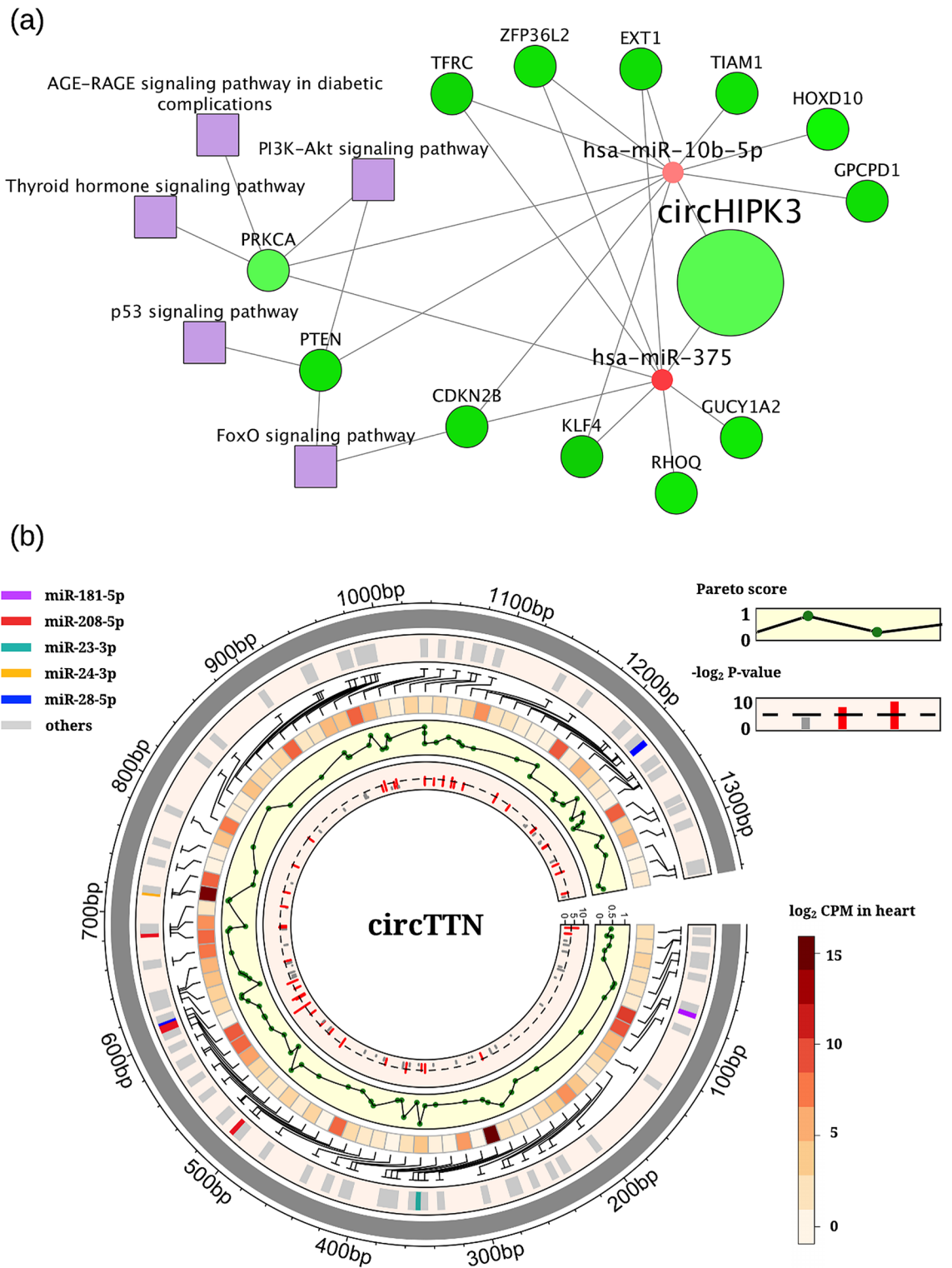


Figure 5. Cerina links circRNAs with potential functions in cancer and cardiovascular disease. **(a)** Network view of predicted circRNA-miRNA-gene-function associations for circHIPK3 in prostate cancer. For circRNA, miRNA, and genes, their colors represent their fold changes (red: up-regulated in prostate cancer; green: down-regulated in prostate cancer). **(b)** Circos plot showing predicted miRNAs of circTTN with Pareto score, expression in heart tissues (\log_2 CPM), and enrichment in immunology functions ($-\log_2 p$ -value). Cytoscape v3.5.1 was used to create the circRNA-miRNA-gene-function network figure (<https://cytoscape.org/>). R package circlize v0.4.8 (<https://cran.r-project.org/web/packages/circlize/index.html>) was used to make circos plot.

