# **Investigation of RNA-RNA Interactions Using the RISE Database**

Yanyan Ju,<sup>1,4</sup> Jing Gong,<sup>1,4</sup> Yucheng T. Yang,<sup>2,3,5</sup> and Qiangfeng Cliff Zhang<sup>1,5</sup>

<sup>1</sup>MOE Key Laboratory of Bioinformatics, Beijing Advanced Innovation Center for Structural Biology, Center for Synthetic and Systems Biology, Tsinghua-Peking Joint Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing, China <sup>2</sup>Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut

<sup>3</sup>Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut

<sup>4</sup>These authors contributed equally to the work

<sup>5</sup>Corresponding authors: *qczhang@tsinghua.edu.cn*; *yucheng.yang@yale.edu* 

RNA-RNA interactions (RRIs) are essential to understanding the regulatory mechanisms of RNAs. Mapping RRIs *in vivo* in a transcriptome-wide manner remained challenging until the recent development of several sequencing-based technologies. However, RRIs generated from large-scale studies had not been systematically collected and analyzed before. This article introduces RISE, a database of the RNA Interactome from Sequencing Experiments. RISE provides a comprehensive collection of RRIs in human, mouse, and yeast, derived from transcriptome-wide sequencing experiments, as well as targeted sequencing studies and other public databases/datasets. To facilitate better understanding of the biological roles of these RRIs, RISE also offers rich functional annotations involving RNAs, and an interactive interface to explore the analysis results. Here, we provide a brief description of the RISE website and a step-by-step protocol for using RISE to study RRIs. © 2018 by John Wiley & Sons, Inc.

Keywords: interaction • RISE • RNA • RNA-RNA • RNA structure • sequencing experiments

## How to cite this article:

Ju, Y., Gong, J., Yang, Y. T., & Zhang, Q. C. (2018). Investigation of RNA-RNA interactions using the RISE database. *Current Protocols* in Bioinformatics, 64, e58. doi: 10.1002/cpbi.58

## INTRODUCTION

RNA molecules can fold into intrinsic complex structures by forming intramolecular and intermolecular base pairs. RNA-RNA interactions (RRIs) induced by intermolecular base pairing play essential roles in gene expression. A canonical example is the base pairing formed between the codons of mRNAs and the anticodons of tRNAs during protein synthesis (Ibba & Soll, 2000). Another example is the recognition of the intronic regions of pre-mRNAs by spliceosomal small nuclear RNAs (snRNAs; Guil & Esteller, 2015). In addition, RRIs are the basis of many post-transcriptional regulatory processes. For example, the 3' UTRs of mRNAs, and lncRNAs as well, can be targeted by miRNAs (Bartel, 2004), which often results in a complex regulatory competing endogenous RNA (ceRNA) network when different sites are targeted by the same set of miRNAs (Salmena, Poliseno, Tay, Kats, & Pandolfi, 2011).





Although RRIs can be formed between all types of RNAs, previous studies have focused on limited RNA species, mainly mRNA-miRNA interactions. Only recently have technologies emerged to characterize comprehensive sets of RRIs involving other types of RNAs. These high-throughput methods can be mainly divided into two groups, i.e., (i) detecting RRIs for a target RNA/protein and (ii) detecting transcriptome-wide RRIs in a cell. The first group includes CLASH (Helwak, Kudla, Dudnakova, & Tollervey, 2013; Kudla, Granneman, Hahn, Beggs, & Tollervey, 2011), hiCLIP (Sugimoto et al., 2015), RIL-seq (Melamed et al., 2016), RIA-seq (Kretz et al., 2013), and RAP-RNA (Engreitz et al., 2014). Briefly, CLASH, hiCLIP, and RIL-seq first crosslink RNA duplexes with associated proteins by UV irradiation. The crosslinked ribonucleoprotein complexes are then affinity purified, and the ligated RNA sequences are subjected to high-throughput sequencing. A modified CLASH method additionally incorporates a small molecule crosslinker [i.e., 4'-aminomethyltrioxsalen (AMT)] to probe direct base pairing simultaneously (Liu et al., 2017). RIA-seq and RAP-RNA use antisense DNA probes of the target RNA to capture RNA duplexes, followed by high-throughput sequencing. The second group includes PARIS (Lu et al., 2016), SPLASH (Aw et al., 2016), LIGR-seq (Sharma, Sterne-Weiler, O'Hanlon, & Blencowe, 2016), and MARIO (Nguyen et al., 2016). The basic strategy of these transcriptome-wide methods is to first crosslink interacting RNA duplexes using nucleic acid crosslinker (e.g., AMT) treatment and UV irradiation, and then purify and capture the enriched RNA duplexes. The RRI regions can be identified by high-throughput sequencing of the ligated duplexes and downstream bioinformatics analysis.

We have recently developed RISE, a database of RNA Interactome from Sequencing Experiments, to aggregate and annotate RRIs derived from these high-throughput sequencing experiments together with those from other databases and publications (Gong et al., 2018). Currently, RISE contains 328,811 RRIs from human, mouse and yeast.

In this article, we will first introduce the basic structure of the RISE database. Then, we will show how to explore RISE by using an example target gene to search for its associated RRIs (Basic Protocol 1). Finally, we will demonstrate how the functional annotations of these RRIs can be useful in understanding their regulatory functions (Basic Protocol 2).

## STRATEGIC PLANNING

RISE can be accessed at *http://rise.zhanglab.net*. The homepage includes a welcome description, a search box, and a navigation bar. The navigation bar is on the left of the homepage, with hyperlinks to other pages (Fig. 1). Users can click the bar to go to any other section and return to the homepage by clicking "Search". The front page contains seven sections, including "News," "Search," "Statistics," "Downloads," "Help", "Links," and "Contact.". Here, we briefly describe five of these sections.

