

# CRAFT: a bioinformatics software for custom prediction of circular RNA functions

Anna Dal Molin, Enrico Gaffo , Valeria Difilippo, Alessia Buratin, Caterina Tretti Parenzan, Silvia Bresolin and Stefania Bortoluzzi 

Corresponding authors: Stefania Bortoluzzi, Department of Molecular Medicine, University of Padova, Padova, Italy; Interdepartmental Research Center for Innovative Biotechnologies (CRIBI), University of Padova, Padova, Italy. Tel.: +39-049-827-6502; Fax: +39-049-827-6208; E-mail: stefania.bortoluzzi@unipd.it; Anna Dal Molin, Department of Molecular Medicine, University of Padova, Padova, Italy. Tel.: +39-049-827-6502; Fax: +39-049-827-6208; E-mail: anna.dalmolin@unipd.it

## Abstract

Circular RNAs (circRNAs), transcripts generated by backsplicing, are particularly stable and pleiotropic molecules, whose dysregulation drives human diseases and cancer by modulating gene expression and signaling pathways. CircRNAs can regulate cellular processes by different mechanisms, including interaction with microRNAs (miRNAs) and RNA-binding proteins (RBP), and encoding specific peptides. The prediction of circRNA functions is instrumental to interpret their impact in diseases, and to prioritize circRNAs for functional investigation. Currently, circRNA functional predictions are provided by web databases that do not allow custom analyses, while self-standing circRNA prediction tools are mostly limited to predict only one type of function, mainly focusing on the miRNA sponge activity of circRNAs. To solve these issues, we developed CRAFT (CircRNA Function prediction Tool), a freely available computational pipeline that predicts circRNA sequence and molecular interactions with miRNAs and RBP, along with their coding potential. Analysis of a set of circRNAs with known functions has been used to appraise CRAFT predictions and to optimize its setting. CRAFT provides a comprehensive graphical visualization of the results, links to several knowledge databases, and extensive functional enrichment analysis. Moreover, it originally combines the predictions for different circRNAs. CRAFT is a useful tool to help the user explore the potential regulatory networks involving the circRNAs of interest and generate hypotheses about the cooperation of circRNAs into the modulation of biological processes.

**Keywords:** CircRNA, bioinformatics, computational pipeline, functional prediction, regulatory network

## Introduction

Circular RNAs (circRNAs) are RNA molecules in which, in a process called backsplicing, a downstream 5' splice site is covalently linked to an upstream 3' splice site giving rise to a circle. The backsplice most frequently, although not always, joins canonical exons, with respect to linear transcript annotation [1].

These peculiar RNAs often exhibit cell type- and tissue-specific expression [2, 3] and can regulate several biological processes with different mechanisms. Moreover, circRNAs are emerging as key oncogenic or tumor suppressor molecules whose expression is dysregulated

in disease, cancer and when genomic translocations occur [4–6]. They represent stable diagnostic and prognostic markers [7, 8] and, most interesting, the identification of new disease mechanisms involving circRNAs has the potential to indicate novel therapeutic targets.

CircRNA functions mostly involve sequence-specific binding with other nucleic acids or proteins, or specific coding potential. One prominent mechanism whereby circRNAs function is by sponging microRNAs (miRNAs), through one or multiple miRNA-response elements (MRE). CircRNAs thus act as competitive endogenous

**Anna Dal Molin** is post-doc at the Computational Genomics Laboratory at the Department of Molecular Medicine, University of Padova. Her research interests include bioinformatics and transcriptomics, circular RNAs in leukemias and development of computational methods for circular RNA function prediction.

**Enrico Gaffo** is post-doc at the Computational Genomics Laboratory at the Department of Molecular Medicine, University of Padova. His research interests include circular RNA, microRNA, advanced methods for RNA-seq data analysis and bioinformatics applied to cancer research.

**Valeria Difilippo** graduated in Industrial Biotechnologies at the University of Padova working on circRNA prediction for her master thesis.

**Alessia Buratin** is PhD student in Biosciences (curriculum Genetics, Genomics and Bioinformatics) of the University of Padova. Her main interests are biostatistics and bioinformatics, transcriptomics of hematologic malignancies and circular RNA biogenesis.

**Caterina Tretti Parenzan** is a PhD student in Pediatric Oncoematology of the University of Padua. Her research is mostly focused on the study of the oncogenic molecular mechanisms involving circRNAs in pediatric acute lymphoblastic leukemia.

**Silvia Bresolin** is assistant professor of Molecular Biology at the Department of Women and Child Health Department of the University of Padova. Her field of interest is focused on pediatric leukemia transcriptomics and genomics alterations, as well as functional genomics and development of novel disease models.

**Stefania Bortoluzzi** is associate professor of Applied Biology at the Department of Molecular Medicine of the University of Padova, where she leads the Computational Genomics Laboratory. Her research interests include cancer genomics and transcriptomics, bioinformatics, systems biology, noncoding RNAs, circular RNAs, exosomal RNAs and hematologic malignancies.

**Received:** November 10, 2021. **Revised:** December 10, 2021. **Accepted:** December 26, 2021

© The Author(s) 2022. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

RNAs (ceRNA) and regulate the expression of miRNA-target genes (TG) [9]. Several key oncogenic circRNA-miRNA-gene axes have been described, impacting all cancer hallmarks [10]. With this mechanism, circRNAs also regulate key epigenetic ‘writers’ [11] controlling DNA methylation and histone modifications.

CircRNAs can also modulate the activity of RNA-binding proteins (RBP), a large class of molecules involved in most biological processes, since they regulate gene expression at the transcriptional and post-transcriptional levels [12–14]. The protein-circRNA interaction is sequence-specific and RBP-response elements (RRE) can be predicted using pattern recognition [15].

Finally, beyond exerting functions typical of long non-coding RNAs, circRNAs can be translated into circRNA-encoded peptides (CEP) [16–18]. CEP generated by translation of open reading frames (ORF) encompassing the backsplice junction, which are not present in linear transcripts, are circRNA-specific. CircRNAs can be also translated using long ORFs continuing along the circle [19, 20] using it as a ‘Möbius strip’ [21] (‘rolling circle translation’).

The same circRNA can carry multiple functional sequences, thus playing diverse functions also with different mechanisms. For instance, for circZNF609 both translation to a peptide and ceRNA activity have been described [17, 22]. Also for circFOXO3 both interaction with proteins and miRNA sponging activity is reported in literature [23, 24].

The function of most of the thousands of circRNAs discovered so far is still undetermined. Therefore, the prediction of circRNA function is instrumental to generate specific hypotheses about their molecular interactions and biological roles to be experimentally investigated, and to interpret the possible impact of circRNA dysregulation in disease.

Several databases have been released with the goal of collecting information regarding known circRNAs. CircBase reports circRNA expression in different biological samples [25], CircR2Disease [26] and Circ2Traits [27] contain, respectively, experimentally validated and potential circRNA association with diseases. CircNet [28] is specific for circRNA-miRNA-mRNA interaction networks, whereas circ2GO [29] merges circRNA expression, gene ontology and miRNA prediction. CircInteractome [30] and CircAtlas [31] collect data about miRNA and RBP binding to circRNAs. RiboCIRC [32] and TransCirc [33] provide information about circRNA coding ability. Importantly, to our knowledge, no databases provide predictions of the three above-described circRNA functions together. In addition, all of them provide precomputed data and do not allow custom predictions. Often, data browsing is difficult and the user has no control about the correspondence between circRNA sequence and functional predictions.

A few tools are available hitherto for circRNA functional prediction. Among them, Cerina R shiny web application [34] and ACT web server [35] predict circRNA functions exclusively based on the ceRNA model.

Instead, CircRNAprofiler [36] and CircCode [37] are both dedicated to the coding potential of circRNAs, and the second needs ribosome profiling (Ribo-seq) data as input. In summary, all of the available methods allow the prediction of a single function of circRNAs and there is a high need for software tools to comprehensively predict circRNA functions. The circRNA sequence is a prerequisite for functional predictions, but most methods to detect circRNA from RNA-seq data return only the backsplice junction coordinates. FcircSEC R package [38] retrieves circRNA sequence using the longer transcript of the gene as reference.

To fill this gap, we developed CRAFT (CircRNA Function prediction Tool), a new computational pipeline to predict circRNA putative molecular interactions with miRNAs and RBP, and circRNA coding potential. Plus, CRAFT allows investigating complex regulatory networks involving circRNAs acting in a concerted way, such as by decoying the same miRNAs or RBP, or miRNAs sharing TG. This new tool facilitates the interpretation of circRNA biological and pathogenetic roles, and can guide the experimental investigation of circRNA-involving mechanisms.

## Methods

### CircRNA sequence reconstruction

CircRNA sequences to be used for the predictions are obtained using the CircExtractor module described below. Bedtools v2.28.0 [39] was used for file manipulation and sequence extraction, as described in the Results section. The circRNA genomic coordinate file and the build genomic reference file are in BED format. The predicted circRNA sequences in FASTA format are saved in the result directory. To run predictions, circRNA sequences are modified in order to also detect functional elements overlapping the backsplice region. A pseudocircular sequence is created repeating the first 20 nt of the sequence at the end of each circRNA for MRE and RRE prediction, whereas the whole sequence is repeated twice to detect ORF crossing the backsplice junction. All the prediction coordinates are relative to the circRNA sequence used.

### MiRNA-binding site prediction

MiRNA sequences downloaded from miRBase release 22.1 are included in the tool. A filter is then applied to retrieve miRNA of the species selected by the user. MiRanda 3.3a [40] and PITA [41] tools are used to predict miRNA binding sites in circRNA sequences. MiRanda parameters set as default by CRAFT are -sc 80 -en -15; PITA is run with default parameters.

AGO2 binding data were downloaded from the doRiNA database [42], merging all 17 files with AGO2 binding information, and converted in hg38 coordinates through UCSC liftOver.

