

Dartmouth College

Dartmouth Digital Commons

Dartmouth Scholarship

Faculty Work

10-7-2016

CircNet: a database of circular RNAs derived from transcriptome sequencing data

Yu-Chen Liu

National Chiao Tung University, HsinChu

Jian-Rong Li

National Chung Hsing University

Chuan-Hu Sun

Dartmouth College

Erik Andrews

Dartmouth College

Rou-Fang Chao

National Chung Hsing University

See next page for additional authors

Follow this and additional works at: <https://digitalcommons.dartmouth.edu/facoa>



Part of the [Genetics and Genomics Commons](#)

Dartmouth Digital Commons Citation

Liu, Yu-Chen; Li, Jian-Rong; Sun, Chuan-Hu; Andrews, Erik; Chao, Rou-Fang; Lin, Feng-Mao; Weng, Shun-Long; Hsu, Sheng-Da; Huang, Chieh-Chen; Cheng, Chao; Liu, Chun-Chi; and Huang, Hsien-Da, "CircNet: a database of circular RNAs derived from transcriptome sequencing data" (2016). *Dartmouth Scholarship*. 3836.

<https://digitalcommons.dartmouth.edu/facoa/3836>

This Article is brought to you for free and open access by the Faculty Work at Dartmouth Digital Commons. It has been accepted for inclusion in Dartmouth Scholarship by an authorized administrator of Dartmouth Digital Commons. For more information, please contact dartmouthdigitalcommons@groups.dartmouth.edu.

Authors

Yu-Chen Liu, Jian-Rong Li, Chuan-Hu Sun, Erik Andrews, Rou-Fang Chao, Feng-Mao Lin, Shun-Long Weng, Sheng-Da Hsu, Chieh-Chen Huang, Chao Cheng, Chun-Chi Liu, and Hsien-Da Huang

CircNet: a database of circular RNAs derived from transcriptome sequencing data

Yu-Chen Liu¹, Jian-Rong Li², Chuan-Hu Sun², Erik Andrews⁴, Rou-Fang Chao², Feng-Mao Lin¹, Shun-Long Weng^{5,6,7}, Sheng-Da Hsu¹, Chieh-Chen Huang³, Chao Cheng^{4,8,9}, Chun-Chi Liu^{2,*} and Hsien-Da Huang^{1,10,11,12,*}

¹Institute of Bioinformatics and Systems Biology, National Chiao Tung University, HsinChu, Taiwan, ²Institute of Genomics and Bioinformatics, National Chung Hsing University, Taiwan, ³Department of Life Sciences, National Chung Hsing University, Taichung, Taiwan, ⁴Institute for Quantitative Biomedical Sciences, Geisel School of Medicine at Dartmouth, Lebanon, NH 03766, USA, ⁵Department of Obstetrics and Gynecology, Hsinchu Mackay Memorial Hospital, Hsinchu, Taiwan, ⁶Mackay Medicine, Nursing and Management College, Taipei, Taiwan, ⁷Department of Medicine, Mackay Medical College, New Taipei City, Taiwan, ⁸Department of Genetics, Geisel School of Medicine at Dartmouth, Hanover, NH 03755, USA, ⁹Norris Cotton Cancer Center, Geisel School of Medicine at Dartmouth, Lebanon, NH 03766, USA, ¹⁰Department of Biological Science and Technology, National Chiao Tung University, HsinChu, Taiwan, ¹¹Center for Bioinformatics Research, National Chiao Tung University, Hsinchu 300, Taiwan and ¹²Department of Biomedical Science and Environmental Biology, Kaohsiung Medical University, Kaohsiung, Taiwan