- (1) The "News" module provides information about updates to the RISE database. Information on additional data or new functional features of RISE can be found here.
- (2) The "Statistics" module provides information on the data source and simple statistics for all collected RRIs in both table view and graph view. RRIs are grouped into three categories: (i) transcriptome-wide sequencing studies, (ii) targeted sequencing studies, and (iii) other databases/datasets. The numbers of RRIs and involved genes in each group are shown in a table view (Fig. 2). More statistics are visualized using several graphs. The pie charts show the number of RRIs from different species, cell lines, and sources/studies (Fig. 3). The Circos plots (Krzywinski et al., 2009) show RRI distribution within and between different types of RNAs in human and mouse (Fig. 3).

🛱 News	Navigation column
🗟 Search	
() Statistics	RISE: RNA Interactome from Sequencing Experiments
Ł Downloads	RBFOX2 Search RNA
🍄 Help	Searching area
Links	O Human O Mouse Veast
Contact	RISE is a comprehensive repository of RNA-RNA interactions. The interactions mainly come from recent transcriptome-wide sequencing-based experiments like PARIS, SPLASH, LIGRseq, and MARIO, and targeted studies like RIAseq, RAP-RNA, and CLASH. RISE also includes interactions aggregated from other (primary) databases and publications. The RISE database currently contains 328,811 RNA-RNA interactions mainly in three species. It features 292,850 Interactions involving messenger and long non-coding RNAs. Introduction of RISE
Contact	RISE is a comprehensive repository of RNA-RNA interactions. The interactions mainly come from recent transcriptome-wide sequencing-based experiments like PARIS, SPLASH, LIGRseq, and MARIO, and targeted studies like RIAseq, RAP-RNA, and CLASH. RISE also includes interactions aggregated from other (primary) databases and publications. The RISE database currently contains 328,811 RNA-RNA interactions mainly in three species. It features 292,850 interactions involving messenger and long non-coding RNAs. Citation
Contact	RISE is a comprehensive repository of RNA-RNA interactions. The interactions mainly come from recent transcriptome-wide sequencing-based experiments like PARIS, SPLASH, LIGRseq, and MARIO, and targeted studies like RIAseq, RAP-RNA, and CLASH. RISE also includes interactions aggregated from other (primary) databases and publications. The RISE database currently contains 328,811 RNA-RNA interactions mainly in three species. It features 292,850 Interactions involving messenger and long non-coding RNAs. Citation
Contact	RISE is a comprehensive repository of RNA-RNA interactions. The interactions mainly come from recent transcriptome-wide sequencing-based experiments like PARIS, SPLASH, LIGRseq, and MARIO, and targeted studies like RIAseq, RAP-RNA, and CLASH. RISE also includes interactions aggregated from other (primary) databases and publications. The RISE database currently contains 328,811 RNA-RNA interactions mainly in three species. It features 292,850 interactions involving messenger and long non-coding RNAs. Introduction of RISE Citation RISE: a database of RNA Interactome from Sequencing Experiments. Jing Gong <sup>*</sup> , Di Shao <sup>*</sup> , Kui Xu, Zhipeng Lu, Zhi John Lu, Yucheng T, Yang <sup>*</sup> , Qiangfeng Cliff Zhang <sup>*</sup>
Contact	RISE is a comprehensive repository of RNA-RNA interactions. The interactions mainly come from recent transcriptome-wide sequencing-based experiments like PARIS, SPLASH, LIGRseq, and MARIO, and targeted studies like RIAseq, RAP-RNA, and CLASH. RISE also includes interactions aggregated from other (primary) databases and publications. The RISE database currently contains 328,811 RNA-RNA interactions mainly in three species. It features 292,850 interactions involving messenger and long non-coding RNAs. Introduction of RISE Citation RISE: a database of RNA Interactome from Sequencing Experiments. Jing Gong <sup>*</sup> , Di Shao <sup>*</sup> , Kui Xu, Zhipeng Lu, Zhi John Lu, Yucheng T, Yang <sup>*</sup> , Qiangfeng Cliff Zhang <sup>*</sup> Citation file
Contact	RISE is a comprehensive repository of RNA-RNA interactions. The interactions mainly come from recent transcriptome-wide sequencing-based experiments like PARIS, SPLASH, LIGRseq, and MARIO, and targeted studies like RIAseq, RAP-RNA, and CLASH. RISE also includes interactions aggregated from other (primary) databases and publications. The RISE database currently contains 328,811 RNA-RNA interactions mainly in three species. It features 292,850 interactions involving messenger and long non-coding RNAs. Introduction of RISE Citation RISE: a database of RNA Interactome from Sequencing Experiments. Jing Gong <sup>*</sup> , Di Shao <sup>*</sup> , Kui Xu, Zhipeng Lu, Zhi John Lu, Yucheng T, Yang <sup>*</sup> , Qiangfeng Cliff Zhang <sup>*</sup> Citation file
Contact	RISE is a comprehensive repository of RNA-RNA interactions. The interactions mainly come from recent transcriptome-wide sequencing-based experiments like PARIS, SPLASH, LIGRseq, and MARIO, and targeted studies like RIAseq, RAP-RNA, and CLASH. RISE also includes interactions aggregated from other (primary) databases and publications. The RISE database currently contains 328,811 RNA-RNA interactions mainly in three species. It features 292,850 interactions involving messenger and long non-coding RNAs. Introduction of RISE Citation RISE: a database of RNA Interactome from Sequencing Experiments. Jing Gong*, Di Shao*, Kui Xu, Zhipeng Lu, Zhi John Lu, Yucheng T. Yang*, Qiangfeng Cliff Zhang* Citation file News *Publication on 4th October 2017 on NAR database issue.