Only miRNA binding sites predicted both by miRanda and PITA, and overlapping with AGO2 binding sites, are kept. The AGO2 binding site file is optional. A different

file with AGO2 binding data can be provided by the user in BED format.

Validated miRNA-TG were retrieved from three different databases (miRecords, mirTarBase and Tarbase) using the *multiMiR* 1.8.0 R package [43], selecting the 'strong' validation category of each database.

### RBP binding site prediction

RBP binding sites were predicted through beRBP [44] tool, modified to give in output the 'voteFrac' score for all predictions, and run in 'general mode' searching for all the 143 RBP and 175 Position Weight Matrices of the tool database.

Result tables are given in plain text tabular format, keeping the output similar to that of the tool's original output, to allow deep inspection to expert users.

### Functional enrichments

The *multiMiR* package is used to perform enrichments of miRNA association with drugs and diseases. Functional enrichments are performed on validated TG of miRNAs with predicted binding sites in circRNA sequences. DOSE 3.18.2 package [45] is used for over-representation analysis, *clusterProfiler* 4.0.5 package [46] for Gene Ontology analysis and KEGG over-representation test, *ReactomePA* 1.30.0 package [47] for Reactome pathway analysis, *meshes* 1.12.0 package [48] for MeSH enrichment analysis. Functional enrichments on validated target genes of miRNAs with predicted binding sites in circRNA sequences can be performed only for *Homo sapiens* (hsa), *Mus musculus* (mmu) and *Rattus norvegicus* (rno) species.

Functional enrichments are performed also on RBP with predicted binding sites in circRNA sequences using the *UniProtR* 2.1.0 R package [49]; Gene Ontology, KEGG enrichment, Reactome pathway and RBP-disease association analyses are obtained. Protein expression, and functional and miscellaneous information associated with RBP were also retrieved.

### Open reading frames prediction

Putative ORF are predicted through ORFfinder tool, searching for ORF with the canonical 'ATG' start codon, with minimal ORF length of 30 nt, and detected on plus strand. All four kinds of ORFfinder output formats are produced.

### Graphical visualization

The R packages *ggplot2* 3.3.5 [50], *phreatmap* 1.0.12, *tidyverse* 1.3.1 [51], *car* 3.0–11 [52] and *enrichplot* 1.12.2 [53] are used for result graphical visualization, whereas tables are printed through the *DT* 0.19 R package.

For the default graphical visualization, CRAFT filters the results with more stringent parameters: miRanda score > 80 and free energy < -15 kcal/mol, PITA  $\Delta G_{\text{duplex}}$  < -15 kcal/mol and  $\Delta G_{\text{open}}$  > -15 kcal/mol are set, beRBP voteFrac  $\geq$  0.15.

### Assessment and sample analyses

For all the presented analyses, known circRNA back-splice coordinates were retrieved from CircBase. CircRNA binding with miRNAs or RBP, and the coding potential information, were collected from literature (Table 1). CircRNA annotation and genomic sequence are based on the Ensembl GRCh38 v.93 human genome. CRAFT was run with default parameters.

CRAFT sensitivity was calculated as true positive predictions (TP, known elements correctly predicted) over the total number of predictions (TP + FN, where FN stands for false negatives). The prediction ratio was calculated as the mean number of predictions at a certain setting over the mean number of predictions obtained at the less stringent condition tested.

## Results

### Overview of the CRAFT workflow

The CRAFT analysis workflow comprises four main steps (Figure 1). First, the circRNA sequence is retrieved by a custom tool, as described in the Results section 'The CircExtractor module'. Second, using the combination of different methods for miRNA and RBP binding sites detection, and ORF finding, complete circRNA functional predictions are obtained. Afterwards, the postprocessing phase integrates the predictions obtained for each circRNA with knowledge databases for functional enrichment analysis of miRNAs, RBP and known miRNA-TG associated with the circRNAs in the previous step. Tabular and graphical visualization of results is provided for each circRNA. In the last CRAFT step, the integration of predictions for all the circRNAs given in input let to identify miRNAs, TG and RBP in common between different circRNAs, and in turn, multiple circRNAs potentially impacting the same biological processes.

CRAFT is a self-standing tool written in Bash and R languages. The Docker [77] containerization makes it usable and portable on all systems. CRAFT accepts circRNAs of any species: the genome sequence and the gene annotation files must be given in input in FASTA and GTF formats, respectively. The user has full control of the parameters used for each prediction tool and can choose to perform either all or only part of the available predictions. Moreover, after the first run, the user can adjust the result visualization by modulating the output filtering parameters, without the need of computing the predictions again, thus saving computational time.

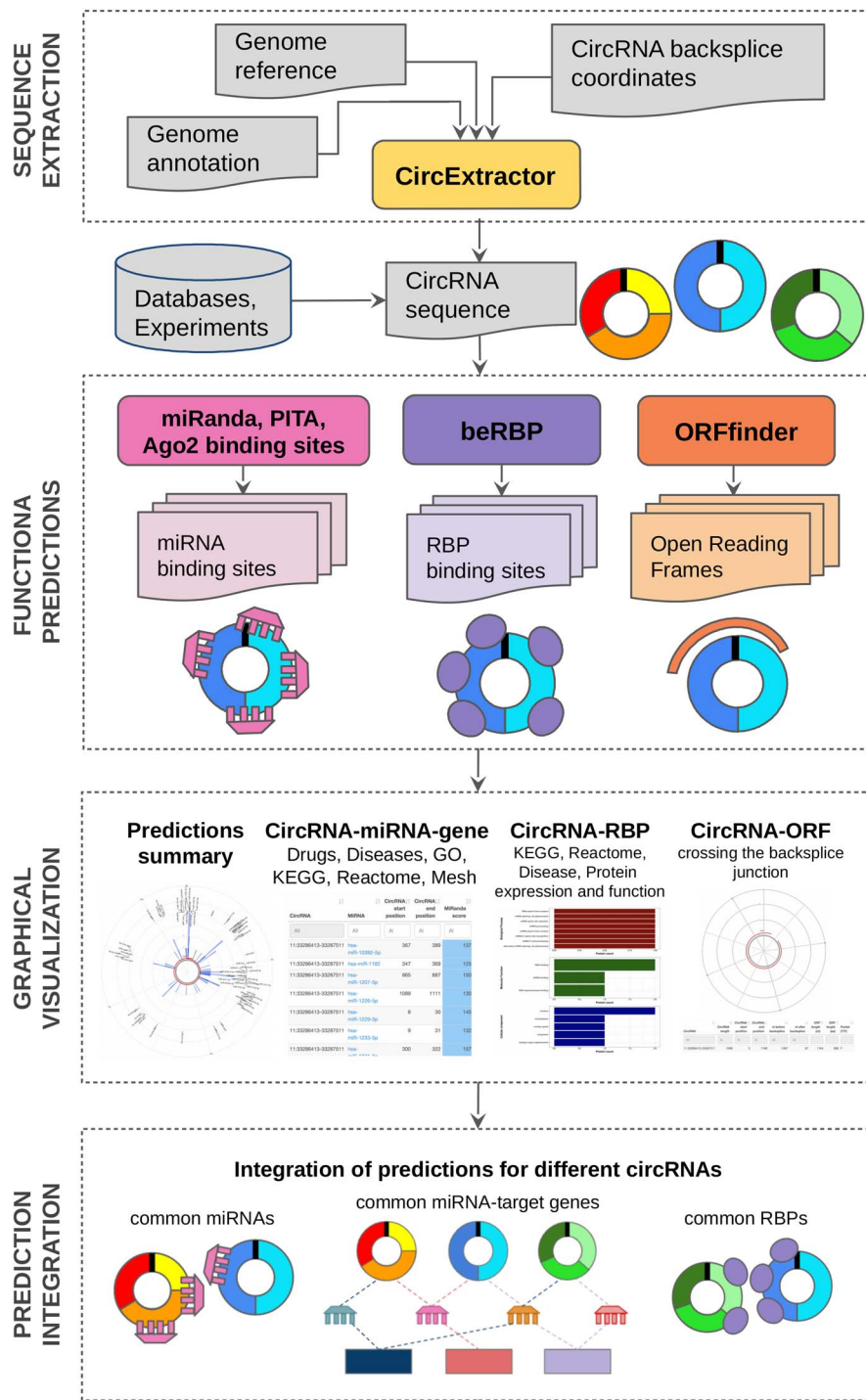
### CRAFT assessment and parameters optimization

We assessed how the parameter tuning and different tool combinations affected the whole number of sites predicted by CRAFT and its capability to detect functional elements, in order to optimize the default setting of the tool. With a literature survey, we collected 20 circRNAs with experimentally determined functions (Table 1), including 26 functional elements, that were used as a test set. For each prediction type, we tested whether

**Table 1.** CircRNAs used for the CRAFT assessment analysis, with previously validated functional elements and biological or pathogenetic roles