179,585,929) had the highest abundance among all differentially expressed circRNAs. circTTN was also exclusively expressed in heart tissues (Supplementary Fig. S4). Cerina analysis revealed that circTTN had binding sites for 82 miRNAs, among them seven mature miRNAs (miR-23b-3p, miR-23a-3p, miR-24-3p, miR-181a-5p, miR-28-5p, miR-181b-5p, miR-208a-5p) with Pareto scores greater than 0.95 (Fig. 5b). Functional enrichment analysis showed that targets of these seven miRNAs were highly associated with cardiovascular and circulatory system development, which were also significantly enriched in several pathways, including TNF, FoxO, and ERBB signaling pathways. In an independent gene expression study (GSE3586)¹²⁵, Barth et al. reported that end-stage DCM was characterized by excessive down-regulation of “immune response”, “inflammatory response”, and “chemokine activity”. Given the apparent under-expression of circTTN in DCM, we hence seek to investigate the possibility of circTTN-mediated suppression of immunity and inflammatory response through ceRNA interactions. Consistently, over-expression of mir-23a, mir-24, mir-28, and mir-208 in DCM patients were previously reported in one or more studies^{126–129}, which suggested a direct link between increased miRNA activity and reduced circTTN expression levels in DCM. Moreover, enrichment analysis showed that target genes of mir-23a/b were significantly associated with the down-regulation of immune-related genes in Barth’s study (Supplementary Methods), exemplified by binding of mir-23a/b to CCL2, a chemokine with the most significant down-regulation in DCM. Taken together, deregulation of circTTN correlates with down-regulation of immune response in end-stage DCM patient likely through modulation of mir-23a/b and others.

Conclusion and discussion

Cerina is the first systematic circRNA functional annotation tool based on integrative analysis of competing endogenous RNA interactions. It has a collection of more than 1.5 million inferred ceRNA interactions between over 33,000 circRNAs and hundreds of miRNAs detected in 11 ENCODE tissues. Although many databases were established for searching MREs in circRNAs, none of them incorporated gene expression data for predicting circRNA-miRNA interactions. Guo et al., raised concerns regarding the functional role of thousands of low-expressing circRNAs, strongly suggesting that the expression levels of circRNAs to be taken into account when interpreting circRNA functions⁶⁶. On a similar note, TargetScan also argued that, in order to mediate consequential repression of its targets, the expression of a miRNA should reach an adequate level, hence recommending removing false-positive interactions based on miRNA expressions levels (http://www.targetscan.org/vert_72/docs/FP_noncons.html). By integrating paired circRNA and miRNA expression data with the predicted MREs using a Pareto Frontier Analysis framework, we have significantly improved the accuracy and functional relevance of the identified ceRNA interactions, validated by data from several mainstream circRNA-disease databases. Through Cerina’s Shiny web interface, users can perform functional query of a circRNA to retrieve information regarding its most likely sponged miRNAs and their tissue-specific expressions, down-stream target genes, and potential enriched biological functions and pathways.

Although applicable to various disease studies, one major limitation of Cerina is that its entire functional prediction is built upon the circRNA-miRNA-mRNA axis. While this miRNA-sponge paradigm is under the spotlight for human circRNA research^{18–39}, increasing evidences have supported a number of alternative biological mechanisms, including alternative splicing regulation, RNA-binding protein sponge, posttranscriptional gene regulation, and protein-coding, among others^{3,7–13}.

In the context of human cancer, circCcnb1 may directly bind H2ax and wild-type p53, which attenuates tumor-suppressor function of p53 and promotes cell proliferation by allowing Bcl2-Bclaf1 binding. On the contrary, in p53 mutant cells, the circCcnb1-H2ax compound binds Bclaf1, hence activates Bclaf1 tumor-suppressor function and leads to apoptosis¹³⁰. In another breast cancer study, co-localization of circ-Amot1 and c-myc was detected, suggesting abnormal levels of circAmot1 to facilitates c-myc nuclear translocation through direct circRNA-protein binding¹³¹. In human glioma, a new protein encoded by circFBXW7 was discovered to have inhibitory effect of cell cycle and proliferation¹³².

Similarly, in non-cancer diseases, various “non-sponging” mechanistic models have been proposed for circRNAs in disease pathogenesis and progression. Examples include circ-Foxo3, a circRNA that promotes cardiac senescence binds CDK2 and p21 to form a ternary complex, blocking cell cycle progression¹³³. Also, in a systemic lupus erythematosus (SLE), degraded circRNAs upon viral induction in monocytes was found to form short RNA duplexes that inhibited abnormal protein kinase R activation cascade, highlighting a new role of circRNA in autoimmune diseases due to its unique structure¹³⁴.

On the other hand, due to the complex ceRNA networking in mammalian cells, circRNAs are a potent family, yet not the only one, of being capable of regulating protein-coding genes by sequestration of miRNAs. Other than circRNAs, small non-coding RNAs, pseudogenes, and lincRNAs, all actively participate in ceRNA interaction network through competing of shared miRNAs^{14,135–137}. Therefore, when Cerina predicts strong ceRNA associations that lead to functional outcomes, experimental validations are needed to further confirm the identified interactions, such as use pull-down assay and dual-luciferase reporter assay to confirm circRNA-miRNA binding^{31,34,138}, or over-express/silence circRNA or their interacting miRNAs to further verify predicted ceRNA interactions and associated phenotypic changes^{31,34,139,140}.