**Figure 1** The homepage of the RISE database. See Strategic Planning for detailed description of the navigation menu, and Basic Protocol 1 for how to search RISE with the search box.

- (3) The "Downloads" module enables users to acquire all the data in RISE. The detailed explanation of the file format can also be found. All RRI data are grouped into different subsets according to species and source. The number of entries and the size of the file for each category are shown in a table. By clicking on a table item, users can download the chosen RRIs. The download files are provided in the BEDPE file format, defined in the BEDTOOLS suite (Quinlan & Hall, 2010) for connecting two relevant genomic regions. A hyperlink to the detailed description of the BEDPE format is provided.
- (4) The "Help" module provides a detailed guidance to RISE, including an introduction, a usage example demonstrated with RNA binding fox-1 homolog 2 (RBFOX2), frequently asked questions, and related references as well. Users are advised to briefly go through this section before exploring the RISE database.
- (5) The "Links" module provides hyperlinks to all additional resources in RISE, including databases used for RRI curations, different resources for RRI annotation, and other related RNA databases/datasets.

## SEARCHING RNA-RNA INTERACTIONS

Users can query the RRIs of a specific gene in the search box on the front page. The result page contains several modules for further exploration, including a summary table of query gene, visualization of all RRIs using a Circos plot, and a table view for detailed information. Here, we use an example, RBFOX2, for a step-by-step demonstration of the RRI search.

BASIC PROTOCOL 1

News	Data sources							
🗟 Search	DNA DNA inter	actions (DDIs) in DISE are from bi	h throughout conuc	nging bacad	owneriments and other	sublic suslishie (priman)	databases and datasets. The f	alloudeg tab
Statistics	gives the detaile them, transcript	d source information. Briefly, RIS ome-wide studies use PARIS, MA	E curated RRIs from RIOS, SPLASH, LIGR	3 types of s -seq techno	ourcestranscriptome logies to discover RRIs,	e-wide studies, targeted st and targeted studies use l	tudies, and other datasets. The RIA-seq, RAP-RNA, CLASH. Th	tasets. Amo
Ł Downloads	databases/datas	ets" include RRIs from the NPInt cieshuman, mouse, yeast, E.coli	er v3.0, RAID v2.0, R i, S.enteria, or 10 diffe	AIN databas erent cell line	ses, and the Lai D, et al. 2 es.	016 datasets (published	in PMID: 26673718). Current	ly, RRIS IN R
다 Help		Category	Method/Resource	Species	Cell line	Number of interactions	Number of involved genes	
				Human	HEK293T	25,824	16,192	
			PARIS	Human	Hela (high RNase)	25,552	19,335	
Contact				Human	Hela (low RNase)	20,330	17,009	
	i	Transcriptome-wide studies		Mouse	mESC	29,514	12,625	
				Mouse	MEF	7,167	2,936	
			MARIO	Mouse	mESC <sup>a</sup>	99,290	15,309	
		number prome while studies		Mouse	mESC <sup>b</sup>	37,441	9,715	
				Human	hESC	3,345	971	
				Human	HeLa	5,799	1,649	
			SPLASH	Human	LCL	4,213	429	
				Human	RA	1,770	671	
			LIGR-seq	Human	HEK293T	641	749	
			RIA-seq	Human	Keratinocytes (TINCR)	3,609	1,815	
				Mouse	mESC (MALAT1)	495	489	
		Targeted studies	KAP-KNA	Mouse	mESC (U1 snRNA)	12,278	8,635	
				Human	HEK293 (miRNAs)	18,508	7,260	
			CLASH	Yeast	BY4741 (miRNAs)	253	47	
				Human		3,691	2,525	
			NPInter v3.0	Mouse		52	83	
			2002 022	Human	¥.	22,521	6,262	
			RAID v2.0	Mouse	<i>*</i>	3,440	2,130	
		From other databases/datasets		Human		2,881	1,189	
			RAIN	Mouse		36	31	
				E. coli		64	68	
			PMID 26673718	S. enterica	a	45	49	
				Yeast		52	45	
		Total				220.044	54 205	

**Figure 2** The statistics table view in the "Statistics" page of RISE. The table provides a summary of the curated RRI data, including the number of RRIs and involved genes from different categories, methods, species, and cell lines.

## Necessary Resources

## Hardware

A computer with internet access

Software

Any up-to-date web browser (e.g., Firefox, Chrome, Opera, Safari, Internet Explorer)

Files

None

## Input gene name and select a species

- 1. First, start the web browser and connect to the website *http://rise.zhanglab.net*. The homepage of the RISE database consists of a brief introduction to RISE, a search box, and a navigation bar with links to other functional modules (Fig. 1). Users can search RRIs of a specific gene on this page with the search box. They can also look for additional information in the RISE database by using the navigation bar.
- 2. Enter the name of the query gene (i.e., RBFOX2) in the search box. The server can automatically complete the full gene name according to the first few letters entered, and provide some candidate gene names in the drop-down box, in which the users



**Figure 3** The statistics plots in the "Statistics" page of RISE. The pie plots (upper) show the distribution of RISE's RRI data in different species, cell lines, and sources. The Circos plots (bottom) show the number of RRIs within/between different types of RNAs.

can select the exact gene name they want to search (Fig. 4). Note that gene-name matching is case-insensitive.

3. After entering the gene name, users need to select a species of interest from human, mouse, and yeast. Note that "Human" is the default selection (Fig. 4). Then, click the "Search RNA" button on the right of the search box. The page will be refreshed to the result layout.