CircRNA	Backsplice	Validated functional element	Validated interaction/peptide (aa)	CircRNA binding validation	Tissue	Function	Reference
circBCRC3	2:231075511-231087181:+	MRE	miR-182-5p	Biotin-labeled RNA pull-down assay, FISH, transfection	Bladder cancer	Tumor suppressor	[54]
circHIPK3*	11:33286413-33287511:+	MRE	miR-124-3p	Luciferase reporter assay, biotin-labeled miRNA pull-down assay	Liver cancer	Oncogenic	[55]
circHIPK3*	11:33286413-33287511:+	MRE	miR-653-5p, miR-338-3p	RIP assay, luciferase reporter assay, biotin-labeled miRNA pull-down assay	Gastric cancer	Oncogenic	[56]
circNR3C1	5:143399656-143400852:-	MRE	miR-27a-3p	Biotin-labeled miRNA pull-down assay, FISH	Bladder cancer	Tumor suppressor	[57]
circNRIP1	21:15014344-15043574:-	MRE	miR-149-5p	Biotin-labeled RNA pull-down assay, biotin-labeled miRNA assay, dual-luciferase reporter assay, FISH, qRT-PCR	Gastric cancer	Oncogenic	[58]
circRNF20	9:101540203-101540975:+	MRE	miR-487a	Luciferase reporter assay, FISH	Breast cancer	Oncogenic	[59]
circSHKBP1	19:40583398-40583717:+	MRE	miR-582-3p	Luciferase reporter assay, biotin-labeled RNA pull-down assay, FISH, qRT-PCR	Gastric cancer	Oncogenic	[60]
circZNF609*	15:64499292-64500166:+	MRE	miR-15a-5p, miR-15b-5p	RNA pull-down assay, luciferase reporter assay	Hepatocellular carcinoma	Oncogenic	[61]
circZNF609*	15:64499292-64500166:+	MRE	miR-138-5p	Luciferase reporter assay, RIP assay	Renal carcinoma	Oncogenic	[62]
circDCUN1D4	4:51863437-51886638:+	RRE	HuR(ELAVL1)	CLIP, RIP assay, biotin RNA pull-down assay, Western blot	Lung adenocarcinoma	Tumor suppressor	[63]
circHIPK3	11:33340968-33341686:+	RRE	FUS	Western blot, correlation analysis, knockdown, qRT-PCR	Cervical cancer	Oncogenic	[64]
circMBNL1	3:152299405-152300367:+	RRE	MBNL1	RIP assay	Neuronal tissues	Gene regulator, circRNA biogenesis	[65]
circNSUN2	5:6623214-6625669:-	RRE	IGF2BP2	RNA pull-down assay, RIP, IF-FISH	Colorectal carcinoma	Oncogenic	[66]
circPABPN1	14:23324138-23324289:+	RRE	HuR(ELAVL1)	RIP, CLIP, biotin RNA pull-down assay	Cervical carcinoma	Negative regulation of PABPN1 mRNA and translation	[67]
circRHOBTB3	5:95755396-95763620:+	RRE	HuR(ELAVL1)	RNA pull-down assay, RNA immunoprecipitation assays	Colorectal cancer	inhibition/endogenous competing RNA	[68]
circSMARCA5*	4:143543509-143543972:+	RRE	SRSF1	eCLIP	Glioblastoma multiforme	Tumor suppressor	[69]
circXIAP	X:123888619-123892773:+	RRE	FUS	Mass spectrometry analysis, RNA pull-down assay	Prostate cancer	Oncogenic	[70]
circZKSCAN1	7:100023419-100024307:+	RRE	FMRP(FMR1)	RIP assay	Hepatocellular carcinoma	Tumor suppressor	[71]
circ-E-Cad	16:68811684-68815759:+	ORF	254	Monoclonal antibody, MS	Glioblastoma	Oncogenic	[72]
circFNDC3B	3:172247533-172251541:+	ORF	218	LC-MS/MS	Colon cancer	Tumor suppressor	[73]
circPICE1	10:94030683-94032252:+	ORF	411	LC-MS	Colorectal carcinoma	Tumor suppressor	[74]
circSHPRH	6:145888020-145894977:-	ORF	146	Antibody, LC-MS/MS	Glioblastoma	Tumor suppressor	[75]
circSMO	7:129205203-129206587:+	ORF	193	Ribosomes enrichment assay, immunoblot (IB), MS	Glioblastoma	Oncogenic	[76]
circZNF609	15:64499292-64500166:+	ORF	251	Polysome association, western blot	Muscle	Implied in myogenesis	[16]

\*CircRNAs used to illustrate CRAFT output. MRE, miRNA recognition element; RRE, RBP recognition element; ORF, open reading frame. Is near to the circRNAs used to illustrate CRAFT output. MRE, miRNA recognition element; RRE, RBP recognition element; ORF, open reading frame.



**Figure 1.** CRAFT workflow, schematizing the pipeline input, the main analysis steps and the produced output.

different strategies and settings can recover the known functional elements in the test set (sensitivity), and how the prediction ratio (Methods section) were affected.

The results based on the analysis of the known MRE is presented in Figure 2A. The sensitivity of predictions obtained by miRanda, PITA and intersection thereof was constantly high ( $\geq 0.9$ ), and the large number of predictions provided by single methods can be controlled by their intersection, as shown by the line plot. At our tests, the addition of information about experimentally

determined AGO2 binding sites was useful. Particularly in combination with the intersection of miRanda and PITA predictions, it kept it with a good sensitivity ( $\geq 0.82$ ), while reducing the amount of predictions to only 6%, respective to the total set of predictions obtained from the union of miRanda and PITA (Figure 2A). The starting number of predictions to be intersected and combined with AGO2 binding sites is obviously reduced when more stringent thresholds for the four parameters provided by miRanda and PITA are set, as shown in Supplementary Figures 1

and 2. The scatterplot of prediction scores provided by miRanda (Supplementary Figure 1) and PITA (Supplementary Figure 2), with indications of the score pairs associated with known MRE, can help the user to understand the effect of stringency on MRE detection. For instance, keeping only the predictions with both scores over the first quartile for both methods, 7/11 known MRE were correctly detected (sensitivity 0.71). Using the median as threshold for the same four scores, the number of predictions decreased (prediction ratio 19%), still identifying 4/11 positives (sensitivity 0.36).

In summary, considering that the intersection of miRanda and PITA predictions and the overlapping with AGO2 binding sites controlled for the prediction number, while keeping in with a good sensitivity, we kept a relatively low stringency on miRanda and PITA parameters in CRAFT default setting ('M&P' in Figure 2A), to give the user a fairly complete set of predictions. Nevertheless, the interactive filtering provided by CRAFT allows the user to customize the number of predictions after inspecting the results.

The results obtained by tests on RRE predictions are shown in Figure 2B and C. All known RRE (9/9, sensitivity = 1) were detected at the lowest stringency of beRBP predictions score ( $\text{voteFrac} > 0$ ), with an average of 316 (range 301–326, median 319) predictions per circRNA (Figure 2B). At a less permissive condition ( $\text{voteFrac} \geq 0.1$ ), sensitivity remained high (0.89; with 8/9 detected) and 188 predictions per circRNA on average were obtained (range 68–316, median 158). Further increasing stringency ( $\text{voteFrac} \geq 0.15$  and  $\geq 0.2$ ), most of the known functional elements were still detected (sensitivity 0.78 and 0.56, respectively), while considerably reducing the number of predictions to 42% and 27% of the total, respectively (Figure 2B). At the most stringent score tested, only 1/9 known RRE element (SRSF1 binding circSMARCA5) was detected, being anyhow the unique passing the filter for circSMARCA5, with an average number of 5 predictions per circRNA (range 0–36, median 0; Figure 2B). Figure 2C shows more detail about the distribution of RBP binding site prediction scores for each circRNA, indicating the scores of known binding sites. Plus, for each  $\text{voteFrac}$  threshold, the boxplot on the right shows the distribution of the number of predictions obtained over the analyzed circRNAs. In summary, the best balance between sensitivity and overall number of predictions was observed at a  $\text{voteFrac} \geq 0.15$ , and this optimal threshold was set as default in the CRAFT RBP analysis.

Coding potential analysis predicted an ORF overlapping the backsplice junction for all six tested circRNAs. In five out of six cases (83.3%), the predicted peptide length exactly corresponded to that of the validated peptide (Figure 2D). The difference in the circ-E-Cad predicted ORF was due to a slight discrepancy between the sequence predicted by CircExtractor based on genome annotation and the sequence considered in the previous study [72]. The ORF and peptide were correctly predicted

by CRAFT when only the exons from the canonical transcript (Ensembl id: ENST00000261769.10) were used to predict the circRNA sequence.

### The CircExtractor module

The circRNA nucleotide sequence is necessary to perform functional predictions. Therefore, we included in our pipeline a module that takes as input the circRNA backsplice coordinates, the genome sequence and the gene annotation, and reconstructs the circRNA sequence (Figure 1). If circRNA sequences are available to the user, this step can be skipped.

The CRAFT CircExtractor module reconstructs the putative circRNA sequence by intersecting the reference annotation with the backsplice coordinates. Four main cases are considered: (i) if the backsplice joins the ends of two annotated exons, the predicted circRNA sequence contains all and only the exons within the backsplice coordinates (Figure 3A); (ii) if one or both backsplice ends map within an exon, only the exon segment within the backsplice coordinates is considered, along with all other entire exons (Figure 3B); (iii) if both backsplice ends map in an intronic or intergenic region, the intronic or intergenic part between the backsplice ends is extracted (Figure 3C); and (iv), if one backsplice end maps into an exon, whereas the other end maps into an intron or an intergenic region, the sequence will contain all the exons within the backsplice coordinates, plus the intronic/intergenic segment between the backsplice start/end and the first/last exon (Figure 3D).