Data availability

A web service of Cerina can be accessed through: <https://www.bswhealth.med/research/Pages/biostat-softwre.aspx>.

Code availability

Source code for Cerina is available through GitHub at <https://github.com/jcardenas14/CERINA>.

Received: 23 June 2020; Accepted: 6 October 2020

Published online: 17 December 2020

References

- Salzman, J. Circular RNA expression: its potential regulation and function. *Trends Genet.* **32**, 309–316. <https://doi.org/10.1016/j.tig.2016.03.002> (2016).
- Chen, L. L. & Yang, L. Regulation of circRNA biogenesis. *RNA Biol.* **12**, 381–388. <https://doi.org/10.1080/15476286.2015.1020271> (2015).
- Zhang, Z., Yang, T. & Xiao, J. Circular RNAs: promising biomarkers for human diseases. *EBioMedicine* **34**, 267–274. <https://doi.org/10.1016/j.ebiom.2018.07.036> (2018).
- Lei, B., Tian, Z., Fan, W. & Ni, B. Circular RNA: a novel biomarker and therapeutic target for human cancers. *Int. J. Med. Sci.* **16**, 292–301. <https://doi.org/10.7150/ijms.28047> (2019).
- Li, S. & Han, L. Circular RNAs as promising biomarkers in cancer: detection, function, and beyond. *Genome Med.* **11**, 15. <https://doi.org/10.1186/s13073-019-0629-7> (2019).
- Bolha, L., Ravnik-Glavac, M. & Glavac, D. Circular RNAs: biogenesis, function, and a role as possible cancer biomarkers. *Int. J. Genomics* **2017**, 6218353. <https://doi.org/10.1155/2017/6218353> (2017).
- Meng, S. *et al.* CircRNA: functions and properties of a novel potential biomarker for cancer. *Mol. Cancer* **16**, 94. <https://doi.org/10.1186/s12943-017-0663-2> (2017).
- Lasda, E. & Parker, R. Circular RNAs: diversity of form and function. *RNA* **20**, 1829–1842. <https://doi.org/10.1261/rna.047126.114> (2014).
- Barrett, S. P. & Salzman, J. Circular RNAs: analysis, expression and potential functions. *Development* **143**, 1838–1847. <https://doi.org/10.1242/dev.128074> (2016).
- Huang, S. *et al.* The emerging role of circular RNAs in transcriptome regulation. *Genomics* **109**, 401–407. <https://doi.org/10.1016/j.ygeno.2017.06.005> (2017).
- Rong, D. *et al.* An emerging function of circRNA-miRNAs-mRNA axis in human diseases. *Oncotarget* **8**, 73271–73281. <https://doi.org/10.18632/oncotarget.19154> (2017).
- Shang, Q., Yang, Z., Jia, R. & Ge, S. The novel roles of circRNAs in human cancer. *Mol. Cancer* **18**, 6. <https://doi.org/10.1186/s12943-018-0934-6> (2019).
- Verduci, L., Strano, S., Yarden, Y. & Blandino, G. The circRNA-microRNA code: emerging implications for cancer diagnosis and treatment. *Mol. Oncol.* **13**, 669–680. <https://doi.org/10.1002/1878-0261.12468> (2019).
- Tay, Y., Rinn, J. & Pandolfi, P. P. The multilayered complexity of ceRNA crosstalk and competition. *Nature* **505**, 344–352. <https://doi.org/10.1038/nature12986> (2014).
- Li, J. H., Liu, S., Zhou, H., Qu, L. H. & Yang, J. H. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* **42**, D92–D97. <https://doi.org/10.1093/nar/gkt1248> (2014).
- Zhong, Y. *et al.* Circular RNAs function as ceRNAs to regulate and control human cancer progression. *Mol. Cancer* **17**, 79. <https://doi.org/10.1186/s12943-018-0827-8> (2018).
- Gomes, C. P. C. *et al.* Circular RNAs in the cardiovascular system. *Noncoding RNA Res.* **3**, 1–11. <https://doi.org/10.1016/j.ncrna.2018.02.002> (2018).
- Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338. <https://doi.org/10.1038/nature11928> (2013).
- Hansen, T. B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384–388. <https://doi.org/10.1038/nature11993> (2013).
- Weng, W. *et al.* Circular RNA ciRS-7-A promising prognostic biomarker and a potential therapeutic target in colorectal cancer. *Clin. Cancer Res.* **23**, 3918–3928. <https://doi.org/10.1158/1078-0432.CCR-16-2541> (2017).