#### Gene summary

The first part of the search results page is an information table about the query gene (Fig. 5). The table lists some basic information, including gene name, gene aliases, species, Ensembl gene ID, genomic location of the RNA, exon count, gene type, RRI number (the total number of RRIs involving this gene), and RefSeq summary. The information on the gene annotation is obtained from Ensembl v84. In the case where multiple isoforms exist, the longest transcript of the gene will be used (Zerbino et al., 2018). The RefSeq summary provides an overall description about the gene and its biological functions (O'Leary et al., 2016).

RBFOX2				Search RNA
1. Input RNA nar	ne			
2. Select a speci	es → 💿 Human	O Mouse	O Yeast	
	Ū			
RBFOX1	~			Search RNA
RBFOX1				
RBFOX2	O Human	O Mouse	O Yeast	
BBEOV/2				

**Figure 4** The search of RRIs for a specific RNA in RISE. Users need to enter a query RNA name and select the species ("Human" is the default species).

Gene name: RBFOX2	Aliases: FOX2, Fox-2, HNRBP2, HRNBP2, RBM9, RTA, dJ106I20.3, fxh
Species: human	Ensembl ID: ENSG00000100320
RNA location: chr22:35738736-35840587,-	Exon count: 12
Gene type: protein_coding	RRI number: 29
<b>RefSeq summary:</b> This gene is one of several hubinding protein that is thought to be a key regula The protein binds to a conserved UGCAUG elem inclusion of the alternative exon in mature transc transcription factor and regulates estrogen rece	uman genes similar to the C. elegans gene Fox-1. This gene encodes an RNA itor of alternative exon splicing in the nervous system and other cell types. ient found downstream of many alternatively spliced exons and promotes cripts. The protein also interacts with the estrogen receptor 1 ptor 1 transcriptional activity. Multiple transcript variants encoding [provided by PofSeq. Jul 2009]

**Figure 5** The gene summary section in a result page. The table provides the basic information of the query gene RBFOX2.

## **Overview of RRIs by Circos plot**

This section provides an integrative visualization of the query gene using a Circos plot, which has been widely used for genomics visualizations (Krzywinski et al., 2009). Here, we show the RRIs between RBFOX2 and its partners (Fig. 6). When exploring the results, users can always adjust their view using the buttons on the upper left. Users can zoom in/out and move the plot by clicking the  $\pm$  buttons and the four arrows for four different directions, respectively. The plot can be reset to its original state by clicking the refresh button. The Circos plot contains six functional annotation tracks to help users explore the genomic context of the interacting regions, and study the potential regulatory roles of the RRIs. Annotations are encoded using different colors in each track. Users can hold the mouse on the label in the track, and a floating hint box with detailed explanation of the label will appear (Fig. 7). Here, we give a brief introduction on each track using this example:

Ju et al.

4. *Gene name*: In this track, the query gene is in bold black, and labeled with asterisks (i.e., \*\*\*RBFOX2\*\*\*). Different types of genes (e.g., protein-coding, lncRNA,



**Figure 6** Interactive visualization of RRIs and annotations using a Circos plot. From the outside to the inside, the six tracks are: (i) gene name, (ii) gene structure, (iii) RBP binding sites, (iv) SNPs/pan-cancer mutations, (v) RNA modification/editing sites, and (vi) RNA-RNA interactions. The track legends list different types of available elements.



Figure 7 Exemplified floating box to describe the feature label under each track.

Gene structure		N				
Genomic element	Start coordinate	End coordinate	Length (bp)	Start coordinate	Length (bp, log <sub>2</sub> scaled)	End coordinate
5' UTR	112,861,222	112,861,376	153	112,861,222	8	112,861,229
CDS 1	112,861,377	112,861,417	41	112,861,230	6	112,861,235
Intron 1	112,861,418	112,862,507	1,090	112,861,236	11	112,861,246
CDS 2	112,862,508	112,862,583	76	112,861,247	7	112,861,253
Intron 2	112,862,584	112,864,456	1,873	112,861,254	11	112,861,264
CDS 3	112,864,457	112,864,528	72	112,861,265	7	112,861,271
Intron 3	112,864,529	112,864,620	92	112,861,272	7	112,861,278
CDS 4	112,864,621	112,864,732	112	112,861,279	7	112,861,285
Intron 4	112,864,733	112,867,403	2,671	112,861,286	12	112,861,297
CDS 5	112,867,404	112,867,534	131	112,861,298	8	112,861,305
3' UTR	112,867,535	112,869,788	2,254	112,861,306	12	112,861,317

**Figure 8** The conversion of the width of different genomic elements from raw coordinates to log<sub>2</sub> scaled for a transcript of SRP19 (ENSG00000153037).

ncRNA, snoRNA, miRNA, rRNA, snRNA, tRNA, pseudogene, transposable element, etc.) are assigned with different colors.

- 5. *Gene structure*: This track depicts the details of the gene structure of the involved RNA using its annotation from Ensembl. For genes with multiple isoforms, the longest transcript is displayed here. Each gene structure element is shown as a block with different colors and varied width. The length of transcripts depicted here is on a log<sub>2</sub> scale. To explain how we did length scaling, we provide an example showing a transcript of SRP19, which contains five exons and four introns (Fig. 8).
- 6. *RBP (RNA binding protein) binding*: This track provides RBP binding information of the interaction regions. The RBP binding sites located in any interacting regions will be shown with a yellow segment. In this example, there are two RBP binding sites located in SRD5A1 and PCDH7, respectively.
- 7. *SNP/mutation site*: This track displays SNPs and Pan-cancer mutations in green and blue segments. Each segment represents a SNP/mutation site overlapping with interacting regions.
- 8. *RNA editing/modification site*: This track contains RNA editing/modification sites within interacting regions, classified into five different types: RNA editing, m<sup>6</sup>A, m<sup>5</sup>C, pseudoU, and others. Each type is represented by segment with different colors. In this example, there is a m<sup>6</sup>A modification site in an interacting region of RBFOX2.