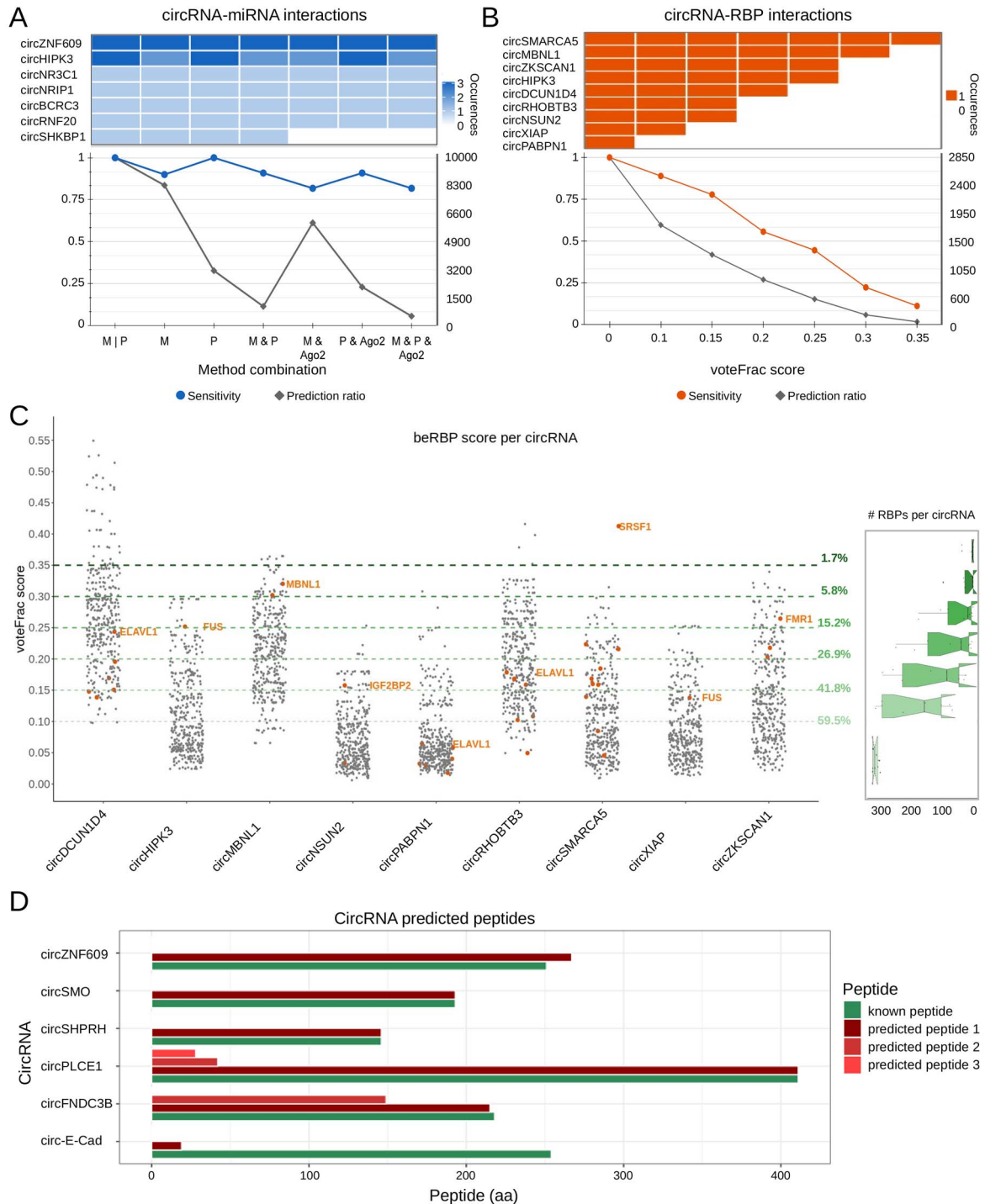
### CRAFT output

To illustrate CRAFT output, post-processing analyses and data display, we applied CRAFT to three previously characterized circRNAs, circSMARCA5, circHIPK3 and circZNF609, from the above-considered group.

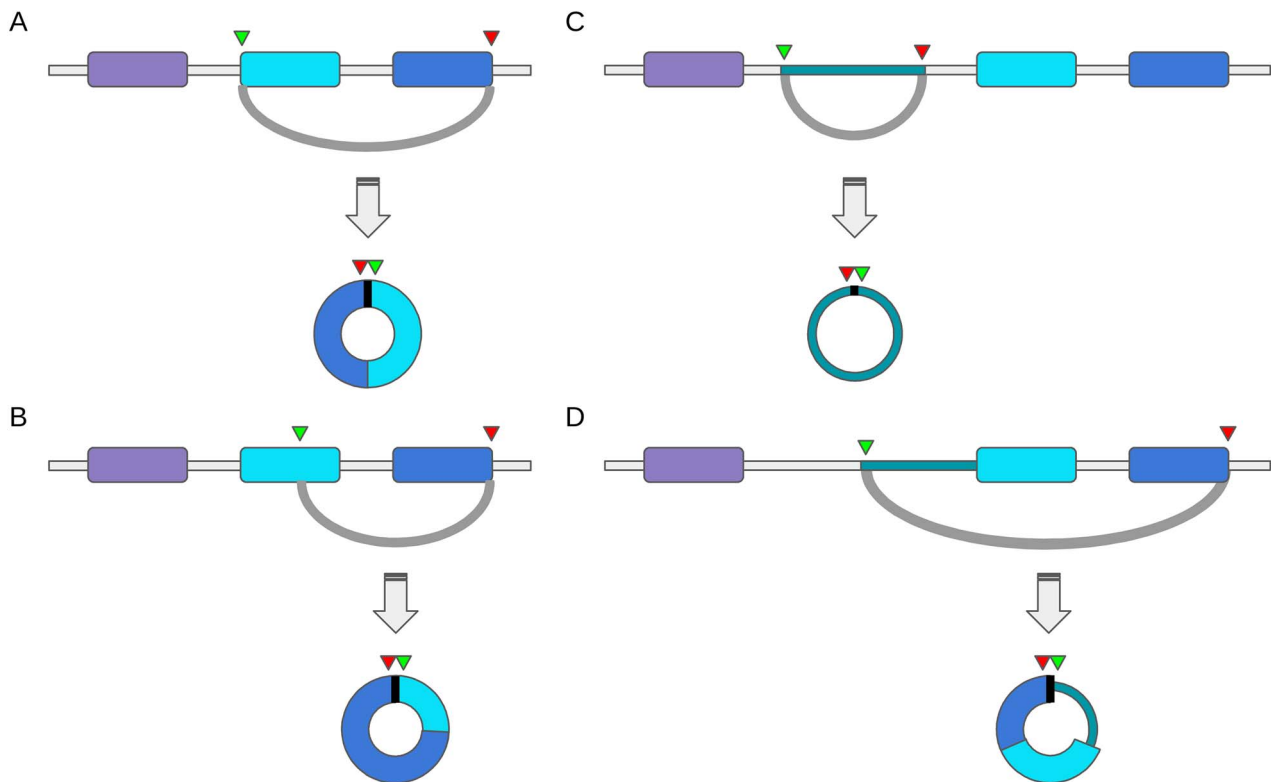
CRAFT outputs multiple HTML pages with interactive tables and figures, describing the predicted miRNA and RBP binding sites, and the putative ORF (Figure 1). Prediction result tables link miRNAs, genes and proteins to several databases (miRBase, GeneCards, NCBI Entrez, Ensembl, Uniprot and NCBI Pubmed). All result tables are downloadable in comma-separated value (CSV) plain text, Excel, and PDF formats, can be copied to clipboard or printed, allow filtering and ordering, and display colored bars to highlight important features.

A summary of functional predictions obtained for circZNF609 was used to illustrate the main features of CRAFT output (Figures 4 and 5). For each circRNA, a dedicated page is given with detailed predictions. At first, a circular plot shows predicted miRNA and RBP binding positions in the circRNA sequence, and the longest predicted ORF (Figure 4A). Next, three sections are dedicated to miRNA, RBP and ORF predictions, respectively.

The miRNA section provides, after a circular plot focusing exclusively on miRNA binding sites, a table with the start and end binding positions and the prediction scores (Figure 4B). Then, the list of all validated TG



**Figure 2.** Assessment analysis of circRNAs with known functions. (A) The upper panel shows the number of validated circRNA-miRNA interactions detected by different methods and combinations thereof for each of the considered circRNAs, whereas the line plot below reports the sensitivity and the prediction ratio in the same conditions (M, miRanda; P, PITA; M | P, union of M and P; M & P, intersection of M and P; the same with AGO2, predicted sites overlapping with experimentally determined AGO2 binding sites); the right axis shows the mean number of predictions for circRNA; (B) the upper panel shows the number of validated circRNA-RBP interactions detected by different beRBP voteFrac thresholds, whereas the line plot below reports the sensitivity and the prediction ratio in the same conditions; the right axis shows the mean number of predictions for circRNA; (C) distribution of RBP binding site prediction scores for each circRNA (known binding sites are shown as orange points, the name is shown for the known site with the highest score; the percentage on the right indicates the prediction ratio of the corresponding voteFrac threshold); the notch boxplot on the right represents the distribution of the number of predicted RBP per circRNA at a certain voteFrac score; (D) barplot of the amino acid (aa) length of known (in green) and predicted (in red) peptides for each circRNA.



**Figure 3.** The CircExtractor module of the CRAFT pipeline reconstructs the circRNA sequence using the backsplice coordinates and the genome annotation and reference. (A–D) show the four cases encompassed by the method, according to the position of the backsplice ends (green and red triangles represent the backsplice start and end, respectively), in exonic (A–B), intronic (C) and intergenic sequences, and combinations thereof (D).

is provided for each miRNA, along with different summarization tables, including the number of TG associated with each miRNA (Figure 4C), the list of miRNAs associated with each TG, and miRNA-TG interactions associated with drugs or diseases. Functional enrichment results on validated TG are displayed, including disease-associated genes (Figure 4D), Gene Ontology, KEGG analysis, Reactome pathways (Figure 4E) and MeSH terms. Of note, miR-15a-5p and miR-15b-5p binding sites were detected for circZNF609, in line with literature data [17, 61, 62].

The RBP section, after a circular plot with RBP binding positions and the corresponding interactive table (Figure 5A), displays results of functional enrichments performed on RBP, considering Gene Ontology, KEGG analysis, Reactome pathways (Figure 5B) and RBP-disease association (Figure 5C). Additional tables provide protein expression, functional and miscellaneous information associated with RBP.