- Pan, H. *et al.* Overexpression of circular RNA ciRS-7 abrogates the tumor suppressive effect of miR-7 on gastric cancer via PTEN/PI3K/AKT signaling pathway. *J. Cell Biochem.* **119**, 440–446. <https://doi.org/10.1002/jcb.26201> (2018).
- Li, R. C. *et al.* CiRS-7 promotes growth and metastasis of esophageal squamous cell carcinoma via regulation of miR-7/HOXB13. *Cell Death Dis.* **9**, 838. <https://doi.org/10.1038/s41419-018-0852-y> (2018).
- Geng, H. H. *et al.* The circular RNA Cdr1as promotes myocardial infarction by mediating the regulation of miR-7a on its target genes expression. *PLoS ONE* **11**, e0151753. <https://doi.org/10.1371/journal.pone.0151753> (2016).
- Chen, J. *et al.* Circular RNA profile identifies circPVT1 as a proliferative factor and prognostic marker in gastric cancer. *Cancer Lett.* **388**, 208–219. <https://doi.org/10.1016/j.canlet.2016.12.006> (2017).
- Wan, L. *et al.* Circular RNA-ITCH suppresses lung cancer proliferation via inhibiting the Wnt/beta-catenin pathway. *Biomed. Res. Int.* **2016**, 1579490. <https://doi.org/10.1155/2016/1579490> (2016).
- Chen, G., Shi, Y., Liu, M. & Sun, J. circHIPK3 regulates cell proliferation and migration by sponging miR-124 and regulating AQP3 expression in hepatocellular carcinoma. *Cell Death Dis.* **9**, 175. <https://doi.org/10.1038/s41419-017-0204-3> (2018).
- Zheng, J. *et al.* TTBK2 circular RNA promotes glioma malignancy by regulating miR-217/HNF1beta/Derlin-1 pathway. *J. Hematol. Oncol.* **10**, 52. <https://doi.org/10.1186/s13045-017-0422-2> (2017).
- Altesha, M. A., Ni, T., Khan, A., Liu, K. & Zheng, X. Circular RNA in cardiovascular disease. *J. Cell Physiol.* **234**, 5588–5600. <https://doi.org/10.1002/jcp.27384> (2019).
- Fan, X. *et al.* Circular RNAs in cardiovascular disease: an overview. *Biomed. Res. Int.* **2017**, 5135781. <https://doi.org/10.1155/2017/5135781> (2017).
- Deng, Y.-Y. *et al.* GW27-e1167 circular RNA related to PPAR γ function as ceRNA of microRNA in human acute myocardial infarction. *J. Am. Coll. Cardiol.* **68**, C51–C52. <https://doi.org/10.1016/j.jacc.2016.07.189> (2016).
- Wang, K. *et al.* Circular RNA mediates cardiomyocyte death via miRNA-dependent upregulation of MTP18 expression. *Cell Death Differ.* **24**, 1111–1120. <https://doi.org/10.1038/cdd.2017.61> (2017).
- Tang, C. M. *et al.* CircRNA_000203 enhances the expression of fibrosis-associated genes by derepressing targets of miR-26b-5p, Col1a2 and CTGF, in cardiac fibroblasts. *Sci. Rep.* **7**, 40342. <https://doi.org/10.1038/srep40342> (2017).
- Zhou, B. & Yu, J. W. A novel identified circular RNA, circRNA_010567, promotes myocardial fibrosis via suppressing miR-141 by targeting TGF-beta1. *Biochem. Biophys. Res. Commun.* **487**, 769–775. <https://doi.org/10.1016/j.bbrc.2017.04.044> (2017).
- Wang, K. *et al.* A circular RNA protects the heart from pathological hypertrophy and heart failure by targeting miR-223. *Eur. Heart J.* **37**, 2602–2611. <https://doi.org/10.1093/eurheartj/ehv713> (2016).
- Chen, J., Cui, L., Yuan, J., Zhang, Y. & Sang, H. Circular RNA WDR77 target FGF-2 to regulate vascular smooth muscle cells proliferation and migration by sponging miR-124. *Biochem. Biophys. Res. Commun.* **494**, 126–132. <https://doi.org/10.1016/j.bbrc.2017.10.068> (2017).
- Lukiw, W. J. Circular RNA (circRNA) in Alzheimer's disease (AD). *Front. Genet.* **4**, 307. <https://doi.org/10.3389/fgene.2013.00307> (2013).
- Piwecka, M. *et al.* Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function. *Science* <https://doi.org/10.1126/science.aam8526> (2017).

