circBase: a database for circular RNAs

PETAR GLAŽAR, PANAGIOTIS PAPAVASILEIOU, and NIKOLAUS RAJEWSKY

Max Delbrück Center for Molecular Medicine, 13125 Berlin, Germany

ABSTRACT

Recently, several laboratories have reported thousands of circular RNAs (circRNAs) in animals. Numerous circRNAs are highly stable and have specific spatiotemporal expression patterns. Even though a function for circRNAs is unknown, these features make circRNAs an interesting class of RNAs as possible biomarkers and for further research. We developed a database and website, "circBase," where merged and unified data sets of circRNAs and the evidence supporting their expression can be accessed, downloaded, and browsed within the genomic context. circBase also provides scripts to identify known and novel circRNAs in sequencing data. The database is freely accessible through the web server at http://www.circbase.org/.

Keywords: gene expression; circular RNA; database

INTRODUCTION

In recent years, several laboratories have reported thousands of circular RNAs (termed circRNAs) expressed in animal cells (Salzman et al. 2012, 2013; Jeck et al. 2013; Memczak et al. 2013). These single-stranded RNA molecules correspond to circular isoforms, often produced from exons in which the 5' and 3' ends are covalently closed to form a "head-to-tail" splice junction (or "backsplice"). It has been shown that in humans, circRNAs are the predominant isoform of exon-scrambling events and that circularization is a widespread and general feature of gene expression (Salzman et al. 2012). Two recent articles provided evidence that the human circRNAs (CDR1as/ciRS-7) can have regulatory function by acting as a miRNA sponge (Hansen et al. 2013; Memczak et al. 2013). However, it is still unknown if circRNAs have biological function since loss-of-function experiments are missing and, in many cases, are difficult to carry out due to the overlap between circular and linear isoforms. Nevertheless, numerous circRNAs seem well expressed (Jeck et al. 2013; Memczak et al. 2013; Salzman et al. 2013), often in a tissue-specific and developmental stage-specific manner (Memczak et al. 2013). Moreover, circRNAs are also unusually stable RNA molecules, presumably because their lack of ends prevents them from regulation by conventional RNA degradation pathways. Thus, circRNAs can be diagnostic of gene expression patterns and may constitute an interesting, novel class of biomarkers.

Thousands of circRNAs in humans, the mouse, and other animals have been reported by computational analysis of sequencing data, and the numbers are likely to grow. For an overview of detection methods used, see articles by Hoffmann et al. (2014) and Jeck and Sharpless (2014). To meet the need for an integrated database that enables the study, comparison, or download of circRNAs, we developed "circ-Base," with the following objectives: (1) For each circRNA, the evidence for its existence and expression should be summarized and accessible; (2) circRNAs should be presented within the genomic context and together with available expression or regulatory data; (3) a search engine should allow flexible queries, for example, by sequence, by gene, by genomic location, etc.; (4) it should be possible to intersect and download data sets in multiple ways and formats; and (5) users should be able to upload tracks with their own circRNA expression and intersect with the available data. In essence, we intend to serve the community by pooling together and making accessible published data sets from different laboratories. In its current implementation (version 1.0), circBase provides a solution to these objectives by (1) a simple searching interface that is convenient for users focusing on a small number of genes or loci; (2) various methods of bulk data retrieval; (3) merging, unifying, and annotating published data sets; and (4) linking circRNAs to the UCSC genome browser (Kent et al. 2002) and doRiNA (Anders et al. 2012), a database for post-transcriptional regulatory elements. Additionally, we (Memczak et al. 2013) have described a method for detecting circRNAs by computational analysis from RNA sequencing data. This method does not depend on preexisting genome annotation (e.g., known transcripts or splice sites).

Corresponding author: rajewsky@mdc-berlin.de

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.043687.113.

^{© 2014} Glažar et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see http://rnajournal.cshlp.org/site/misc/terms.xhtml). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/ by-nc/4.0/.

The latest version of software needed to perform this analysis is also available on circBase.

The database currently hosts data from various *Homo sapiens*, *Mus musculus*, *Caenorhabditis elegans*, and *Latimeria*

samples. We are expecting the range of organisms and samples to extend in the future. circBase will not cover viroids, which are already collected in other resources (Rocheleau and Pelchat 2006; Hamilton et al. 2011).