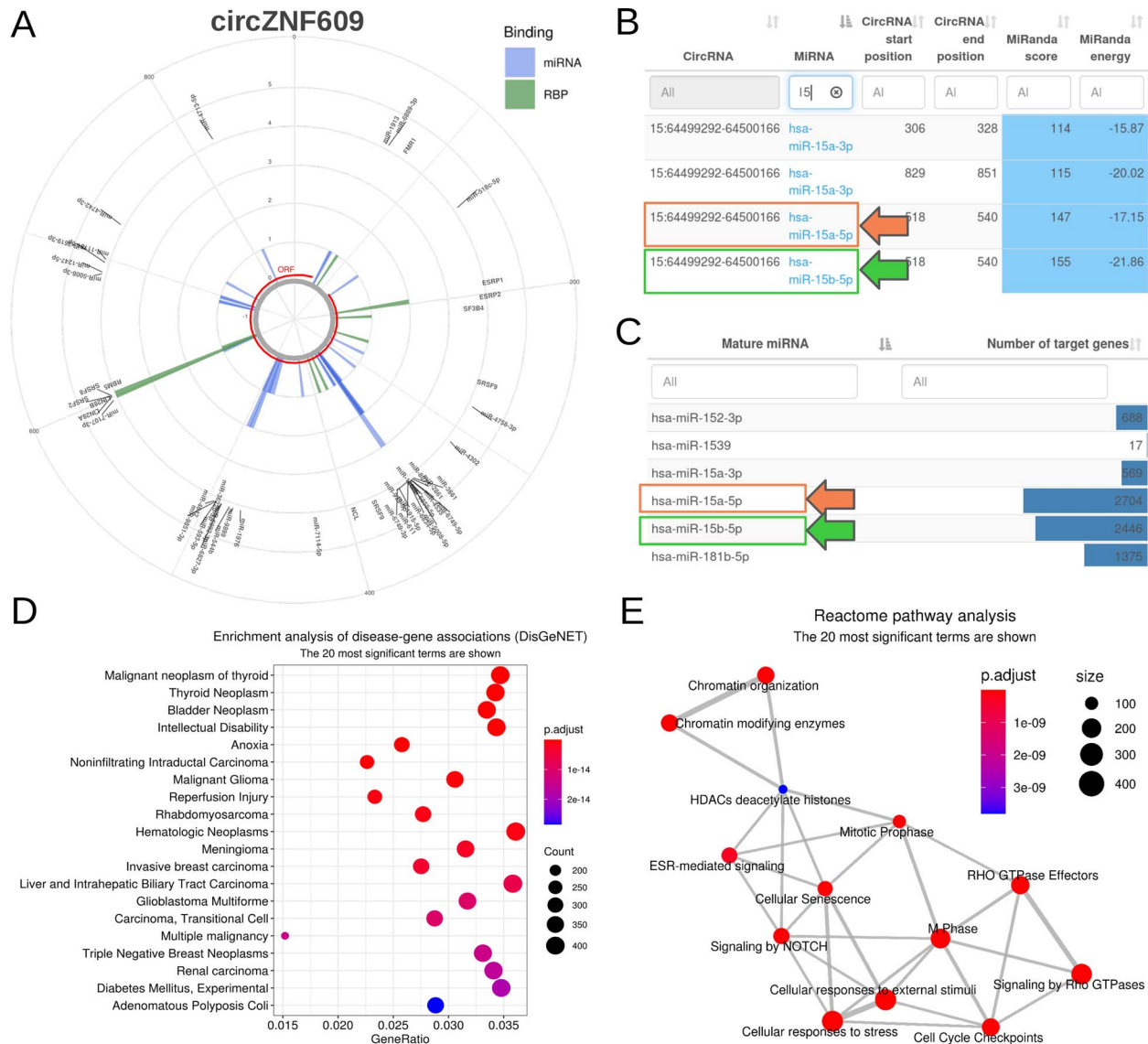
In the last section, the longest predicted ORF overlapping to the backsplice junction is shown by a circular plot (Figure 5D), while a table reports, for all the predicted ORF, the start and end positions, the distance from the backsplice junction, and the predicted coding and peptide sequences in FASTA format (Figure 5E).

A default set of high-scoring predictions is reported and visualized in the output HTML pages (see Methods).

The user can tune the filter stringency and regenerate the HTML pages with updated tables and figures with minimal computation time. Two filtering modes are allowed: (i) a ‘direct mode’, by directly inputting each prediction tool threshold parameter; (ii) a ‘quantile mode’, to filter a subset of predictions based on a selected quantile (f.i. display and analyze only the top 25% high-scoring predictions for each tool).

In principle, different circRNAs can act as sponge for the same miRNA, reinforcing the desuppression of the miRNA targets. Multiple circRNAs can regulate the same gene also through different miRNAs (Figure 1). Thus, if more than one circRNA is given in input, CRAFT integrates altogether the predictions obtained for the different circRNAs. The common predictions are displayed through tables and figures, to provide valuable hints about the possible cooperation of circRNAs to the same regulatory axes. The 13 miRNAs with binding sites in the top 45% high-scoring predictions for both tools and common to multiple circRNAs, are shown in Figure 6A. Moreover, in the CRAFT output, circRNAs are connected to miRNA-validated TG through their predicted interactions with miRNAs. In addition to tables with TG potentially shared by circRNAs, CRAFT provides a network visualization of predicted circRNA-miRNA-TG axes in which multiple circRNAs are connected to the same TG through the same or different miRNAs. For instance, circHIPK3





**Figure 4.** CRAFT miRNA prediction output based on circZNF609 analysis. (A) Summary plot with predicted miRNA and RBP binding sites, and the longest ORF (with a stop codon), along the circRNA sequence; (B) table of CRAFT predicted miRNA binding sites in circZNF609 sequence, with start and end positions, and scores (orange and green boxes highlight predicted miR-15a-5p and miR-15b-5p binding sites, respectively); (C) table summarizing the number of TG per miRNA (box colors as in B); (D) plot of disease-miRNA TG association enrichment analysis; (E) network showing Reactome pathway enrichment analysis results on miRNA TG. In panels (B) and (C) only a small part of the corresponding table is shown.

and circZNF609 can potentially control the expression of a group of eleven genes by sponging miR-1207-5p (circHIPK3) or miR-5581-3p and miR-611 (circZNF609) (Figure 6B).

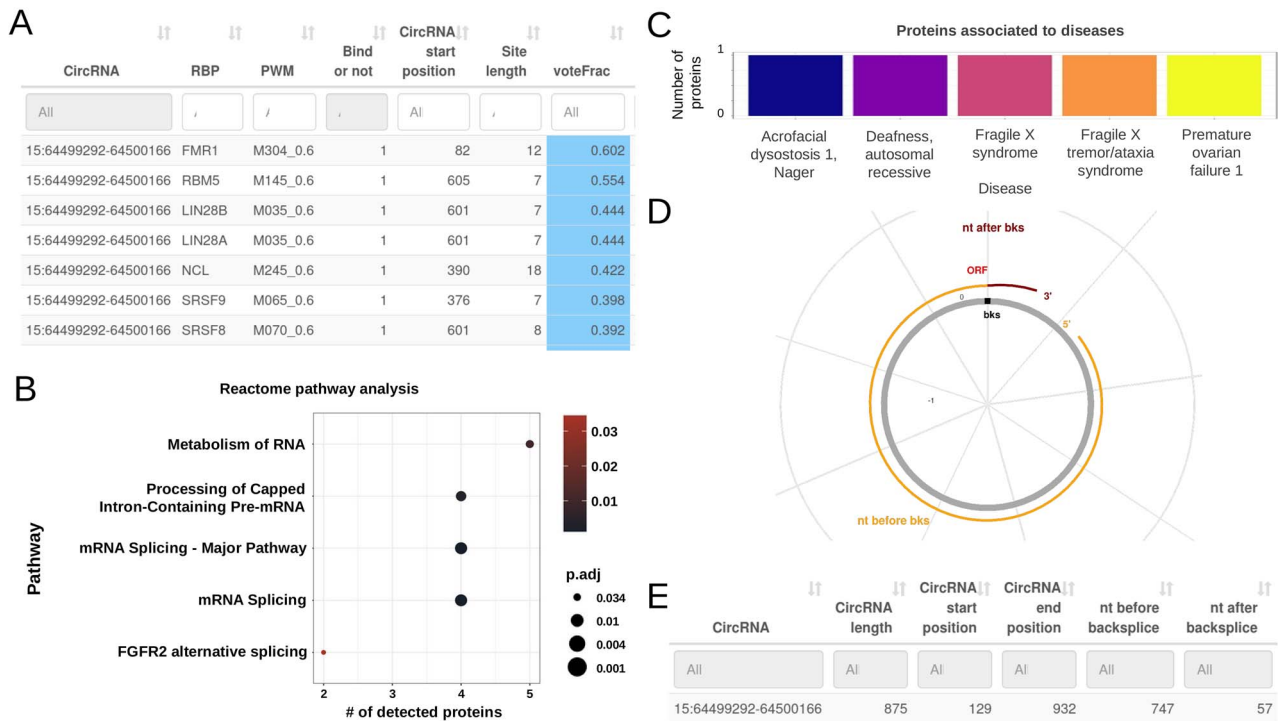
Similarly, CRAFT displays RBP associated with RRE predicted in multiple circRNAs. Considering all the predictions obtained with CRAFT default stringency, 48 RBP are commonly associated with all the three circRNAs. Of note, an interaction with the FUS RBP is predicted by CRAFT for circHIPK3, in line with literature data [78], as well as for circSMARCA5. The RBP with binding sites in the top 10% high-scoring predictions and common for multiple of the three circRNAs considered in our sample analysis are shown in Figure 6C. All the considered circRNAs have strongly predicted binding sites for SRSF1,

including circSMARCA5 whose interaction with SRSF1 has been previously demonstrated [69].

Finally a barplot (Figure 6D) and a detailed table present the number of predicted ORF for each circRNA, distinguishing between ORF with and without a stop codon.