38. Wang, Z. *et al.* Identifying circRNA-associated-ceRNA networks in the hippocampus of Abeta1-42-induced Alzheimer's disease-like rats using microarray analysis. *Aging (Albany, N.Y.)* **10**, 775–788. <https://doi.org/10.18632/aging.101427> (2018).
39. Ma, N. *et al.* Whole-transcriptome analysis of APP/PS1 mouse brain and identification of circRNA-miRNA-mRNA networks to investigate AD pathogenesis. *Mol. Ther. Nucleic Acids* **18**, 1049–1062. <https://doi.org/10.1016/j.omtn.2019.10.030> (2019).
40. Dudekula, D. B. *et al.* CircInteractome: a web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biol.* **13**, 34–42. <https://doi.org/10.1080/15476286.2015.1128065> (2016).
41. Ji, P. *et al.* Expanded expression landscape and prioritization of circular RNAs in mammals. *Cell Rep.* **26**, 3444–3460. <https://doi.org/10.1016/j.celrep.2019.02.078> (2019).
42. Ghosal, S., Das, S., Sen, R., Basak, P. & Chakrabarti, J. Circ2Traits: a comprehensive database for circular RNA potentially associated with disease and traits. *Front. Genet.* **4**, 283. <https://doi.org/10.3389/fgene.2013.00283> (2013).
43. Liu, Y. C. *et al.* CircNet: a database of circular RNAs derived from transcriptome sequencing data. *Nucleic Acids Res.* **44**, D209–215. <https://doi.org/10.1093/nar/gkv940> (2016).
44. Xia, S. *et al.* Comprehensive characterization of tissue-specific circular RNAs in the human and mouse genomes. *Brief Bioinform.* **18**, 984–992. <https://doi.org/10.1093/bib/bbw081> (2017).
45. Xia, S. *et al.* CSCD: a database for cancer-specific circular RNAs. *Nucleic Acids Res.* **46**, D925–D929. <https://doi.org/10.1093/nar/gkx863> (2018).
46. Liu, M., Wang, Q., Shen, J., Yang, B. B. & Ding, X. Circbank: a comprehensive database for circRNA with standard nomenclature. *RNA Biol.* **16**, 899–905. <https://doi.org/10.1080/15476286.2019.1600395> (2019).
47. Zang, J., Lu, D. & Xu, A. The interaction of circRNAs and RNA binding proteins: an important part of circRNA maintenance and function. *J. Neurosci. Res.* **98**, 87–97. <https://doi.org/10.1002/jnr.24356> (2020).
48. Chen, X. *et al.* circRNADb: a comprehensive database for human circular RNAs with protein-coding annotations. *Sci. Rep.* **6**, 34985. <https://doi.org/10.1038/srep34985> (2016).
49. Meng, X., Hu, D., Zhang, P., Chen, Q. & Chen, M. CircFunBase: a database for functional circular RNAs. *Database (Oxford)* <https://doi.org/10.1093/database/baz003> (2019).
50. Davis, C. A. *et al.* The encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* **46**, D794–D801. <https://doi.org/10.1093/nar/gkx1081> (2018).
51. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 3. <https://doi.org/10.14806/ej.17.1.200> (2011).
52. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360. <https://doi.org/10.1038/nmeth.3317> (2015).
53. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> (2009).
54. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930. <https://doi.org/10.1093/bioinformatics/btt656> (2014).
55. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773. <https://doi.org/10.1093/nar/gky955> (2019).
56. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. <https://doi.org/10.1186/s13059-014-0550-8> (2014).
57. Hansen, T. B. Improved circRNA identification by combining prediction algorithms. *Front. Cell Dev. Biol.* **6**, 20. <https://doi.org/10.3389/fcell.2018.00020> (2018).
58. Zeng, X., Lin, W., Guo, M. & Zou, Q. A comprehensive overview and evaluation of circular RNA detection tools. *PLoS Comput. Biol.* **13**, e1005420. <https://doi.org/10.1371/journal.pcbi.1005420> (2017).
59. Gao, Y., Wang, J. & Zhao, F. CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biol.* **16**, 4. <https://doi.org/10.1186/s13059-014-0571-3> (2015).
60. Gao, Y., Zhang, J. & Zhao, F. Circular RNA identification based on multiple seed matching. *Brief Bioinform.* **19**, 803–810. <https://doi.org/10.1093/bib/bbx014> (2018).
61. Zhang, X. O. *et al.* Complementary sequence-mediated exon circularization. *Cell* **159**, 134–147. <https://doi.org/10.1016/j.cell.2014.09.001> (2014).
62. Rozowsky, J. *et al.* exceRpt: a comprehensive analytic platform for extracellular RNA profiling. *Cell Syst.* **8**, 352–357. <https://doi.org/10.1016/j.cels.2019.03.004> (2019).
63. Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* <https://doi.org/10.7554/eLife.05005> (2015).
64. Chou, C. H. *et al.* miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* **46**, D296–D302. <https://doi.org/10.1093/nar/gkx1067> (2018).
65. Vo, J. N. *et al.* The landscape of circular RNA in cancer. *Cell* **176**, 869–881. <https://doi.org/10.1016/j.cell.2018.12.021> (2019).
66. Guo, J. U., Agarwal, V., Guo, H. & Bartel, D. P. Expanded identification and characterization of mammalian circular RNAs. *Genome Biol.* **15**, 409. <https://doi.org/10.1186/s13059-014-0409-z> (2014).
67. Bleazard, T., Lamb, J. A. & Griffiths-Jones, S. Bias in microRNA functional enrichment analysis. *Bioinformatics* **31**, 1592–1598. <https://doi.org/10.1093/bioinformatics/btv023> (2015).
68. Phipson, B. & Smyth, G. K. Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. *Stat. Appl. Genet. Mol. Biol.* **9**, 39. <https://doi.org/10.2202/1544-6115.1585> (2010).
69. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
70. Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. & Tanabe, M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* **47**, D590–D595. <https://doi.org/10.1093/nar/gky962> (2019).
71. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951. <https://doi.org/10.1002/pro.3715> (2019).
72. Ulgen, E., Ozisik, O. & Sezerman, O. U. pathfindR: an R Package for comprehensive identification of enriched pathways in omics data through active subnetworks. *Front. Genet.* **10**, 858. <https://doi.org/10.3389/fgene.2019.00858> (2019).
73. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29. <https://doi.org/10.1038/75556> (2000).
74. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still going strong. *Nucleic Acids Res.* **47**, D330–D338. <https://doi.org/10.1093/nar/gky1055> (2019).
75. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2020).
76. Chang, W., Cheng, J., Allaire, J. & McPherson, J. *shiny: Web Application Framework for R* (2020).
77. Chang, W. *shinydashboard: Create Dashboards with 'Shiny'* (2018).
78. Attali, D. *shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds* (2020).
79. Attali, D. *shinycssloaders: Add Loading Animations to a 'Shiny' Output While It's Recalculating* (2020).
80. Bailey, E. *shinyBS: Twitter Bootstrap Components for Shiny* (2015).
81. Xie, Y., Cheng, J. & Tan, X. *DT: A Wrapper of the JavaScript Library 'DataTables'* (2020).

82. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686. <https://doi.org/10.21105/joss.01686> (2019).
83. Galili, T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31**, 3718–3720. <https://doi.org/10.1093/bioinformatics/btv428> (2015).
84. Almende, B. V., Thieurmel, B. and Robert, T. *visNetwork: Network Visualization using 'vis.js' Library* (2019).
85. Galili, T., O'Callaghan, A., Sidi, J. & Sievert, C. heatmaply: an R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics* **34**, 1600–1602. <https://doi.org/10.1093/bioinformatics/btx657> (2018).
86. Bates, D. & Maechler, M. *Matrix: Sparse and Dense Matrix Classes and Methods* (2019).
87. Mullner, D. fastcluster: fast hierarchical, agglomerative clustering routines for R and python. *J. Stat. Softw.* **53**, 1–18 (2013).
88. Cheng, J., Sievert, C., Chang, W., Xie, Y. & Allen, J. *htmltools: Tools for HTML* (2020).
89. Wickham, H. Reshaping data with the reshape package. *J. Stat. Softw.* **21**, 1–20 (2007).
90. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *Interf. Complex Syst.* **1695**, 1–9 (2006).
91. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, Cham, 2016).
92. Lemon, J. Plotrix: a package in the red light district of R. *R-News* **6**, 8–12 (2006).
93. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812. <https://doi.org/10.1093/bioinformatics/btu393> (2014).
94. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504. <https://doi.org/10.1101/gr.1239303> (2003).
95. Lei, W. *et al.* Signature of circular RNAs in human induced pluripotent stem cells and derived cardiomyocytes. *Stem Cell Res. Ther.* **9**, 56. <https://doi.org/10.1186/s13287-018-0793-5> (2018).
96. Huang, Z. *et al.* HMDD v3.0: a database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res.* **47**, D1013–D1017. <https://doi.org/10.1093/nar/gky1010> (2019).
97. Siede, D. *et al.* Identification of circular RNAs with host gene-independent expression in human model systems for cardiac differentiation and disease. *J. Mol. Cell Cardiol.* **109**, 48–56. <https://doi.org/10.1016/j.yjmcc.2017.06.015> (2017).
98. Fan, C., Lei, X., Fang, Z., Jiang, Q. & Wu, F. X. CircR2Disease: a manually curated database for experimentally supported circular RNAs associated with various diseases. *Database (Oxford)*. <https://doi.org/10.1093/database/bay044> (2018).
99. Yao, D. *et al.* Circ2Disease: a manually curated database of experimentally validated circRNAs in human disease. *Sci. Rep.* **8**, 11018. <https://doi.org/10.1038/s41598-018-29360-3> (2018).
100. Liu, W. & Wang, X. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. *Genome Biol.* **20**, 18. <https://doi.org/10.1186/s13059-019-1629-z> (2019).
101. Chen, Y. & Wang, X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.* **48**, D127–D131. <https://doi.org/10.1093/nar/gkz757> (2020).
102. Hsu, J. B. *et al.* miRTar: an integrated system for identifying miRNA-target interactions in human. *BMC Bioinform.* **12**, 300. <https://doi.org/10.1186/1471-2105-12-300> (2011).
103. Sticht, C., De La Torre, C., Parveen, A. & Gretz, N. miRWalk: an online resource for prediction of microRNA binding sites. *PLoS ONE* **13**, e0206239. <https://doi.org/10.1371/journal.pone.0206239> (2018).
104. Shao, Y. *et al.* Global circular RNA expression profile of human gastric cancer and its clinical significance. *Cancer Med.* **6**, 1173–1180. <https://doi.org/10.1002/cam4.1055> (2017).
105. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209. <https://doi.org/10.1038/nature13480> (2014).
106. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell* **163**, 1011–1025. <https://doi.org/10.1016/j.cell.2015.10.025> (2015).
107. Wang, L. H. *et al.* Inactivation of SMAD4 tumor suppressor gene during gastric carcinoma progression. *Clin. Cancer Res.* **13**, 102–110. <https://doi.org/10.1158/1078-0432.CCR-06-1467> (2007).