le browser	ase.org/cgi-bin/tablebro	user coi		- 01	公白 4	=	(Gwv	w.circbase.org/cgi-bin/s	simplesearch.cgi							• ° ☆ I	b
	www.wrg/cgrom/cablebro	mashAyi		÷C	H U 4												
iroE	Base						cir	cBase									
IICE	base						home		table browser		downloads	help					_
iome	list search table	prowser blat	download	s help			Expo	rt results: xisx	bat cs ^a	v fasta							
							organism	browser link)	strand o circRN	weige		samples o			tation •	transcript ^e gene symb	• 10
	Organism:	Human (hg19)	\$				hsa	chr1:36367548-36367938		rc_0000052 390	257	cd_34, K562, Ag04450		CDS, INTER	.coding. RNAL	NM_012199 EIF20	
	circRNA study:	Memczak 2013	\$				hsa	chr1:36939396-36941065		rc_0000055 1669		cd_34			coding. RNAL	NM_156039 CSF3	
		cd_19	0				hsa	chv1:40319641-40323089	i - hsa_cii	rc_0000057 3448	240	cd_34, Hs68_RNase, Hs68_control	11.67.1 N	ione ANNO CDS. INTER OVCO OVEX	OTATED. coding. RNAL,	NM_017646 TRITI	
	Samples:	cd_34 HEK293 neutr												OVER	KON		
		ANNOTATED	0														
	Annotation:	ANTISENSE CDS															
		coding downstream_end															
	Experimental evidence:		experimentally a														
	Repeats:	exclude re		select only repea			С						splice ju				
	Region: Genomic length:	genome max	position	chrX:139,865,2	240-139,880,92		3	' genomic reference se						ic reference			-
	Spliced seq length:	min max					A	ARAGCTGCCARGTTG					AAGTTGGT	100			тст
	Score:	min max	- T					-50 -40	-30	-20	-		1	10	20	30	
	Unique reads:	[min] [max	- T														
			_														
	Search	Reset Form	?					A		GN			•••				
																	••••
VID	BERLIN-BUCH	circBase is devel Systems Biology,	oped by the Raje	wsky lab at the l	Berlin Institute f	or Medical								. 16			••••
Base	lla Firefox	brd.cgi?id=hsa_circ_1	1000052	* @	() () () () () () () () () () () () () (● ●● +	Hu Hu	n chrX:139,865,339 man chrX:139,865,3 dorina.mdc-berlin.de	3 +		me Browse	r v265 - Mo	ozilla Firef		ट <u>द</u> े	â ∔ 1	
Base - Mozi Base Dwww.circb	+		000052	* C	() () () () () () () () () () () () () (• • • =	Home	iman chrX:139,865,3 dorina.mdc-berlin.do Genomes	e/cgi-bin/hgTr Blat	acks Tables	Gene Sor	ter P(CR [T DNA	Convert	自 🖡 🏦 PDF	E
Base - Mozi Base Www.circb	+ ase.org/cgi-bin/singlered	ord.cgi?id=hsa_circ_	downloads		☆ 6 4			ıman chrX:139,865,3 dorina. mdc-berlin.d	8 + le/cgi-bin/hgTr Blat e Brows	racks Tables Ser on	Gene Sor Huma	ter Po In Feb	cr (. 200	DNA (GR	convert RCh3'		• • •
Base - Mozi Base Www.circt	+ ase.org/cgi-bin/singlered	ord.cgi?id=hsa_circ_] ☆ 自 - I			iman chrX:139,865,3 dorina.mdc-berlin.do Genomes C Genome	a + le/cgi-bin/hgTr Blat e Brows < > >>	Tables Ser On Ses Zoo: 9,866,824	Gene Sor Huma m in 1.5x	ter Po n Feb 3x 10 gene	cr (. 200	DNA 19 (GR 200m o	Convert RCh3' out 1.5x		F E
Base - Mozi Base Www.circt irce ome	* ase.org/cgl-bin/singlered asec asec bist search bsa_circ_0000052	ord.cgi?id=hsa_circ_			<u>}</u> ☆ □ 1			anan chrX:139,865,3 dorina.mdc-berlin.d Genomes GC Genomes vve <<< << · · ·	a + le/cgi-bin/hgTr Blat e Brows < > >>	Tables Ser on >>> Zoor 9,866,824	Gene Sor Huma m in 1.5x bp. conf	ter PC In Feb 3x 10 gene igure	CR [0. 200 0x base	DNA 99 (GR 200m o	Convert RCh3' out 1.5x jump c	PDF PDF 7/hg19 3x 10x lear size 1	r E) A
Base - Mozi Base www.circt irce ome ID Alias	* ase.org/cgl-bin/singlered ase.org/cgl-bin/singlered ase.org/cgl-bin/singlered bsa_circ_0000052 bsa_circ_000045	ord.cgi?id=hsa_circ_ rowser blat			∂ ∂ 1 1 1 1 1 1 1 1			aman chrX:139,865,3 dorina.mdc-berlin.do Genomes C Genomes we <<< <<	8 + le/cgi-bin/hgTr Blat e Brows < > >> 39,865,339-139	Tables Ser On >>> Zoo 9,866,824	Gene Sor Huma m in 1.5x bp. conf	ter Po n Feb 3x 10 gene igure	CR C CR C C 200 Dx base		Convert RCh3' out 1.5x jump c	PDF PDF 7/hg19 3x 10x lear size 1	• • •
Base - Mozi Base Swww.circt irce ome	* ase.org/cgl-bin/singlered asec asec bist search bsa_circ_0000052	ord.cgi?id=hsa_circ_ rowser blat			· 소 @ 4			man chrX:139,865,3 dorina mdc-berlin.d Cenomes C Genomes C Genomes con/search chrX:13 (ores (ai7.1))	8 + le/cgi-bin/hgTr Blat e Brows < > >> 39,865,339-139	Tables Ser on >>> Zoor 9,866,824	Gene Sor Huma m in 1.5x bp. conf	ter Po n Feb 3x 10 gene igure	CR C CR C C 200 Dx base		Convert RCh3' out 1.5x jump c	PDF PDF 7/hg19 3x 10x lear size 1	r E) A