## Discussion

Robust evidence on circRNA functions, involvement in biological processes, and oncogenic potential make these molecules extremely attractive for both fundamental and cancer research [79]. CircRNAs can be detected and quantified from RNA-seq data using appropriate software tools, including [80, 81]. Next, the challenge faced



**Figure 5.** CRAFT RBP and ORF prediction output based on circZNF609 analysis. (A) Table of CRAFT predicted RBP binding sites in circZNF609 sequence, with start position, binding probability and prediction score; (B) plot of Reactome enrichment analyses on RBP potentially interacting with circZNF609, according to CRAFT predictions; (C) barplot of RBP-disease associations; (D) plot of the longest predicted ORF with a stop codon in circZNF609 sequence; (E) table with start and end position, length and frame of predicted ORF. In panels (A) and (E) only a small part of the corresponding table is shown.

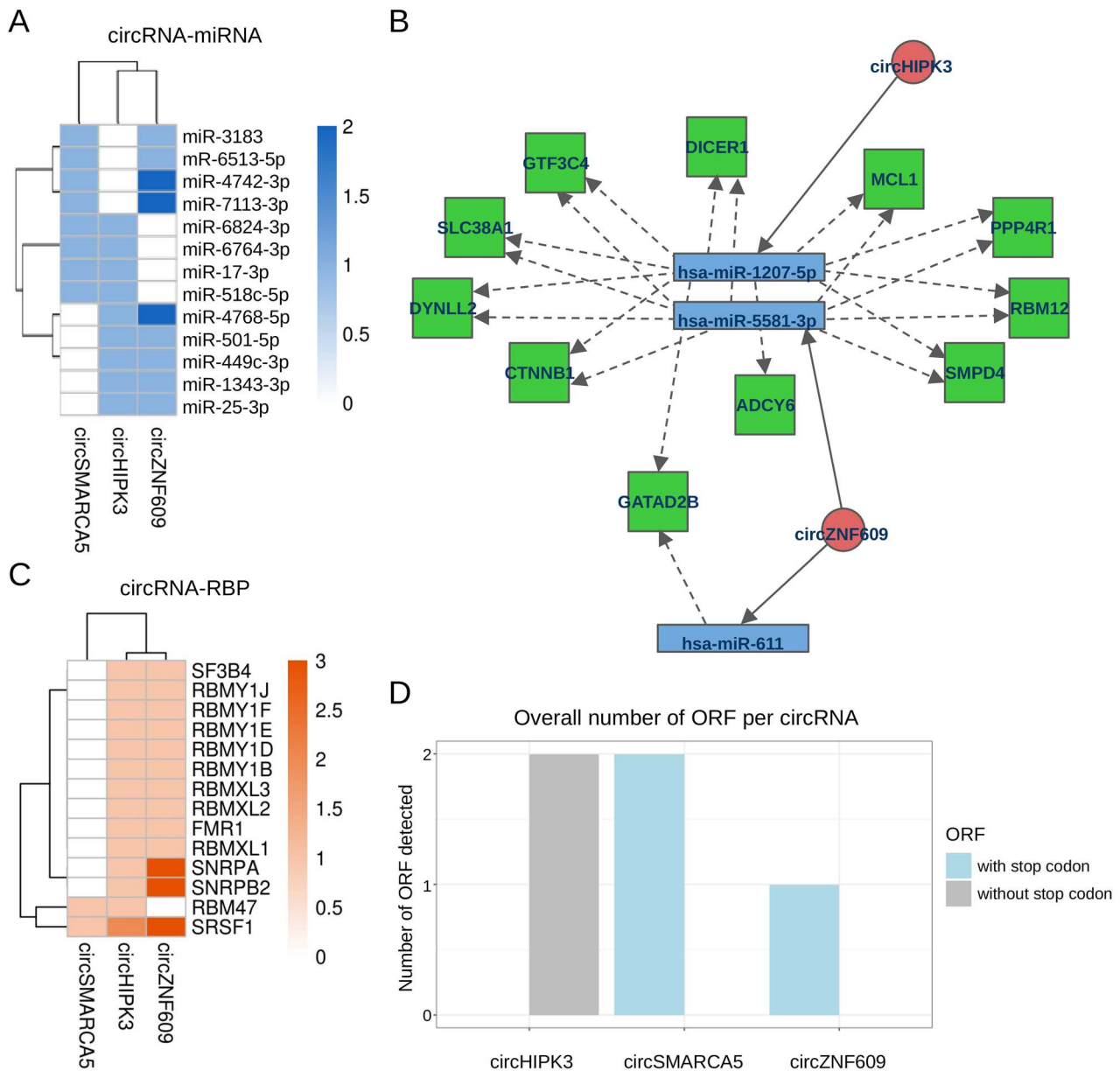
by studies aiming at defining circRNA roles in disease and cancer is the identification of circRNA-involving mechanisms. CircRNA screening by massive silencing or over-expression studies and experimental investigation can be of great help. Nevertheless, the prediction of circRNA potential interactions and functions is a precondition for the prioritization of circRNAs for functional experimental studies, for the interpretation of experiment results and the design of targeted mechanistic investigations. Indeed, even once a circRNA has been demonstrated to significantly impact cell features, such as proliferation, apoptosis, metabolism or differentiation, there is the need to disclose the involved molecules and regulatory axes.

CRAFT has been developed to overcome the limitations of the currently available circRNA function prediction methods. It allows the user to explore putative regulatory networks involving one or more circRNAs of interest, facilitating the interpretation of their biological and pathogenetic role. Moreover, CRAFT is also a highly portable tool, thanks to the containerization of the software, easy to use, and it requires minimal input data.

CRAFT presents four main advantages over the existing tools. First, it provides the putative circRNA sequence. The sequence extraction is based on annotated exons, following the observation that linear splicing patterns of circRNAs mainly join canonical exons [82]. The CircExtractor module in CRAFT merges all the overlapping exons within the circRNA coordinates. This works perfectly well for most cases, as indirectly demonstrated by our test on circRNAs proved to encode for peptides.

Of note, CRAFT correctly detected the known ORF of circSHPRH, which is formed by the circularization of four exons of the SHPRH gene. In specific instances, the sequence reconstructed by CircExtractor can be different from the sequence produced by the splicing of the exons included in the canonical transcript. This is exemplified by circ-E-Cad, whose backsplice ends include a region overlapped by 11 different transcripts that use alternative 5'- and 3'-exon ends in the linear splicing, while the validated circ-E-Cad follows the linear splicing pattern of the canonical longer coding transcript of the gene. In general, the uncertainty associated with circRNA sequence reconstruction based on exon annotations tends to increase with the size of the genomic region within the backsplice ends and, in particular, with the complexity of the alternative linear splicing pattern in the region. In this regard, our recommendation is to use the CRAFT sequence reconstruction module if the circRNA sequences are not known, or for massive analyses of several circRNAs. Nevertheless, in projects focusing on the functional experimental investigation of one or a few specific circRNAs, we advise using experimentally determined circRNA sequences, that provide a firmer ground to build predictions on.

The second main CRAFT asset is that it allows prediction of the three most important circRNAs functions, whose biological relevance is supported by robust literature [9–14, 16, 18–20]. The most investigated function of circRNAs is the ability to bind miRNAs. One limitation of miRNA binding site prediction algorithms



**Figure 6.** Predictions of different circRNAs are combined to facilitate hypothesis generation. (A) Heatmap of the miRNAs common to at least two circRNAs, considering the top 45% high-scoring miRNA binding site predictions; (B) network showing the 11 common validated TG, targeted by miR-1207-5p, miR-5581-3p and miR-611, respectively associated to circHIPK3 (the first) and circZNF609 (the last two); high-scoring predictions was obtained with miRanda score  $> 122$  and energy  $< -24$ , PITA  $\Delta G_{\text{duplex}} < -20$  and  $\Delta G_{\text{open}} > -11$ ; (C) heatmap of the RBP common to at least two circRNAs, considering the top 10% high-scoring RBP binding site predictions; (D) summary barplot of the number of ORF predicted for the group of circRNAs considered by CRAFT, binning ORF by the presence or absence of a stop codon in the reading frame. Canberra distance-based clustering has been used for heatmaps in (A) and (C).

is the low specificity of the predictions. To face this problem, we used a dual strategy: (i) the combination of two algorithms, miRanda [40, 83] and PITA [41], and (ii) the integration of predictions with experimentally determined AGO2 binding sites. The assessment of the CRAFT predictions on a sizable set of circRNAs with known functions showed that this approach can control the number of predictions, while holding good sensitivity.

Also the interaction of circRNAs with RBP can be associated with important functions [23]. In principle, circRNAs acting as decoys can counteract RBP functions, whereas in other cases circRNAs can serve as scaffolds for the assembly of molecular complexes. CRAFT allows

inspecting the possible interactions between circRNAs and RBP, and also the identification of the RBP possibly bound by multiple circRNAs. Of note, beyond the beRBP database of position weight matrices embedded in CRAFT, custom matrices can be provided by the user, extending the analysis.

CRAFT predicts the ORF in the circularized sequence, allowing the user to obtain the peptides potentially encoded by the circRNA and not by the linear transcript. To what extent circRNAs are translated it is still a matter of debate [84], but CEP have been identified in different studies and, in several cases, were also proven to be biologically active [85–87]. CRAFT aims

at the prediction of circRNA-specific peptides, thus it focuses on those ORF overlapping the backsplice site, which are not contained in linear transcripts. To undergo translation without a 5' cap, circRNAs can require specific sequence elements (internal ribosome entry sites (IRES) [88] and N<sup>6</sup>-methyladenosine modification [18, 89]) whose prediction can be envisaged as an interesting CRAFT future development.