Search:	result	table 100 show 10 \$	Dow entries	nload resul	t table	miRNA ncRNA protein_coo pseudogen	ding e	HEK293 HEK293T Hela(highRNa Hela(lowRNas	PAF PAF RAII se)	NS Dv2.0
Interacting region in RBFOX2	Genomic context	Ensembl ID of Interacting gene	Gene name	Gene type	Interacting region	Genomic context	Species	Cell line	() Method/Source	PMID
chr22:35809944- 35809990	CDS	ENSG00000112655	РТК7	protein_coding	chr6:43161525- 43161567	CDS	human	Hela(highRNase)	PARIS	2718090
chr22:35759936+ 35759961	CDS	ENSG00000225630	MTND2P28	pseudogene	chr1:630610- 630642	Exon	human	HEK293T	PARIS	271809
chr22:35744164- 35744204	CDS	ENSG00000204345	CD300LD	protein_coding	chr17:74588663- 74588689	CDS	human	Hela(lowRNase)	PARIS	271809
chr22:35809944- 35809988	CDS	ENSG00000112655	РТК7	protein_coding	chr6:43161525- 43161567	CDS	human	Hela(lowRNase)	PARIS	271809
chr22:35809797- 35809850	CDS	ENSG00000207721	MIR186	miRNA	chr1:71067694- 71067716	Exon	human	HEK293	CLASH	236222
chr22:35746497- 35746550	CDS	EN5G00000197182	MIRLET7BHG	IncRNA	chr22:46112731- 46112753	Exon	human	HEK293	CLASH	236222
chr22:35746492- 35746540	CDS	ENSG00000199165	MIRLET7A1	miRNA	chr9:94175951- 94175973	Exon	human	HEK293	CLASH	236222
chr22:35744217- 35744235	CDS	ENSG00000169851	PCDH7	protein_coding	chr4:30722514- 30722536	CDS	human	Hela(highRNase)	PARIS	271809
chr22:35840561- 35840580	SUTR	ENSG00000261020	RP11- 744K17	IncRNA	chr17:22423406- 22423423	Exon	human	HEK293T	PARIS	271809
chr22:35740046- 35740082	<b>3UTR</b>	ENSG00000145545	SRD5A1	protein_coding	chr5:6668720- 6668754	<b>3UTR</b>	human	Hela(highRNase)	PARIS	271809
Showing 1 to 10 of	29 entries								Previous 1 2	3 Ne

**Figure 9** The integrated table view of the RRIs for a query RNA, RBFOX2. The table provides regional details of the query RNA and the interacting RNA partner. See text and next few figures for details.

- 9. *RNA-RNA interaction*: This track illustrates the links connecting the interacting regions of the query gene and its interaction partners. The colors of the links are assigned according to their sources, i.e., transcriptome-wide sequencing studies (red), targeted sequencing studies (blue), and other databases or datasets (green).
- 10. For all these six tracks, find a detailed description of the annotations by holding the mouse on the blocks, segments, or interaction links (Fig. 7). Taking the interacting region as an example, a floating window will appear when putting the mouse cursor over the link. The information is given in the following order: RNA name, Ensembl gene ID, gene type, and duplex region of two interacting RNAs, together with experimental method, species, and cell line.

## RRI table

This section provides details for the query gene and all RRIs involving this gene (Fig. 9). Above the table is a search box with a selection box to set the number of entries shown on each page, and a data export button.

- 11. Use the search box to filter the results. For example, to check whether another RNA (e.g., TP53 and MIRLET7A1) has specific interactions with the query RNA (i.e., RBFOX2), the user can input the target in the search box (Fig. 10). Searching with more than one term is also supported, in which case users need to type multiple terms separated by spaces, thus keeping entries that meet all query terms. The table will be refreshed in real time, automatically changing the displayed entries. If there are a large number of queries, users can adjust the number of entries in the selection box to view more in one page.
- 12. Download the result table by clicking "Export data to CSV file" for more downstream analysis (Fig. 10).

Search	n:		S	how 10	¢e	entries	Export dat	ta to CSV fi	le				
NA-RNA interac	tions												
Search: TP53	TP53	Show 10 \$ en	tries	xport data to C	5V file								
Interacting region	Genomic	Ensembl ID of	Gene	Genet	/pe	Interacting	Genomic	Species	Cellline	÷	Method/Se CLASH	ource	PMID
in RBFOX2	context	interacting gene	name		÷	region	context				) (		0.000000
in RBFOX2	context	interacting gene	name		\$ No mat	tching records found	context				/		
in RBFOX2 Showing 0 to 0 of 0	context	interacting gene	name		\$ No mat	tching records found	concar					Previou	s Nex
In RBFOX2 Showing 0 to 0 of 0 tNA-RNA interac Search: Search: Interacting region in RBFOX2	context entries (filtered to tions LET7, Genomic context	Interacting gene from 29 total entries)	tries I Gene name	xport data to C Gene typ	SV file	Interacting region	Genomic context	Species	Cell line	\$	Method/Sour CLASH \$	Previou:	s Nex



This RRI table (Fig. 9) contains eleven columns shown in three different background colors: (i) the first two columns (with red background) denote the query RNA information; (ii) the third to seventh columns (with blue background) denote the RNA information of interacting partners; and (iii) eighth to the last columns (with white background) denote other information. The first part includes the interacting regions of the query gene (i.e., RBFOX2) and its genomic context. The second part indicates the Ensembl gene ID of the interacting partner gene, the gene name, the gene type, the interacting region, and genomic context. The third part includes species, cell line, experimental method/source, and PMID. Among these eleven columns, "Gene type," "Cell line" (note that the RRIs from other databases do not have cell line information), and "Method/source" have their own selection buttons.