Base - Mozi Base Downw.circt Dirce Dome ID Alias Position Strand Genomic	* ase.org/cgl-bin/singlered ase.org/cgl-bin/singlered ase.org/cgl-bin/singlered bsa_circ_0000052 bsa_circ_000045	ord.cgi?id=hsa_circ_ rowser blat] ☆ 血 - 4			man chr.: 139,865,3 dorina.mdc-berlin.d Genomes Gen	8 + le/cgi-bin/hgTr Blat e Brows < > >> 39,865,339-139	Tables Ser On >>> Z001 9,866,824	Gene Sor Huma m in 1.5x bp. conf bp. conf 1596664 1596664 1596664 1596664	ter Po In Feb 3x 10 gene igure down satzam refse one refse one refse one	CR C . 200 0x base 12413 thinks a con- 14 14 14 14 14 14 14 14 14 14		Convert RCh3' out 1.5x jump c	PDF PDF 7/hg19 3x 10x lear size 1	r E) A
Base - Mozi Base Base Down circle Dirce Di Dirce Dirce Dirce Dirce Dirce Dirce Dirco	+ ase.org/cg-bin/singlerce base ist search table b hsa_circ_000052 hsa_circ_000045 chr1103057548-3036783 +	ord.cgi?id=hsa_circ_ rowser blat			() () () () () () () () () () () () () (man chr.X.139,865,3 dorina mdc-berlin.dr Cenomes CC Genomes CVV e ecc ecc jows car7.13 jows car7.13 series car7	Blat Blat Blat Blat Blat C Blat Blat Blat Blat Blat C Blat Blat Blat Blat Blat Blat Blat Blat	acks Tables Set On >>> Z00; >>> Z00; >>> Z00; Set Dase Set Dase USEC Genes (Rec	Gene Sor Huma m in 1.5x bp. conf bp. conf source conf the source the the source the source the the the the the the the the the th	ter Po n Feb 3x 10 gene igure el el el statasson a MMB (ross a MMB (ross a MMB (ross a Statasson a MMB (ross a Statasson a MMB (ross a Statasson a Statasson	CR E 2000 0x base 12913 10945 £ Cor 12913 10945 £ Cor 12913 12915 1291	DNA DNA DO (GR 2 ZOOM O 2 ZOOM O 1000 100	Convert CCh3' put 1.5x jump c seein	POF 7/hg19 3x 10x lear size 1	• • •
Base - Mozi Base Base Ince Ince Alias Position Strand Genomic Length Spliced	Asse.org/rg/bin/kindered Asse.org/rg	ord.cgi?id=hsa_circ_ rowser blat			☆ 値 4		Home UCSS moster	man chr.X.139,865,3 dorina mdc-berlin,dr Cenomes CCenomes CCenomes CCenomes CCenomes CCenomes CCenomes Sector (Control Sector (Control (Control (Control Sector (Control (Control (Control Sector (Control (Control (Control (Control Sector (Control (Blat Blat Blat Blat Blat C Blat Blat Blat Blat Blat C Blat Blat Blat Blat Blat Blat Blat Blat	acks Tables Set On Set On Set Set Control Tables Set Control Tables Set Control Tables Set Control Tables T	Gene Sor Huma m in 1.5x bp. conf bp. conf st sources conf tax tax tax tax tax tax tax tax tax tax	ter PO In Feb 3x 10 gene igure i	CR E , 2000 DX base CRUES 6 CRUES 6 CRUES 6 CRUES C	2000 COR 2 ZOOM O 2 ZOOM O 1 7015 1 7015	Convert CCh3' put 1.5x jump c seein	POF 7/hg19 3x 10x lear size 1	E) A
Base - Mozi Base Base Internet Base Internet Base Base Base Base Base Base Base Base	Ass.org/cg-bin/singlerce CASCE list search table b hsa_cire_000045 chr130567548-36367831 + 390 257 ANNOTATED	ord.cgi?id=hsa_circ_ rowser blat) 会 自 4 4 4 4 4 4 4 4 4 4 4 4 4		Hom UCSS ma positi	man chr.X.139,865,3 dorina mdc-berlin,dr Cenomes Cenom	Blat e Brows sogs, 865, 339-135 reserved	acks Tables Set On 9,866,824 Set base S	Gene Sor Human min 1.5x bp. conf Street, United Hans Hans Hans Hans Hans Hans Hans Hans	ter Po n Feb 3x 11 gene gure 1000 Falance 1000 Falan	CR E , 2000 0x base 0x base	200A 29 (GR 2 ZOOM O 2	Convert CCh3' put 1.5x jump c seein	POF 7/hg19 3x 10x lear size 1	• • •
Base - Mozi Base Democirct Incession Alias Position Strand Genand Genand Genand Genand Genand Genand Genand Genand Genand Genand	Asse.org/cg-bin/singlerce ass.org/cg-bin/singlerce ass.org/cg-bin/sing	ord.cgi?id=hsa_circ_ rowser blat] ☆ 白 4		Hom UCSS ma positi	wman chr.X:139,865,3 dorina.mdc-berlin.dr Cenomes Cenomes We exec	Blat e Brows sogs, 865, 339-135 reserved	acks Tables Set On >>> Z00; >,866,824 Et base UCSC Gens (Re L Orgen Fg)	Gene Sor Huma min 1.5x bp. conf sorted, unifed to the Manan est Manan est Ma	ter PA Defense gene igure	CR C CR CR C CR	CONA O	Convert CCh3' put 1.5x jump c seein	POF 7/hg19 3x 10x lear size 1	• • •
Base - Mozi Base Www.circt Ince Ince Alias Position Strand Genomic Length Spliced sequence Length	Asse.org/cg-bin/kingkree SASE ist/search kable b is	ord.cgi?id=hsa_circ_ rowser blat] ☆ 白 4		Hom UCSS ma positi	man chr.X.139,865,3 dorina mdc-berlin,dr Cenomes Cenom	Blat e Brows sogs, 865, 339-135 reserved	acks Tables Set On >>> Z00; >,866,824 Et base UCSC Gens (Re L Orgen Fg)	Gene Sor Huma min 1.5x bp. conf sorted, unifed to the Manan est Manan est Ma	ter PA Defense gene igure	CR C CR CR C CR	CONA O	Convert CCh3' put 1.5x jump c seein	POF 7/hg19 3x 10x lear size 1	• • •