Received August 14, 2015; Accepted September 08, 2015

ABSTRACT

Circular RNAs (circRNAs) represent a new type of regulatory noncoding RNA that only recently has been identified and cataloged. Emerging evidence indicates that circRNAs exert a new layer of post-transcriptional regulation of gene expression. In this study, we utilized transcriptome sequencing datasets to systematically identify the expression of circRNAs (including known and newly identified ones by our pipeline) in 464 RNA-seq samples, and then constructed the CircNet database (<http://circnet.mbc.nctu.edu.tw/>) that provides the following resources: (i) novel circRNAs, (ii) integrated miRNA-target networks, (iii) expression profiles of circRNA isoforms, (iv) genomic annotations of circRNA isoforms (e.g. 282 948 exon positions), and (v) sequences of circRNA isoforms. The CircNet database is to our knowledge the first public database that provides tissue-specific circRNA expression profiles and circRNA–miRNA-gene regulatory networks. It not only extends the most up to date catalog of circRNAs but also provides a thorough expression analysis of both previously reported and novel circRNAs. Furthermore, it generates an integrated regulatory network that illustrates the regulation between circRNAs, miRNAs and genes.

INTRODUCTION

Circular RNAs (circRNAs) are a family of RNAs whose head 3' and tail 5' ends covalently bond together to result in a circular form. Identified more than thirty years ago, circRNAs were first hypothesized to be relics of pre-cellular evolution (1) with the phenomenon of the 3' and 5' ends covalently joining called 'exon scramble' (2) or 'back-spliced junctions'. In plant cells, circRNAs were known to be pathogenic and deemed viroid (3,4). In contrast, circRNAs found expressed in mammal cells (5–8) were generally considered to be rare products of splicing error.

In recent years, high throughput sequencing technology has dramatically expanded the scope of transcriptomics research (9) and made it possible to more accurately investigate the expression of non-coding RNA genes. In 2012, Salzman *et al.* (10,11) developed an algorithm to detect scrambled exons in RNA-seq datasets, and reported that circular RNA isoforms are actually predominant in many human gene isoforms. By 2013, Memczak *et al.* (12) developed an improved version of the algorithm and found that some circRNAs serve as natural microRNA 'sponges'. The reliability of the back-splicing junction detection algorithm can be validated through a protocol developed by Jeck *et al.* (13), which includes examining RNA-seq samples treated with RNase R (14). For instance, a circRNA named CDR1as (15) expressed in human and mouse brain was shown to negatively regulate miR-7 in a post-transcriptional manner (12,15,16); this mechanism appears to be evolutionarily conserved (12). Extended identifica-

*To whom correspondence should be addressed. Tel: +886 3 5712121 (Ext 56952); Email: bryan@mail.nctu.edu.tw
Correspondence may also be addressed to Chun-Chi Liu. Tel: +886 4 22840338 (Ext 7031); Email: jimliu@nchu.edu.tw

tion of circRNAs in mouse (17), fly (18) and other animals (19) suggests that circRNA ubiquity is evolutionally-conserved. Further evidence indicates that human circRNA expression exhibits tissue specificity, and now tens of thousands of circRNAs have been found and reported across human tissues (17,20–25). CircRNAs functional correlation with brain function, synaptic plasticity (26,27) and fetal development (28) has recently been reported. The fact that cell free circRNAs were found stable in saliva (29) as well as exosomes (30) makes circRNA a promising diagnosis biomarker.

Given the emerging understanding of the biological importance of circRNAs and research efforts to understand them, we constructed a database called CircNet to extend the catalog of reported circRNAs and provide resources to aid in their study. Previously reported and newly identified human circRNAs are cataloged, with CircRNA expression metadata in the form of a heatmap illustrating circRNA expression profiles across 464 human transcriptome samples additionally provided. Embracing the idea that circRNAs are enriched with conserved miRNA binding sites (31) and function as natural miRNA sponges (12), CircNet maps circRNA–miRNA–mRNA interactions into regulatory networks. An integrated genome browser illustrates the position of circRNAs in the human genome, and relative locations of repeat sequences are presented to address the issue of circRNA biogenesis as the biogenesis of circRNAs correlates to repeats in the flanking area (13,32). A provided search function enables the exploration of CircNet. Users can choose an interested gene or miRNA to access its associated network, expression profile, and genome position simultaneously in CircNet. Overall, CircNet provides interactive tools for users to easily access comprehensive information regarding expression profiles across many conditions, genome loci, close repeat sequences, post-transcription regulation networks, and references to previous studies for circRNAs.