108. Pan, J. Y. *et al.* miR-134: a human cancer suppressor?. *Mol. Ther. Nucleic Acids* **6**, 140–149. <https://doi.org/10.1016/j.omtn.2016.11.003> (2017).
109. Dong, Y. & Qiu, G.-B. Biological functions of miR-590 and its role in carcinogenesis. *Front. Lab. Med.* **1**, 173–176. <https://doi.org/10.1016/j.flm.2017.11.002> (2017).
110. Lee, B. *et al.* Homozygous deletions at 3p22, 5p14, 6q15, and 9p21 result in aberrant expression of tumor suppressor genes in gastric cancer. *Genes Chromosomes Cancer* **54**, 142–155. <https://doi.org/10.1002/gcc.22226> (2015).
111. Chen, X. X. *et al.* Methylation of CTNNA1 promoter: frequent but not an adverse prognostic factor in acute myeloid leukemia. *Leuk Res.* **38**, 613–618. <https://doi.org/10.1016/j.leukres.2014.03.002> (2014).
112. Fanjul-Fernandez, M. *et al.* Cell-cell adhesion genes CTNNA2 and CTNNA3 are tumour suppressors frequently mutated in laryngeal carcinomas. *Nat. Commun.* **4**, 2531. <https://doi.org/10.1038/ncomms3531> (2013).
113. Rophina, M., Sharma, D., Poojary, M. & Scaria, V. Circad: a comprehensive manually curated resource of circular RNA associated with diseases. *Database (Oxford)*. <https://doi.org/10.1093/database/baaa019> (2020).
114. Xie, Y. *et al.* The circular RNA HIPK3 (circHIPK3) and its regulation in cancer progression: review. *Life Sci.* <https://doi.org/10.1016/j.lfs.2019.117252> (2020).
115. Lu, Y. C. *et al.* Oncogenic function and early detection potential of miRNA-10b in oral cancer as identified by microRNA profiling. *Cancer Prev. Res. (Phila)* **5**, 665–674. <https://doi.org/10.1158/1940-6207.CAPR-11-0358> (2012).
116. Tu, H. F., Lin, S. C. & Chang, K. W. MicroRNA aberrances in head and neck cancer: pathogenetic and clinical significance. *Curr. Opin. Otolaryngol. Head Neck Surg.* **21**, 104–111. <https://doi.org/10.1097/MOO.0b013e32835e1d6e> (2013).
117. Zhu, Q. *et al.* miR-10b exerts oncogenic activity in human hepatocellular carcinoma cells by targeting expression of CUB and sushi multiple domains 1 (CSMD1). *BMC Cancer* **16**, 806. <https://doi.org/10.1186/s12885-016-2801-4> (2016).
118. Bahena-Ocampo, I. *et al.* miR-10b expression in breast cancer stem cells supports self-renewal through negative PTEN regulation and sustained AKT activation. *EMBO Rep.* **17**, 648–658. <https://doi.org/10.15252/embr.201540678> (2016).
119. Sheedy, P. & Medarova, Z. The fundamental role of miR-10b in metastatic cancer. *Am. J. Cancer Res.* **8**, 1674–1688 (2018).
120. Zhang, X., Sun, Y., Wang, P., Yang, C. & Li, S. Exploration of the molecular mechanism of prostate cancer based on mRNA and miRNA expression profiles. *Onco Targets Ther.* **10**, 3225–3232. <https://doi.org/10.2147/OTT.S135764> (2017).
121. Wang, J. *et al.* Prognostic value and function of KLF4 in prostate cancer: RNAi and vector-mediated overexpression identify KLF4 as an inhibitor of tumor cell growth and migration. *Cancer Res.* **70**, 10182–10191. <https://doi.org/10.1158/0008-5472.CAN-10-2414> (2010).
122. Jamaspishvili, T. *et al.* Clinical implications of PTEN loss in prostate cancer. *Nat. Rev. Urol.* **15**, 222–234. <https://doi.org/10.1038/nrurol.2018.9> (2018).
123. Yan, J. W., Lin, J. S. & He, X. X. The emerging role of miR-375 in cancer. *Int. J. Cancer* **135**, 1011–1018. <https://doi.org/10.1002/ijc.28563> (2014).
124. Khan, M. A. *et al.* RBM20 regulates circular RNA production from the titin gene. *Circ. Res.* **119**, 996–1003. <https://doi.org/10.1161/CIRCRESAHA.116.309568> (2016).
125. Barth, A. S. *et al.* Identification of a common gene expression signature in dilated cardiomyopathy across independent microarray studies. *J. Am. Coll. Cardiol.* **48**, 1610–1617. <https://doi.org/10.1016/j.jacc.2006.07.026> (2006).
126. Li, M. *et al.* MiR-1-3p that correlates with left ventricular function of HCM can serve as a potential target and differentiate HCM from DCM. *J. Transl. Med.* **16**, 161. <https://doi.org/10.1186/s12967-018-1534-3> (2018).