lase - Moziase Base www.circtcles iirccles iirccles anne log Alias Spliced sequence length spliced sequence length monotation Repeats Best	Asse.org/cg-bin/kingkree SASE ist/search kable b is	ord.cgi?id=hsa_circ_ rowser blat] ☆ 白 4		Hom UCSS ma positi	man chr.X.139,865,3 dorina mdc-berlin,dr Cenomes Cenom	Blat e Brows sogs, 865, 339-135 reserved	acks Tables Set On >>> Z00; >,866,824 Et base UCSC Gens (Re L Orgen Fg)	Gene Sor Huma min 1.5x bp. conf sort to the total total text text text text text text text tex	ter PA Defense gene igure	CR C CR CR C CR	CONA O	Convert CCh3' put 1.5x jump c seein	POF 7/hg19 3x 10x lear size 1	• • •
asse - Mozizase Base Www.circtce iircce b Mias Position Strand Spliced length Spliced length spliced length Repeats Best Best Best Strancpi Composition Composition Strand Spliced Composition Strand Spliced Composition Strand Spliced Composition Strand Spliced Composition Composition Co	Assc.org/rgb-bin/kingkreed Assc.org/rgb-bin/kingkreed Bas_cric_000052 hsa_cric_000052 hsa_cric_000055 hsa_cric_00055 hsa_cric_0005 hsa_cric_00055 hsa_crintatt hsa_cric_00055 hsa_cr	ord.cgi?id=hsa_circ_ rowser blat] ☆ 血		Hom UCSS ma positi	man chr.X 139,865,3 dorina mdc-berlin.di Canones Canon	Blat e Brows sogs, 865, 339-135 reserved	acks Tables SET ON SET ON SET ON SET ON CONTRACTOR FOR UTIC CONTRACTOR FOR Ought	Gene Sor Human bp. con Con Con Con Con Con Con Con Con Con C	ter Per n Febb 3x 10 gene igure 1022-11 10	CR C CR CR C CR	CONA	Convert Cch3' out 1.5x jump c Steel non (cz)	POF 7/hg19 3x 10x lear size 1	480
Asse - Moziase Base www.dircter iircces wme ID Alias Strand Genomic length Spliced sequence length annotation Repeats Best transcript	Association of the second sec	ord.cgitid=hsa_circ_t	downloads	: help		Î	Hom UCSS ma positi	man chrX:139,865,3 dorina md-berlin.d Cenomes Cenomes Ve eee e o over cear.is over cear.is over cear.is norsearch chrX:13 merse dorig merse dorig balane data mers the370° Cenames the const of the cons		acks Tables SET ON SET ON SET ON SET ON CONTRACTOR FOR UTIC CONTRACTOR FOR Ought	Gene Sor Human bp. con Con Con Con Con Con Con Con Con Con C	ter Per n Febb 3x 10 gene igure 1022-11 10	CR C CR CR C CR	CONA	Convert Cch3' out 1.5x jump c Steel non (cz)	POF 7/hg19 3x 10x kear size 1	480
asse - Mozizase Base Www.circtce iircce b Mias Position Strand Spliced length Spliced length spliced length Repeats Best Best Best Strancpi Composition Composition Strand Spliced Composition Strand Spliced Composition Strand Spliced Composition Strand Spliced Composition Composition Co	Asscarging-bin/kingkered asscarging-bin/kingk	ord.cgitid=hsa_circ_t	downloads	i help	unique reads	unique reads		man chrX:139,865,3 dorina md-berlin.d Cenomes Cenomes Ve eee e o over cear.is over cear.is over cear.is norsearch chrX:13 merse dorig merse dorig balane data mers the370° Cenames the const of the cons	 el/cgi-bin/hgTr el/gi-bin/hgTr el/gi-bin/hgTr el/gi-bin/hgTr el/gi-bin/hgTr state	acks Tables Set On O ,866,824 Res Dase Comparison O regit	Cente Societa Marine Societa Control Contrel Contrel Contrel Contrel Con	EXT P(I) The second s	CR CR C , 2000 Dx base deal The state of the state o	V V	Convert CCh3' Jut 1.5x Jump c Seel Seel Seel Seel Seel Seel Seel See	the second	
asse - Mozizase Base Www.circtce iircce b Mias Position Strand Spliced length Spliced length spliced length Repeats Best Best Best Strancpi Composition Composition Strand Spliced Composition Strand Spliced Composition Strand Spliced Composition Strand Spliced Composition Composition Spliced Composition Compositio	Asscarg(rg)bin/kingtered Ascarg(rg)bin/kingtered Ascarg(rg)bin/kingtered Asscarg(rg)bi	rowser blat	downloads	i help	unique reads li	unique reads		man chrX:139,865,3 dorina mdc-berlin.d. Cenomes Cenome	 el/cgi-bin/hgTr el/gi-bin/hgTr el/gi-bin/hgTr el/gi-bin/hgTr el/gi-bin/hgTr state	acks Tables Set On 9,866,824		ter PP Provide a second	CCR C C C C C C C C C C C C C C C C C C	You (1) Y	Convert CCh3' Jut 1.5x Jump c Seel Seel Seel Seel Seel Seel Seel See	the second	