The third CRAFT advantage for the user is that it provides a rich output with graphical visualizations, leveraging functional enrichments and results linking to several knowledge databases. The output can be explored, filtered and reanalyzed as long as the user prefers to generate hypotheses. For instance, the connection of circRNAs with miRNAs, validated miRNA-target genes, and specific pathways, functions and diseases can be of great help to identify regulatory networks to be further scrutinized. This can be pursued with experimental investigation, as well as integrating CRAFT results with circRNA and gene expression data. For instance, positive correlations between a circRNA and the validated target genes of miRNAs potentially sponged by the circRNA can be indicative of functional regulatory axes [4, 5].

Finally, we believe that the combination of predictions for different circRNAs provided by CRAFT is a highly useful and original feature. In research projects, when circRNAs dysregulated in cancer or disease tissue compared to a normal counterpart are identified, subsequent experimental investigation often focuses only on the elucidation of the functions of one specific circRNA. Instead, cooperative effects of multiple dysregulated circRNAs in pathogenetic mechanisms can be hypothesized. Thus, the potential interaction of different circRNAs with the same RBP can be inspected in CRAFT. Plus, CRAFT allows the exploration of regulatory networks in which multiple circRNAs can act as sponge for the same miRNA, reinforcing the desuppression of the miRNA-target genes, or sponge different miRNAs with a common target gene, possibly coregulating its expression.

In this study, we appraised CRAFT predictions by using, as benchmarking, a set of circRNAs with known functions, including the three types of functional elements predicted by CRAFT. CRAFT sensitivity was assessed in relation to the total number of predictions to be handled by the user at different settings. Our sensitivity estimates were highly conservative, based on a pejorative scenario in which all the functional elements are known and thus all the predictions not corresponding with known functional elements are counted as false positives, even if some of them can be functional still not known. The conducted tests informed the effects of different parameters, allowing us to optimize the CRAFT strategy and default settings and providing intelligence about the analysis customization and results filtering for CRAFT users.

In conclusion, we are handing over the researchers in the circRNA field with a new software tool facilitating the exploration of circRNA functions, and indicating potential regulatory axes involving one or more circRNAs

of interest. The generation of new hypotheses about the potential impact of circRNAs in biological processes, pathways and diseases can help circRNA prioritization for further study and the interpretation of their biological and pathogenetic role.

#### Key Points

- CRAFT is a self-standing tool for comprehensive circRNA function prediction.
- CRAFT functions include circRNA sequence reconstruction, microRNA and RNA-binding protein response elements and coding potential prediction.
- Predictions for multiple circRNAs are connected to infer possible cooperation networks and illustrate the potential impact of circRNAs on biological and disease processes.

## Supplementary data

Supplementary data are available online at <https://academic.oup.com/bib>.

## Data availability

CRAFT tool is available on GitHub at <https://github.com/annadalmolin/CRAFT> and on DockerHub at <https://hub.docker.com/repository/docker/annadalmolin/craft>.

## Authors' contributions

AD and SBo did the conceptualization; AD performed data curation; AD and SBo did the formal analysis; EG, SBr and SBo did the funding acquisition; AD and EG did the methodology; SBo did project administration; SBo collected the resources; AD and VD operated the software; EG and SBo did the supervision; AD and SBo did the visualization; AD, EG and SBo had written the original draft of the manuscript; Writing – Review & Editing, AD, EG, AB, CTP, SBr and SBo reviewed and edited the original draft of the manuscript. All authors read and approved the final manuscript.

## Funding

The Fondazione AIRC per la Ricerca sul Cancro, Milan, Italy [Investigator Grant 2017 20052 to SBo]; the Italian Ministry of Education, Universities, and Research [PRIN 2017 2017PPS2X4\_003 to SBo]; Fondazione Umberto Veronesi, Milan, Italy [Fellowship 2020 to EG]; Department of Molecular Medicine of The University of Padova [to SBo]; PhD in Biosciences of The University of Padova [to AB], Fondazione Cariparo [to SBr and SBo].

## References

1. Memczak S, Jens M, Elefsinioti A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 2013;**495**: 333–8.

2. Maass PG, Glažar P, Memczak S, et al. A map of human circular RNAs in clinically relevant tissues. *J Mol Med* 2017;**95**:1179–89.
3. Gaffo E, Boldrin E, Dal Molin A, et al. Circular RNA differential expression in blood cell populations and exploration of circRNA deregulation in pediatric acute lymphoblastic leukemia. *Sci Rep* 2019;**9**:14670.
4. Dal Molin A, Hofmans M, Gaffo E, et al. CircRNAs dysregulated in juvenile myelomonocytic leukemia: CircMCTP1 stands out. *Front Cell Dev Biol* 2020;**8**:613540.
5. Buratin A, Paganin M, Gaffo E, et al. Large-scale circular RNA deregulation in T-ALL: unlocking unique ectopic expression of molecular subtypes. *Blood Adv* 2020;**4**:5902–14.
6. Dal Molin A, Bresolin S, Gaffo E, et al. CircRNAs are here to stay: a perspective on the MLL recombinome. *Front Genet* 2019;**10**.
7. Dahl M, Daugaard I, Andersen MS, et al. Enzyme-free digital counting of endogenous circular RNA molecules in B-cell malignancies. *Lab Invest* 2018;**98**:1657–69.
8. Ng WL, Mohidin TBM, Shukla K. Functional role of circular RNAs in cancer development and progression. *RNA Biol* 2018;**1**:1–11.
9. Hansen TB, Jensen TI, Clausen BH, et al. Natural RNA circles function as efficient microRNA sponges. *Nature* 2013;**495**:384–8.
10. Rong D, Sun H, Li Z, et al. An emerging function of circRNA-miRNAs-mRNA axis in human diseases. *Oncotarget* 2017;**8**:73271–81.
11. Chen N, Zhao G, Yan X, et al. A novel FLI1 exonic circular RNA promotes metastasis in breast cancer by coordinately regulating TET1 and DNMT1. *Genome Biol* 2018;**19**:218.
12. Martin KC, Ephrussi A. mRNA localization: gene expression in the spatial dimension. *Cell* 2009;**136**:719–30.
13. Sonenberg N, Hinnebusch AG. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* 2009;**136**:731–45.
14. Fu X-D, Ares M, Jr. Context-dependent control of alternative splicing by RNA-binding proteins. *Nat Rev Genet* 2014;**15**:689–701.
15. Galarneau A, Richard S. Target RNA motif and target mRNAs of the quaking STAR protein. *Nat Struct Mol Biol* 2005;**12**:691–8.
16. Pamudurti NR, Bartok O, Jens M, et al. Translation of CircRNAs. *Mol Cell* 2017;**66**:9–21.e7.
17. Legnini I, Di Timoteo G, Rossi F, et al. Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis. *Mol Cell* 2017;**66**:22–37.e9.
18. Yang Y, Fan X, Mao M, et al. Extensive translation of circular RNAs driven by N6-methyladenosine. *Cell Res* 2017;**27**:626–41.
19. Abe N, Matsumoto K, Nishihara M, et al. Rolling circle translation of circular RNA in living human cells. *Sci Rep* 2015;**5**:16435.
20. Mo D, Li X, Raabe CA, et al. A universal approach to investigate circRNA protein coding function. *Sci Rep* 2019;**9**:11684.
21. Bonizzato A, Gaffo E, Te Kronnie G, et al. CircRNAs in hematopoiesis and hematological malignancies. *Blood Cancer J* 2016;**6**:e483.
22. Liu Z, Liu F, Wang F, et al. CircZNF609 promotes cell proliferation, migration, invasion, and glycolysis in nasopharyngeal carcinoma through regulating HRAS via miR-338-3p. *Mol Cell Biochem* 2021;**476**:175–86.
23. Du WW, Yang W, Liu E, et al. Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2. *Nucleic Acids Res* 2016;**44**:2846–58.
24. Xiang T, Jiang H-S, Zhang B-T, et al. CircFOXO3 functions as a molecular sponge for miR-143-3p to promote the progression of gastric carcinoma via upregulating USP44. *Gene* 2020;**753**:144798.
25. Glažar P, Papavasileiou P, Rajewsky N. circBase: a database for circular RNAs. *RNA* 2014;**20**:1666–70.
26. Yao D, Zhang L, Zheng M, et al. Circ2Disease: a manually curated database of experimentally validated circRNAs in human disease. *Sci Rep* 2018;**8**:11018.
27. Ghosal S, Das S, Sen R, et al. Circ2Traits: a comprehensive database for circular RNA potentially associated with disease and traits. *Front Genet* 2013;**4**:283.
28. Liu YC, Li JR, Sun CH, Andrews E, et al. CircNet: a database of circular RNAs derived from transcriptome sequencing data. *Nucleic Acids Res* 2016;**44**(Database issue):D209–D215.
29. Lyu Y, Caudron-Herger M, Diederichs S. circ2GO: a database linking circular RNAs to gene function. *Cancer* 2020;**12**.
30. Dudekula DB, Panda AC, Grammatikakis I, et al. CircInteractome: a web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biol* 2016;**13**:34–42.
31. Wu W, Ji P, Zhao F. CircAtlas: an integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes. *Genome Biol* 2020;**21**:101.
32. Li H, Xie M, Wang Y, et al. riboCIRC: a comprehensive database of translatable circRNAs. *Genome Biol* 2021;**22**:79.
33. Huang W, Ling Y, Zhang S, et al. TransCirc: an interactive database for translatable circular RNAs based on multi-omics evidence. *Nucleic Acids Res* 2021;**49**:D236–42.
34. Cardenas J, Balaji U, Gu J. Cerina: systematic circRNA functional annotation based on integrative analysis of ceRNA interactions. *Sci Rep* 2020;**10**:22165.
35. Lin Y-C, Lee Y-C, Chang K-L, et al. Analysis of common targets for circular RNAs. *BMC Bioinformatics* 2019;**20**.
36. Aufiero S, Reckman YJ, Tijssen AJ, et al. circRNAprofiler: an R-based computational framework for the downstream analysis of circular RNAs. *BMC Bioinformatics* 2020;**21**.
37. Sun P, Li G. CircCode: a powerful tool for identifying circRNA coding ability. *Front Genet* 2019;**10**:981.
38. Hossain MT, Peng Y, Feng S, et al. FcircSEC: an R package for full length circRNA sequence extraction and classification. *Int J Genomics* 2020;**2020**:1–11.
39. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;**26**:841–2.
40. Enright AJ, John B, Gaul U, et al. MicroRNA targets in drosophila. *Genome Biol* 2003;**5**:R1.
41. Kertesz M, Iovino N, Unnerstall U, et al. The role of site accessibility in microRNA target recognition. *Nat Genet* 2007;**39**:1278–84.
42. Blin K, Dieterich C, Wurmus R, et al. DoRiNA 2.0—upgrading the doRiNA database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res* 2015;**43**:D160–7.
43. Ru Y, Kechris KJ, Tabakoff B, et al. The multiMiR R package and database: integration of microRNA-target interactions along with their disease and drug associations. *Nucleic Acids Res* 2014;**42**:e133.
44. Yu H, Wang J, Sheng Q, et al. beRBP: binding estimation for human RNA-binding proteins. *Nucleic Acids Res* 2019;**47**:e26.
45. Yu G, Wang L-G, Yan G-R, et al. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* 2015;**31**:608–9.
46. Yu G, Wang L-G, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS: J Integrative Biol* 2012;**16**:284–7.
47. Yu G, He Q-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol Biosyst* 2016;**12**:477–9.
48. Yu G. Using meshes for MeSH term enrichment and semantic analyses. *Bioinformatics* 2018;**34**:3766–7.