127. Ikeda, S. *et al.* Altered microRNA expression in human heart disease. *Physiol. Genomics* **31**, 367–373. <https://doi.org/10.1152/physiolgenomics.00144.2007> (2007).
128. Onrat, S. T., Onrat, E., Ercan Onay, E., Yalim, Z. & Avsar, A. The genetic determination of the differentiation between ischemic dilated cardiomyopathy and idiopathic dilated cardiomyopathy. *Genet. Test Mol. Biomark.* **22**, 644–651. <https://doi.org/10.1089/gtmb.2018.0188> (2018).
129. Satoh, M., Minami, Y., Takahashi, Y., Tabuchi, T. & Nakamura, M. Expression of microRNA-208 is associated with adverse clinical outcomes in human dilated cardiomyopathy. *J. Cardiol. Fail.* **16**, 404–410. <https://doi.org/10.1016/j.cardfail.2010.01.002> (2010).
130. Fang, L. *et al.* Enhanced breast cancer progression by mutant p53 is inhibited by the circular RNA circ-Ccnb1. *Cell Death Differ.* **25**, 2195–2208. <https://doi.org/10.1038/s41418-018-0115-6> (2018).
131. Yang, Q. *et al.* A circular RNA promotes tumorigenesis by inducing c-myc nuclear translocation. *Cell Death Differ.* **24**, 1609–1620. <https://doi.org/10.1038/cdd.2017.86> (2017).
132. Yang, Y. *et al.* Novel role of FBXW7 circular RNA in repressing glioma tumorigenesis. *J. Natl. Cancer Inst.* <https://doi.org/10.1093/jnci/djx166> (2018).
133. Du, W. W. *et al.* Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2. *Nucleic Acids Res.* **44**, 2846–2858. <https://doi.org/10.1093/nar/gkw027> (2016).
134. Liu, C. X. *et al.* Structure and degradation of circular RNAs regulate PKR activation in innate immunity. *Cell* **177**, 865–880. <https://doi.org/10.1016/j.cell.2019.03.046> (2019).
135. Salmena, L., Poliseno, L., Tay, Y., Kats, L. & Pandolfi, P. P. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?. *Cell* **146**, 353–358. <https://doi.org/10.1016/j.cell.2011.07.014> (2011).
136. Karreth, F. A. & Pandolfi, P. P. ceRNA cross-talk in cancer: when ce-bling rivalries go awry. *Cancer Discov.* **3**, 1113–1121. <https://doi.org/10.1158/2159-8290.CD-13-0202> (2013).
137. Chen, L. L. Linking long noncoding RNA localization and function. *Trends Biochem. Sci.* **41**, 761–772. <https://doi.org/10.1016/j.tibs.2016.07.003> (2016).
138. Zhang, X. *et al.* Circular RNA circNRP1 acts as a microRNA-149-5p sponge to promote gastric cancer progression via the AKT1/mTOR pathway. *Mol. Cancer* **18**, 20. <https://doi.org/10.1186/s12943-018-0935-5> (2019).
139. Du, W. W. *et al.* Foxo3 circular RNA promotes cardiac senescence by modulating multiple factors associated with stress and senescence responses. *Eur. Heart J.* **38**, 1402–1412. <https://doi.org/10.1093/eurheartj/ehw001> (2017).
140. Gupta, S. K. *et al.* Quaking inhibits doxorubicin-mediated cardiotoxicity through regulation of cardiac circular RNA expression. *Circ. Res.* **122**, 246–254. <https://doi.org/10.1161/CIRCRESAHA.117.311335> (2018).

Acknowledgements

We thank Dr Nicole Baldwin from Baylor Scott & White Research Institute for her technical help with the Cerina web server. We also thank Dr Xuan Wang from Baylor Scott & White Research Institute for her constructive comments of the manuscript.

Author contributions

J.G. and U.B. designed the study and circRNA detection framework. U.B. collected and processed the data. J.C. and J.G. developed the statistical methodology. J.C. implemented the method and developed the software. J.G. wrote the manuscript and interpreted the data with contribution from U.B. and J.C. All authors reviewed and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-020-78469-x>.

Correspondence and requests for materials should be addressed to J.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020