Base - Mozi Base Base Base Come ID Alias Position Alias Position Alias Source Length Alias Sepliced Sequence Length Alias Sepliced Sequence Length Repeats Best Best Care Care Care Care Care Care Care Care	Asscarg/cg-bin/kingkered asscarg/cg-bin/kingk	nple	downloads total reads circular junction	nique reads circular junction	uniquo reado linea-5 l junctions jun	reads mears inears 2 61		man chr.X 139,865,3 dorina mdc-berlin.d Cenomes Cenome	 el/cgi-bin/hgTr el/gi-bin/hgTr el/gi-bin/hgTr el/gi-bin/hgTr el/gi-bin/hgTr state	acks Tables Set On O ,866,824 Res Dase Comparison O regit		ter PC	CCR C C C C C C C C C C C C C C C C C C	You (1) Y	Convert CCh3' Jut 1.5x Jump c Seel Seel Seel Seel Seel Seel Seel See	the second	
asse - Mozi Base Base Base Come Dome Dome Alias Alias Position Alias Sopied Genomic length Annotation Repeats Best Best Best Come Come Come Come Come Come Come Come	Assc.org/cg-bin/kindered assc.org/cg-bin/kindered assc.org/cg-bin/kindered assc.org/cg-bin/kindered assc.org/cg-bin/kindered baa.circ_0000052 thaa_circ_0000052 thaa_circ_000005 thaa_circ_000005 thaa_circ_000005 thaa_circ_000005 thaa_circ_000005 thaactirc_000005 thaactirc_00005 thaactirc_0005 thaactirc_0005 thaactirc_00005 thaactirc_00005 thaactirc_0005 thaattirc_0005 thaa	nple	total reads circular junction 25 2 1	: help	unique reads linear-5 lunctions jur	reads near-3' n near-3' n near-3' n n n n n n n n n n n n n n n n n n n		man chrX:139,865,3 dorina mdc-berlin.d. Cenomes Cenome	 el/cgi-bin/hgTr el/gi-bin/hgTr el/gi-bin/hgTr el/gi-bin/hgTr el/gi-bin/hgTr state	acks Tables Set On O ,866,824 Res Dase Comparison O regit		ter PP Provide a second	CCR C C C C C C C C C C C C C C C C C C	You (1) Y	Convert CCh3' Jut 1.5x Jump c Seel Seel Seel Seel Seel Seel Seel See	the second	
asse - Mozi Base Base Base Come Dome Dome Alias Alias Position Alias Sopied Genomic length Annotation Repeats Best Best Best Come Come Come Come Come Come Come Come	Asscarg/cg-bin/kingkered asscarg/cg-bin/kingk	nple	total reads circular junction 25 2 1	tunique reads circular junction 5	unique reads linear-5' li junctions jun 0 66	reads mears interns 2 61		man chr.X 139,865,3 dorina mdc-berlin.d Genomes C Genomes C Gen	 el/cgi-bin/hgTr el/gi-bin/hgTr el/gi-bin/hgTr el/gi-bin/hgTr el/gi-bin/hgTr state	acks Tables Set On O ,866,824 Res Dase Comparison O regit		ter PP Provide a second	CCR C C C C C C C C C C C C C C C C C C	You (1) Y	Convert CCh3' Jut 1.5x Jump c Seel Seel Seel Seel Seel Seel Seel See	the second	
Base - Mozi Base Base Base Come ID Alias Position Alias Position Alias Source Length Alias Sepliced Sequence Length Alias Sepliced Sequence Length Repeats Best Best Care Care Care Care Care Care Care Care	Assc.org/cg-bin/kindered assc.org/cg-bin/kindered assc.org/cg-bin/kindered assc.org/cg-bin/kindered assc.org/cg-bin/kindered baa.circ_0000052 thaa_circ_0000052 thaa_circ_000005 thaa_circ_000005 thaa_circ_000005 thaa_circ_000005 thaa_circ_000005 thaactirc_000005 thaactirc_00005 thaactirc_0005 thaactirc_0005 thaactirc_00005 thaactirc_00005 thaactirc_0005 thaattirc_0005 thaa	nple	total reads circular junction 25 2 1	tunique reads circular junction 5	unique reads linear-5' li junctions jun 0 66	reads mears interns 2 61		man chr.X 139,865,3 dorina mdc-berlin.d Genomes C Genomes C Gen		acks Tables Set On Set On O ,866,824 Set Date Comparison O Gaugett Set Date Set Date Comparison O Gaugett Set Date Co		ter PC	CR C C C C C C C C C C C C C C C C C C	Control of the second sec	Convert CCh3' Jut 1.5x Jump c Seel Seel Seel Seel Seel Seel Seel See	the second	
asse - Moziase - Mozi - Moziase - Mo	Assc.org/cg-bin/kindered assc.org/cg-bin/kindered assc.org/cg-bin/kindered assc.org/cg-bin/kindered assc.org/cg-bin/kindered baa.circ_0000052 thaa_circ_0000052 thaa_circ_000005 thaa_circ_000005 thaa_circ_000005 thaa_circ_000005 thaa_circ_000005 thaactirc_000005 thaactirc_00005 thaactirc_0005 thaactirc_0005 thaactirc_00005 thaactirc_00005 thaactirc_0005 thaattirc_0005 thaa	nple	total reads circular junction 25 2 1	tunique reads circular junction 5	unique reads linear-5' li junctions jun 0 66	reads mears interns 2 61		man chr.X 139,865,3 dorina mdc-berlin.d Genomes C Genomes C Gen	electional and a second and a s	acks Tables Set One Set One Set One Comparison	Cores Soon Human by conf core core core core core core core core	ter PC	CR () CR ()	V Try USP	Convert CCh3' Jut 1.5x Jump c Seel Seel Seel Seel Seel Seel Seel See	PDF 7/hg19 3x 10x ear size 1 20 20 20 20 20 20 20 20 20 20 20 20 20	