- 13. Filter the search results by clicking on these buttons, and choose one from the floating selection box.
- 14. Finally, click a gene in the "Gene name" column to be directed to the result page of this gene. In addition, users can access the reference paper in PubMed by clicking the PMID in the "PMID" column.

# BASIC EXPLORING

EXPLORING FUNCTIONAL ANNOTATIONS

To facilitate functional exploration of interacting regions, RISE provides multiple types of molecular annotations: (i) RBP binding, (ii) RNA editing, (iii) RNA modification, (iv) SNP, (v) pan-cancer mutation, and (vi) gene expression level.

## Necessary Resources

Hardware

A computer with internet access

Ju et al.

10 of 15

RBP bindin	g RNA edi	ting/modification	SNPs	/Pan-Cancer n	nutations	Gene express	ion levels		
Your target	t RNA has no RBP bir	idng search results	in the RRI inter	ractome.					
NA annotat	ion								
RBP bindin	RNA ed	iting/modification	SNP	s/Pan-Cancer I	nutations	Gene expres	sion levels		
Search:		Show 10 \$	entries Exp	ort data to CSV file	]				
RBP name	RBP position	Genomic context	Binding score	Method	Analysis method	Species	Cell line	Source	PMID
ELAVL1	chrX:73833332- 73833353	Exon	0.700	PAR-CLIP	PARalyzer	human	HEK293	CLIPdb	24393468
ELAVL1	chrX:73833356- 73833382	Intron	0.940	PAR-CLIP	PARalyzer	human	HEK293	CLIPdb	24393468
ELAVL1	chrX:73831232- 73831246	Exon	0.649	PAR-CLIP	PARalyzer	human	HEK293	CLIPdb	21572407
ELAVL1	chrX:73831223- 73831249	Exon	0.791	PAR-CLIP	PARalyzer	human	HEK293	CLIPdb	24393468
HNRNPC	chrX:73833355- 73833376	Intron	9.000	iCLIP	CIMS	human	HeLa	CLIPdb	23374342
HNRNPC	chrX:73831224- 73831245	Exon	3.000	iCLIP	CIMS	human	HeLa	CLIPdb	23374342
PTBP1	chrX:73833359-	Exon	8.000	iCLIP	CIMS	human	HEK293T	CLIPdb	25219497

Figure 11 The "RBP binding" annotation sub-module for two query RNAs, RBFOX2 (top) and XIST (bottom).

#### Software

Any up-to-date web browser (e.g., Firefox, Chrome, Opera, Safari, Internet Explorer)

Files

None

#### Functional annotations of RRIs

- 1. Use the "RBP binding" module to identify RBP binding sites located in the interacting regions (Fig. 11). In the case of RBFOX2, there are no RBP binding sites in the interacting regions. Here, we demonstrate the usage of this sub-module with another transcript, XIST, with interacting regions covering several RBP binding sites. From left to right, the table shows the name of the RBP, the position of the RBP binding region on the gene, the genomic context of the binding site, the binding score calculated in CLIPdb (Yang et al., 2015), the CLIP-seq experimental method to detect the binding site, the computational method for binding site identification, the species, the cell line, the source database (i.e., CLIPdb), and the PMID of the reference that describes the experiment.
- 2. Use the "RNA editing" sub-module, which shows RNA editing sites located in the RRI regions (Fig. 12). Again there is no RNA editing site in its RRI regions for RBFOX2. So we use another example, TP53, whose interacting regions cover three RNA editing sites. There are five columns in the RNA editing table, including the type of RNA editing (i.e., the change of base before and after editing), the editing

RBP binding	RNA editing/modification	SNPs/Pan-Cancer mutation	s Gene expression le	evels
RNA editing				
Search:	Show 10 \$	entries Export data to CSV file		
Туре	Position	Genomic context	Source	PMID
A-to-I	chr17:7668822-7668823	3UTR	RADAR	24163250
C-to-T	chr17:7668829-7668830	3UTR	DARNED	22484847
C-to-T	chr17:7668835-7668836	3UTR	DARNED	22484847
Showing 1 to 3	3 of 3 entries			Previous 1 Next
RNA modificatio	n			
Search:	Show 10 \$	entries Export data to CSV file		
Туре	Position	Strand Genomic context	Source	PMID
m6A	chr17:7668822-7668823	- 3UTR	RMBase	25719671
Chaudian data	1 of 1 entries			Previous 1 Next

Figure 12 The "RNA editing/modification" annotation sub-module for a query RNA, TP53.

site position, the genomic context of the editing site, the source database, and the PMID of the reference that describes the experiment.