49. Soudy M, Anwar AM, Ahmed EA, et al. UniprotR: retrieving and visualizing protein sequence and functional information from universal protein resource (UniProt knowledgebase). *J Proteomics* 2020;**213**:103613.
50. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*, 2009.
51. Wickham H, Averick M, Bryan J, et al. Welcome to the Tidyverse. *J Open Source Software* 2019;**4**:1686.
52. Fox J, Weisberg S. *An R Companion to Applied Regression*, 2018.
53. Yu G. *enrichplot: Visualization of Functional Enrichment Result. R package version 1.14.1*, 2021.
54. Xie F, Li Y, Wang M, et al. Circular RNA BCRC-3 suppresses bladder cancer proliferation through miR-182-5p/p27 axis. *Mol Cancer* 2018;**17**:144.
55. Zheng Q, Bao C, Guo W, et al. Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nat Commun* 2016;**7**:11215.
56. Jin Y, Che X, Qu X, et al. CircHIPK3 promotes metastasis of gastric cancer via miR-653-5p/miR-338-3p-NRP1 Axis under a long-term hypoxic microenvironment. *Front Oncol* 2020;**10**:1612.
57. Zheng F, Wang M, Li Y, et al. CircNR3C1 inhibits proliferation of bladder cancer cells by sponging miR-27a-3p and downregulating cyclin D1 expression. *Cancer Lett* 2019;**460**:139–51.
58. Zhang X, Wang S, Wang H, et al. Circular RNA circNRIP1 acts as a microRNA-149-5p sponge to promote gastric cancer progression via the AKT1/mTOR pathway. *Mol Cancer* 2019;**18**:20.
59. Cao L, Wang M, Dong Y, et al. Circular RNA circRNF20 promotes breast cancer tumorigenesis and Warburg effect through miR-487a/HIF-1 $\alpha$ /HK2. *Cell Death Dis* 2020;**11**:145.
60. Xie M, Yu T, Jing X, et al. Exosomal circSHKBP1 promotes gastric cancer progression via regulating the miR-582-3p/HUR/VEGF axis and suppressing HSP90 degradation. *Mol Cancer* 2020;**19**:112.
61. He Y, Huang H, Jin L, et al. CircZNF609 enhances hepatocellular carcinoma cell proliferation, metastasis, and stemness by activating the hedgehog pathway through the regulation of miR-15a-5p/15b-5p and GLI2 expressions. *Cell Death Dis* 2020;**11**:358.
62. Xiong Y, Zhang J, Song C. CircRNA ZNF609 functions as a competitive endogenous RNA to regulate FOXP4 expression by sponging miR-138-5p in renal carcinoma. *J Cell Physiol* 2019;**234**:10646–54.
63. Liang Y, Wang H, Chen B, et al. circDCUN1D4 suppresses tumor metastasis and glycolysis in lung adenocarcinoma by stabilizing TXNIP expression. *Mol Ther Nucleic Acids* 2021;**23**:355–68.
64. Chen R-X, Liu H-L, Yang L-L, et al. Circular RNA circRNA\_0000285 promotes cervical cancer development by regulating FUS. *Eur Rev Med Pharmacol Sci* 2019;**23**:8771–8.
65. Ashwal-Fluss R, Meyer M, Pamudurti NR, et al. circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell* 2014;**56**:55–66.
66. Chen R-X, Chen X, Xia L-P, et al. N6-methyladenosine modification of circNSUN2 facilitates cytoplasmic export and stabilizes HMG2 to promote colorectal liver metastasis. *Nat Commun* 2019;**10**.
67. Abdelmohsen K, Panda AC, Munk R, et al. Identification of HuR target circular RNAs uncovers suppression of PABPN1 translation by CircPABPN1. *RNA Biol* 2017;**14**:361–9.
68. Chen J, Wu Y, Luo X, et al. Circular RNA circRHOBTB3 represses metastasis by regulating the HuR-mediated mRNA stability of PTBP1 in colorectal cancer. *Theranostics* 2021;**11**:7507–26.
69. Barbagallo D, Caponnetto A, Cirnigliaro M, et al. CircS-MARCA5 inhibits migration of glioblastoma multiforme cells by regulating a Molecular Axis involving splicing factors SRSF1/SRSF3/PTB. *Int J Mol Sci* 2018;**19**.
70. Feng Y, Yang Y, Zhao X, et al. Circular RNA circ0005276 promotes the proliferation and migration of prostate cancer cells by interacting with FUS to transcriptionally activate XIAP. *Cell Death Dis* 2019;**10**:792.
71. Zhu Y-J, Zheng B, Luo G-J, et al. Circular RNAs negatively regulate cancer stem cells by physically binding FMRP against CCAR1 complex in hepatocellular carcinoma. *Theranostics* 2019;**9**:3526–40.
72. Gao X, Xia X, Li F, et al. Circular RNA-encoded oncogenic E-cadherin variant promotes glioblastoma tumorigenicity through activation of EGFR-STAT3 signalling. *Nat Cell Biol* 2021;**23**:278–91.
73. Pan Z, Cai J, Lin J, et al. A novel protein encoded by circFND3B inhibits tumor progression and EMT through regulating snail in colon cancer. *Mol Cancer* 2020;**19**:71.
74. Liang Z-X, Liu H-S, Xiong L, et al. A novel NF- $\kappa$ B regulator encoded by circPLCE1 inhibits colorectal carcinoma progression by promoting RPS3 ubiquitin-dependent degradation. *Mol Cancer* 2021;**20**:103.
75. Zhang M, Huang N, Yang X, et al. A novel protein encoded by the circular form of the SHPRH gene suppresses glioma tumorigenesis. *Oncogene* 2018;**37**:1805–14.
76. Wu X, Xiao S, Zhang M, et al. A novel protein encoded by circular SMO RNA is essential for Hedgehog signaling activation and glioblastoma tumorigenicity. *Genome Biol* 2021;**22**:33.
77. Mout A. *Using Docker: Developing and Deploying Software with Containers*, 2015.
78. Wang J, Zhang Y, Song H, et al. The circular RNA circSPARC enhances the migration and proliferation of colorectal cancer by regulating the JAK/STAT pathway. *Mol Cancer* 2021;**20**:81.
79. Kristensen LS, Hansen TB, Venø MT, et al. Circular RNAs in cancer: opportunities and challenges in the field. *Oncogene* 2018;**37**:555–65.
80. Gaffo E, Bonizzato A, Kronnie GT, et al. CirComPara: a multi-method comparative bioinformatics pipeline to detect and study circRNAs from RNA-seq data. *Noncoding RNA* 2017;**3**.
81. Gaffo E, Buratin A, Dal Molin A, et al. Sensitive, reliable, and robust circRNA detection from RNA-seq with CirComPara2. *Brief Bioinform* 2022 Jan 17;**23**(1):bbab418.
82. Chen L-L, Yang L. Regulation of circRNA biogenesis. *RNA Biol* 2015;**12**:381–8.
83. John B, Enright AJ, Aravin A, et al. Human MicroRNA targets. *PLoS Biol* 2004;**2**:e363.
84. Ho-Xuan H, Glažar P, Latini C, et al. Comprehensive analysis of translation from overexpressed circular RNAs reveals pervasive translation from linear transcripts. *Nucleic Acids Res* 2020;**48**:10368–82.
85. Zhang M, Zhao K, Xu X, et al. A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma. *Nat Commun* 2018;**9**:4475.
86. Ye F, Gao G, Zou Y, et al. circFBXW7 inhibits malignant progression by sponging miR-197-3p and encoding a 185-aa protein in triple-negative breast cancer. *Mol Ther Nucleic Acids* 2019;**18**:88–98.
87. Lu Y, Li Z, Lin C, et al. Translation role of circRNAs in cancers. *J Clin Lab Anal* 2021;**35**:e23866.
88. Chen C-K, Cheng R, Demeter J, et al. Structured elements drive extensive circular RNA translation. *Mol Cell* 2021.
89. Zhang L, Hou C, Chen C, et al. The role of N6-methyladenosine (m6A) modification in the regulation of circRNAs. *Mol Cancer* 2020;**19**.