FIGURE 1. (Legend on next page)

RESULTS AND DISCUSSION

Searching circBase

There are three main ways to query circBase: (1) Simple search, (2) List search, and (3) Table browser.

Simple search, available from the server homepage, is intended for simple queries by identifiers, genomic location, sequence, gene description, or Gene Ontology term identifiers. Apart from internal identifiers, we provide support for RefSeq transcript identifiers, HGNC gene symbols, and names of particular circRNAs from the literature (e.g., CDR1as). We cannot exclude the possibility of different investigators using the same name for multiple circRNAs. Therefore, only circBase internal identifiers, which are based on the genomic location of a head-to-tail splice site, uniquely identify database records. In the search field, genomic locations can be given in a UCSC compatible format (e.g., chrX:134,619,072-134,756,686) or with whitespace-separated chromosome, start and end coordinates. When searching by genomic coordinates, the list of results will contain all circRNAs that overlap the specified genomic region. Search by gene description terms (e.g., "apoptosis") will return circRNAs that are produced from transcripts of genes matching the description. Description terms are derived from official full gene names in NCBI Entrez Gene database (Maglott et al. 2011). Users can also query the database by DNA or RNA sequence, in which case the number of exact sequence matches per circRNA will be shown. This search option is limited to sequences >6 nucleotides (nt), and only exact matches of queried sequence, or its reverse complement, will be reported. Searches can be limited to a particular animal species. For the nonexact sequence search, users may use Blat (Kent 2002), available from the main menu. circBase hosts Blat references of human, mouse, and C. elegans circRNAs. circRNAs are cut opposite to the head-to-tail splice junction to keep it intact and allow looking for continuous matches that overlap the junction.

List search allows the intersection of a large number of search terms with database contents. Upon selecting the organism and assembly, the user can paste or upload a list of any identifier type supported by circBase. It is also possible to submit a list of genomic regions (for example in BED format) and retrieve all overlapping circRNAs in the database. The Table browser can be used for conditional data retrieval (Fig. 1A). After selecting the organism and experiment of interest, the user can further refine the selection by a number of options, such as presence in a particular sample, range of genomic or spliced sequence lengths, number of reads supporting the head-to-tail splice junction, and many more. This interface was developed to enable simple generation of complex queries such as "Retrieve all the circRNAs from experiment A that are present in HEK293 cells, intergenic, between 1000 and 2000 bp long and do not overlap repeat sequences."

All search results are returned as a hyperlinked table (Fig. 1B). The table contains basic information on every circRNA that matched the search criteria, while additional information can be accessed by following the related hyperlink. Clicking the genomic position links to the local copy (Anders et al. 2012) of UCSC genome browser, while gene symbols and best transcripts are linked to the NCBI Entrez Gene (Maglott et al. 2011) and Reference Sequence (RefSeq) (Pruitt et al. 2007) databases, respectively (with the exception of C. elegans where they are linked to WormBase) (Chen et al. 2005). Data set and circRNA IDs link to pages with additional information. Apart from data visible in a results table, circRNA single-record pages (Fig. 1D) contain evidence for the occurrence of a particular circRNA, such as prediction score, read counts, alignment of reads to a head-to-tail splice junction (Fig. 1C), and the literature-based evidence for circRNAs that was verified experimentally. Genomic positions are linked to our local UCSC Genome browser, which is a part of the doRiNA database. Therefore, in addition to information retrieved from circBase, the user can explore RNA binding protein target sites, predicted miRNA binding sites and ribosome profiling tracks (Fig. 1E).

Data export

Results can be exported using options in the "Export Results" ribbon below the main menu. Results tables are available in comma-separated values (csv), tab-delimited text (tsv), and Excel Workbook (xlsx) formats, while nucleotide sequence of particular circRNAs can be downloaded as a compressed fasta file. Upon clicking the "fasta" option from the results

FIGURE 1. (*A*) circBase table browser. Table browser enables the user to design queries for conditional data retrieval. In this example, the user will retrieve all circRNAs from a CD34+ sample, 100–500 nt in length, that are transcribed from coding regions and do not overlap repeat sequences. (*B*) circBase results page. The user is presented with basic information about each circRNA that matched the search query. Columns, from *left* to *right*, contain information on the following: (1) the organism data came from; (2) genomic position; (3) the strand circRNA is transcribed from; (4) a unique identifier; (5) genomic length; (6) length of an in silico predicted spliced form; (7) list of samples the circRNA was observed in; (8) number of reads overlapping the circular junction; (9) repeat sequences overlapping the 3' and 5' ends that give rise to a head-to-tail splice junction; (10) circRNA annotation (annotation terms are described in documentation available on the circBase website); (11) an overlapping transcript; (12) gene symbol of the overlapping gene; and (13) a list of studies in which the particular circRNA was detected. (*C*) Reads mapped to a head-to-tail splice junction. Alignments of reads that support a head-to-tail splice junction can be retrieved from a single record page. In addition to data available form a results page, the user can explore more information about a particular circRNA on a single record page. (*E*) UCSC genome browser on doRiNA, a database for post-transcriptional regulatory elements. A region surrounding the *CDR1as* circRNA is shown. In addition to the basic genome browser tracks, the user can explore RNA binding protein PAR-CLIP tracks, ribosome profiling data, miRNA target prediction tracks, and much more.