- 3. Use the "RNA modification" sub-module to show RNA modification sites located in the interacting regions (Fig. 12). There are six columns in this table, including modification type (m<sup>6</sup>A, m<sup>5</sup>C, pseudoU and others), the modification site position, the strand (+ or -), the genomic context, the source database, and the PMID of the reference that describes the experiment.
- 4. Use the "SNP" sub-module to identify SNPs located in the interacting regions (Fig. 13). Here again, we show the example from TP53. There are six columns in this table, including the SNP position, the genomic context, the reference base, the alternative base, the source database (i.e., dbSNP), and the PMID of the reference.
- 5. Explore the "pan-cancer mutation" sub-module, which provides cancer somatic mutations located in the interacting regions (Fig. 13). This sub-module includes somatic mutations from more than 40 cancer types. There are seven columns in this table, including the coordinate of the mutation site, the genomic context of the mutation site, the reference base, the mutated base, the cancer type (i.e., full name and its abbreviation), the source (i.e., where the mutation comes from), and the PMID of the reference.
- 6. The "gene expression level" sub-module provides the expression levels of the query gene across various cell lines and tissue types. We provide expression levels from 34 cell lines and tissue types in human, and 18 cell lines and tissue types in mouse. Expression levels are shown in a bar plot. The *x* axis represents cell and tissue types in the corresponding species, and the height of a bar represents the expression level (FPKM value). Check the FPKM value of a certain cell and or a tissue type by holding the mouse on the bar. The data source is showed below the bar plot with a hyperlink pointing to the paper in PubMed. (Fig. 14).

	RNA editing/modific	ation	SNPs/P	an-Cancer mutations	Gene expression	levels	
Gene SNPs							
Search:	Show	10 🛊 entries	Expo	rt data to CSV file			
P. 11				2.0			115
cbr17.7668835-7668836	311	P		G A/G	dbSNI	2 111	25122
Showing 1 to 1 of 1 entrie	is					Previous	1 Next
Pan-Cancer mutations							
Search:	Show	10 🗘 entries	Expo	rt data to CSV file			
Position	Genomic context	Ref base	Alt base	Cancer		Source	PMID
chr17:7674195-7674196	CDS	G	A	Head and Neck squamous ce	ll carcinoma [HNSC]	Pan-Cancer	26619011
chr17:7674196-7674197	CDS	т	G	Breast invasive carcinoma [B	IRCA]	Pan-Cancer	26619011
chr17:7674198-7674199	CDS	А	с	Breast invasive carcinoma [B	IRCA]	Pan-Cancer	26619011
chr17:7674198-7674199	CDS	A	т	Breast invasive carcinoma [B	IRCA]	Pan-Cancer	26619011
chr17:7674198-7674199	CDS	А	с	Esophageal carcinoma [ESC/	A]	Pan-Cancer	26619011
chr17:7674198-7674199	CDS	A	с	Glioblastoma multiforme [Gl	BM]	Pan-Cancer	26619011
chr17:7674198-7674199	CDS	А	т	Pancreatic Adenocarcinoma	[PAAD]	Pan-Cancer	26619011
chr17:7674198-7674199	CDS	A	G	Pancreatic Adenocarcinoma	[PAAD]	Pan-Cancer	26619011
chr17:7674199-7674200	CDS	т	A	Breast invasive carcinoma [B	RCA]	Pan-Cancer	26619011
chr17:7674199-7674200	CDS	т	A	Chronic Lymphocytic Leuker	nia [CLL]	Pan-Cancer	26619011
Showing 1 to 10 of 66 ent	tries			Previous	1 2 3	4 5 6	7 Next
13 The "SN	IP/pan-canc	er mutat	ion" ar	notation sub-n	nodule for a	ı query F	RNA, TP
NA annotation							
RBP binding	RNA editing/modifi	cation	SNPs/P	an-Cancer mutations	Gene expression	levels	
Export data to CSV file		RE	BFOX2 Ex	pression Values			
80							

Figure 14 The "gene expression level" annotation sub-module for a query RNA, RBFOX2.

#### COMMENTARY

#### **Background Information**

Databases that aim at collecting RRIs have been developed in recent years, including NPInter v3.0 (Hao et al., 2016), RAID v2.0 (Yi et al., 2017), and RAIN (Junge et al., 2017). The main limitation of these databases is that the RRIs in them are largely miRNA related. In addition, they usually contain little information on the collected RRIs, such as the cell and tissue types, the experimental methods, etc. NPInter is more informative, but still only provides limited molecular annotations for the RRIs. RISE differs from these existing similar databases mainly in two aspects: (i) RISE provides a comprehensive view on RRIs of diverse RNA types; (ii) RISE annotates RRIs with extensive functional annotations, including RNA editing/modification sites, pan-cancer mutations, and gene expression level, etc.

As mentioned above, there have been several sequencing-based experiments for transcriptome-wide RRI detection using different protocols. It is thus interesting to perform systematic comparison on the RRIs identified from different experiments. Indeed, we found substantial heterogeneity in the RRI datasets derived from different experiments (Gong et al., 2018). We observed limited overlaps of RRIs among different experimental technologies even for the same cell line. These results may suggest the dynamic and cellspecific nature of RNA interactions-and they also suggest that different sequencing-based experiments may have certain biases towards the RRIs they can identify. They may also indicate that the RRIs from these sequencingbased experiments are far from saturation. In the future, we need more powerful technologies and more comprehensive datasets to derive a complete understanding of the RNA interactions.

#### Troubleshooting

RISE was implemented with Hyper Text Markup Language (HTML), Cascading Style Sheets (CSS), and Hypertext Preprocessor (PHP), and has been tested on all major web browsers. Currently, RISE does not support searching with gene IDs or genomic coordinates. We also suggest that users try more gene aliases for failed searches. The gene aliases can be accessed at *https://www.ncbi.nlm.nih.gov/ gene*.

#### Acknowledgments

Ju et al.

This work was supported by the National Natural Science Foundation of China (Grant Numbers. 31671355, 91740204, and 31761163007) and the National Thousand Young Talents Program of China (to Q.C.Z.).