page, the user can select between genomic or inferred spliced sequence, define a number of upstream and downstream flanking nucleotides, or download an arbitrary number of nucleotides around the 5' and 3' splice sites.

Additional flat file export options are available from the "downloads" page.

Conclusions and future work

circBase facilitates circRNA research by (1) providing users with a single repository that collects, unifies, and annotates circRNA data in standardized formats; (2) extending public data sets with additional basic analysis; (3) integrating this data with external resources, such as the UCSC Genome Browser and NCBI databases; and (4) enabling users to interact with the database contents through a simple interface. While we developed circBase, a special emphasis was put on building a simple and self-explanatory interface, as well as writing thorough user documentation that covers all the advanced options that may not be used routinely. We intend to regularly update the database with newly published data. Direct data submission by the users is currently not supported, but the users are encouraged to contact us with requests for adding their data to circBase.

Availability

The database is freely accessible through the web server at http://www.circbase.org/.

MATERIALS AND METHODS

Implementation and contents

Our web server user interface is implemented using HTML, CSS, and JavaScript. Data are stored in a MySQL database, which is

TABLE 1. Circular RNA studies available in circBase version 1.0

			No. of circRNA	Sequencing study			
circRNA study	Organism	Sample	candidates	Reference	Data access		
Jeck et al. (2013)	H. sapiens	Hs68	7771	Jeck et al. (2013)	SRA050270		
Memczak et al. (2013)	H. sapiens	CD19+	1303	Salzman et al. (2012)	GSM835231		
		CD34+	528	Salzman et al. (2012)	GSM835232		
		HEK293	239	Memczak et al. (2013)	GSM1065666		
		Neutrophils	417	Salzman et al. (2012)	GSM835233		
	M. musculus	Adult brain 1	970	Vivancos et al. (2010)	SRA009091		
		Adult brain 2	537	Vivancos et al. (2010)	SRA009091		
		ESC	761	Huang et al. (2011)	GSE22959		
		Fetal head	1184	Huang et al. (2011)	GSE22959		
	C. elegans	1 cell	217	0			
	Ŭ	2cell	178				
		fem1 (oocytes)	316				
		spe9 (oocytes)	364				
		Sperm	270				
		sperm_act	297				
Salzman et al. (2013)	H. sapiens	A549	14,970	Rosenbloom et al. (2013)	ENCODE ^a		
		AG04450	16,808	Rosenbloom et al. (2013)	ENCODE		
		BJ	12,705	Rosenbloom et al. (2013)	ENCODE		
		GM12878	18,274	Rosenbloom et al. (2013)	ENCODE		
		H1-hESC	16,017	Rosenbloom et al. (2013)	ENCODE		
		HeLa S3	15,195	Rosenbloom et al. (2013)	ENCODE		
		HepG2	14,024	Rosenbloom et al. (2013)	ENCODE		
		HMEC	1107	Rosenbloom et al. (2013)	ENCODE		
		HSMM	2940	Rosenbloom et al. (2013)	ENCODE		
		HUVEC	8143	Rosenbloom et al. (2013)	ENCODE		
		K562	27,307	Rosenbloom et al. (2013)	ENCODE		
		MCF-7	5331	Rosenbloom et al. (2013)	ENCODE		
		NHEK	13,760	Rosenbloom et al. (2013)	ENCODE		
		NHLF	2209	Rosenbloom et al. (2013)	ENCODE		
		SK-N-SH RA	12,745	Rosenbloom et al. (2013)	ENCODE		
Zhang et al. (2013)	H. sapiens	H9	103	Zhang et al. (2013)	GSE48003		
Nitsche et al. (2013)	L. chalumnae	Muscle	767	Amemiya et al. (2013)	SRR401852		
	L. menadoensis	Liver	588	Pallavicini et al. (2013)	SRR576100		
		Testis	816	Pallavicini et al. (2013)	SRR576101		

Listed are references to circRNA screening publications, samples screened, number of detected circRNAs, and references to sequencing data. ^aENCODE data deposited at http://hgdownload.cse.UCSC.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/. interfaced to the front end by a series of server-side Perl CGI scripts. A web browser fully compatible with HTML5 and CSS3 is required for optimal performance. We have thoroughly tested the web server in Google Chrome 31.0, Mozilla Firefox 25.0, and Internet Explorer 10.0.

To the best of our knowledge, circBase contains data from all studies of large-scale circRNA identification published to date (Table 1; Jeck et al. 2013; Memczak et al. 2013; Nitsche et al. 2013; Salzman et al. 2013; Zhang et al. 2013). In addition to the information made available by the investigators, we have annotated all circRNA transcripts, predicted their putative spliced forms, and, where applicable, provided alignments of reads spanning head-totail junctions. Transcript reannotation was necessary to standardize the content and level the amount of information contained across some publications (Jeck et al. 2013; Salzman et al. 2013). Unique identifiers are assigned to circRNAs and will provide fixed references for future circBase releases.