#### **Literature Cited**

- Aw, J. G., Shen, Y., Wilm, A., Sun, M., Lim, X. N., Boon, K. L., ... Wan, Y. (2016). In vivo mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Molecular Cell*, 62, 603–617. doi: 10.1016/j.molcel.2016.04.028.
- Bartel, D. P. (2004). MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell*, *116*, 281–297. doi: 10.1016/S0092-8674(04) 00045-5.
- Engreitz, J. M., Sirokman, K., McDonel, P., Shishkin, A. A., Surka, C., Russell, P., ... Lander, E. S. (2014). RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell*, 159, 188–199. doi: 10.1016/j.cell.2014.08.018.
- Gong, J., Shao, D., Xu, K., Lu, Z., Lu, Z. J., Yang, Y. T., & Zhang, Q. C. (2018). RISE: A database of RNA interactome from sequencing experiments. *Nucleic Acids Research*, 46, D194–D201.
- Guil, S., & Esteller, M. (2015). RNA-RNA interactions in gene regulation: The coding and noncoding players. *Trends in Biochemical Sciences*, 40, 248–256. doi: 10.1016/j.tibs.2015.03.001.
- Hao, Y., Wu, W., Li, H., Yuan, J., Luo, J., Zhao, Y., & Chen, R. (2016). NPInter v3.0: An upgraded database of noncoding RNA-associated interactions. *Database*, 2016, baw057. doi: 10.1093/database/baw057.
- Helwak, A., Kudla, G., Dudnakova, T., & Tollervey, D. (2013). Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*, 153, 654–665. doi: 10.1016/j.cell.2013.03.043.
- Ibba, M., & Soll, D. (2000). Aminoacyl-tRNA synthesis. Annual Review of Biochemistry, 69, 617– 650. doi: 10.1146/annurev.biochem.69.1.617.
- Junge, A., Refsgaard, J. C., Garde, C., Pan, X., Santos, A., Alkan, F., ... Gorodkin, J. (2017). RAIN: RNA-protein association and interaction networks. *Database*, 2017, baw167. doi: 10.1093/database/baw167.
- Kretz, M., Siprashvili, Z., Chu, C., Webster, D. E., Zehnder, A., Qu, K., ... Khavari, P. A. (2013). Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature*, 493, 231–U245. doi: 10.1038/nature11661.
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., ... Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, 19, 1639–1645. doi: 10.1101/gr.092759.109.
- Kudla, G., Granneman, S., Hahn, D., Beggs, J. D., & Tollervey, D. (2011). Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. *Proceedings of the*

National Academy of Sciences of the United States of America, 108, 10010–10015. doi: 10.1073/pnas.1017386108.

- Liu, T., Zhang, K., Xu, S., Wang, Z., Fu, H., Tian, B., ... Li, W. (2017). Detecting RNA-RNA interactions in E. coli using a modified CLASH method. *BMC Genomics*, *18*, 343. doi: 10.1186/s12864-017-3725-3.
- Lu, Z. P., Zhang, Q. C., Lee, B., Flynn, R. A., Smith, M. A., Robinson, J. T., ... Chang, H. Y. (2016). RNA duplex map in living cells reveals higherorder transcriptome structure. *Cell*, 165, 1267– 1279. doi: 10.1016/j.cell.2016.04.028.
- Melamed, S., Peer, A., Faigenbaum-Romm, R., Gatt, Y. E., Reiss, N., Bar, A., ... Margalit, H. (2016). Global mapping of small RNA-target interactions in bacteria. *Molecular Cell*, 63, 884– 897. doi: 10.1016/j.molcel.2016.07.026.
- Nguyen, T. C., Cao, X. Y., Yu, P. F., Xiao, S., Lu, J., Biase, F. H., ... Zhong, S. (2016). Mapping RNA-RNA interactome and RNA structure in vivo by MARIO. *Nature Communications*, *7*, 12023. doi: 10.1038/ncomms12023.
- O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., ... Pruitt, K. D. (2016). Reference sequence (Ref-Seq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 44, D733–745. doi: 10.1093/nar/gkv1189.
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing ge-

nomic features. *Bioinformatics*, 26, 841–842. doi: 10.1093/bioinformatics/btq033.

- Salmena, L., Poliseno, L., Tay, Y., Kats, L., & Pandolfi, P. P. (2011). A ceRNA hypothesis: The Rosetta Stone of a hidden RNA language? *Cell*, *146*, 353–358. doi: 10.1016/j.cell.2011.07. 014.
- Sharma, E., Sterne-Weiler, T., O'Hanlon, D., & Blencowe, B. J. (2016). Global Mapping of Human RNA-RNA Interactions. *Molecular Cell*, 62, 618–626. doi: 10.1016/j.molcel. 2016.04.030.
- Sugimoto, Y., Vigilante, A., Darbo, E., Zirra, A., Militti, C., D'Ambrogio, A., ... Ule, J. (2015). hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1. *Nature*, 519, 491–494. doi: 10.1038/nature14280.
- Yang, Y. C., Di, C., Hu, B., Zhou, M., Liu, Y., Song, N., ... Lu, Z. J. (2015). CLIPdb: A CLIP-seq database for protein-RNA interactions. *BMC Genomics*, 16, 51. doi: 10.1186/s12864-015-1273-2.
- Yi, Y., Zhao, Y., Li, C., Zhang, L., Huang, H., Li, Y., ... Wang, D. (2017). RAID v2.0: An updated resource of RNA-associated interactions across organisms. *Nucleic Acids Research*, 45, D115– D118. doi: 10.1093/nar/gkw1052.
- Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., ... Flicek, P. (2018). Ensembl 2018. *Nucleic Acids Research*, 46, D754–D761. doi: 10.1093/nar/ gkx1098.