ACKNOWLEDGMENTS

We thank all members of the N. Rajewsky laboratory for helpful discussions and comments on circBase interface and contents. We thank Christine Kocks and Benedikt Obermayer for their valuable suggestions on improving the manuscript. Marvin Jens is thanked for providing the software for transcript annotation and splice site predictions. P.G. was partially funded by the DFG Graduate School "Computational Systems Biology" (CSB-GRK 1772) and DEEP–German Epigenome Programme. P.P. was funded by the MDC-NYU PhD Exchange Programme.

Received November 28, 2013; accepted July 8, 2014.

REFERENCES

- Amemiya CT, Alfoldi J, Lee AP, Fan S, Philippe H, Maccallum I, Braasch I, Manousaki T, Schneider I, Rohner N, et al. 2013. The African coelacanth genome provides insights into tetrapod evolution. *Nature* **496**: 311–316.
- Anders G, Mackowiak SD, Jens M, Maaskola J, Kuntzagk A, Rajewsky N, Landthaler M, Dieterich C. 2012. doRiNA: a database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res* 40: D180–D186.
- Chen N, Harris TW, Antoshechkin I, Bastiani C, Bieri T, Blasiar D, Bradnam K, Canaran P, Chan J, Chen CK, et al. 2005. WormBase: a comprehensive data resource for *Caenorhabditis* biology and genomics. *Nucleic Acids Res* 33: D383–D389.
- Hamilton JP, Neeno-Eckwall EC, Adhikari BN, Perna NT, Tisserat N, Leach JE, Levesque CA, Buell CR. 2011. The Comprehensive Phytopathogen Genomics Resource: a web-based resource for data-mining plant pathogen genomes. *Database (Oxford)* 2011: bar053.

- Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. 2013. Natural RNA circles function as efficient microRNA sponges. *Nature* 495: 384–388.
- Hoffmann S, Otto C, Doose G, Tanzer A, Langenberger D, Christ S, Kunz M, Holdt L, Teupser D, Hackermueller J, et al. 2014. A multi-split mapping algorithm for circular RNA, splicing, trans-splicing, and fusion detection. *Genome Biol* 15: R34.
- Huang R, Jaritz M, Guenzl P, Vlatkovic I, Sommer A, Tamir IM, Marks H, Klampfl T, Kralovics R, Stunnenberg HG, et al. 2011. An RNA-Seq strategy to detect the complete coding and non-coding transcriptome including full-length imprinted macro ncRNAs. *PLoS One* 6: e27288.
- Jeck WR, Sharpless NE. 2014. Detecting and characterizing circular RNAs. *Nat Biotechnol* **32:** 453–461.
- Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, Marzluff WF, Sharpless NE. 2013. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 19: 141–157.
- Kent WJ. 2002. BLAT: the BLAST-like alignment tool. *Genome Res* 12: 656–664.
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* 12: 996–1006.
- Maglott D, Ostell J, Pruitt KD, Tatusova T. 2011. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res* **39**: D52–D57.
- Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M, et al. 2013. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495: 333–338.
- Nitsche A, Doose G, Tafer H, Robinson M, Saha NR, Gerdol M, Canapa A, Hoffmann S, Amemiya CT, Stadler PF. 2013. Atypical RNAs in the coelacanth transcriptome. J Exp Zool B Mol Dev Evol doi: 10.1002/jez.b.22542.
- Pallavicini A, Canapa A, Barucca M, Alf Ldi J, Biscotti MA, Buonocore F, De Moro G, Di Palma F, Fausto AM, Forconi M, et al. 2013. Analysis of the transcriptome of the Indonesian coelacanth *Latimeria menadoensis*. BMC Genomics 14: 538.
- Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35:** D61–D65.
- Rocheleau L, Pelchat M. 2006. The Subviral RNA Database: a toolbox for viroids, the hepatitis *delta virus* and satellite RNAs research. *BMC Microbiol* **6**: 24.
- Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, Wong MC, Maddren M, Fang R, Heitner SG, et al. 2013. ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res* **41**: D56–D63.
- Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. 2012. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* 7: e30733.
- Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. 2013. Cell-type specific features of circular RNA expression. PLoS Genet 9: e1003777.
- Vivancos AP, Guell M, Dohm JC, Serrano L, Himmelbauer H. 2010. Strand-specific deep sequencing of the transcriptome. *Genome Res* 20: 989–999.
- Zhang Y, Zhang XO, Chen T, Xiang JF, Yin QF, Xing YH, Zhu S, Yang L, Chen LL. 2013. Circular intronic long noncoding RNAs. *Mol Cell* **51**: 792–806.



circBase: a database for circular RNAs

Petar Glazar, Panagiotis Papavasileiou and Nikolaus Rajewsky

RNA 2014 20: 1666-1670 originally published online September 18, 2014 Access the most recent version at doi:10.1261/rna.043687.113

References	This article cites 22 articles, 4 of which can be accessed free at: http://rnajournal.cshlp.org/content/20/11/1666.full.html#ref-list-1
Creative Commons License	This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see http://rnajournal.cshlp.org/site/misc/terms.xhtml). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/ .
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.



To subscribe to RNA go to: http://rnajournal.cshlp.org/subscriptions