DATA COLLECTION AND DATABASE CONTENT

CircNet's data collection, cataloging and analysis are summarized in Figure 1. The process started from identifying back-spliced junction sites in the collected datasets to find a total of 212 950 circRNAs, including 53 687 novel RNAs. From them, we defined 34 000 high-confidence circRNAs by selecting back-spliced junction sites with supporting experiments ≥ 3 . To associate circRNA isoforms with miRNAs, mature human miRNA sequences and experimentally verified miRNA target genes were acquired from miR-Base (33) and miRTarBase (34). Perfect alignments of 6mer, 7mer-A1, 7mer-m8 and 8mer lengths between miRNA target seeds and circRNA isoforms were collected in CircNet, where 6mer represents the match of position 2 to 7; 7mer-A1 represents a 6mer alignment plus one A in position 1; 7mer-m8 represents a matched from position 2 to 8; and 8mer represents match of all 1 to 8 position (35). Appearance frequency of these miRNA target sites were normalized by isoform length. From the distribution of this frequency across all transcripts, including linear isoforms of the genes, only those isoforms with a *P* value less than 0.001 are reported to sponge miRNA in CircNet.

Data collection

Reported human back-spliced junction sites were also collected from the supplemental information provided in other work (10–13,20–25,27). In addition, 464 RNA-seq samples were collected from a wide range of independent experiments across 26 human tissues and 104 disease conditions from the NCBI Sequence Read Archive (36). These samples cover research topics ranging from detecting gene variation (37–41) to searching for long non-coding RNAs (42). Datasets used in recent publications (17,20–25) regarding circRNA identification were also included. CircRNA back-spliced junction sites were identified within these samples. To acquire the expression patterns of circRNAs within these samples, we designed a pipeline to perform expression profiling.

CircRNA identification

CircRNA back-spliced junction sites in the collected datasets were identified using the algorithm proposed by Memczak *et al.* (16), which has also been used by the circBase database to detect back-spliced junction sites (43). As shown in Figure 1, reads spanning junction sites that cannot be fully aligned onto the reference genome were analyzed to identify back-spliced junction sites. Junction sites meeting the same threshold criteria in Memczak *et al.* (16) were cataloged in CircNet.

To acquire the full length nucleotide sequence from RNA-seq reads, back-spliced junction sites were compared with the hg19 human genome annotation as obtained from UCSC genome browser (44). Given recent research, we recognized that multiple circRNA isoforms might originate from the same back-spliced junction site (11,26). Based on the method proposed by Salzman *et al.* (11), we identified multiple circRNA isoforms with annotated transcripts if the 'head' and 'tail' positions of a detected back-spliced junction were located exactly at the position of the exon junction sites belonging to the same annotated transcripts. Since back-spliced junction site flanking exons could potentially be composed of multiple isoforms, we included all the possible isoforms in the annotation for the expression analysis and miRNA target search. In addition, we found many reported and novel back-spliced junction sites that did not align exactly to well-annotated exon locations. For those circRNAs associated with these back-spliced junction sites having small misalignments to exon locations, flanked exons and a small portion of intron sequence were considered as parts of the isoforms. Still other junction sites were found to be located in intergenic positions while others, despite overlapping with certain genes, localized to their antisense strands. For these circRNAs with no apparent flanked exons, we took the entire flanked sequence between the 'head' and 'tail' position of the back-spliced junction sites for expression analysis and miRNA target search. Given the possibility of flanking intron areas as part of the circRNAs as described in Gao *et al.* (20), we took the full flanked sequences into consideration for expression and miRNA target search. The process is illustrated in Supplementary Figure S1. Expression profiling was performed through integrating the tool Cufflinks (45,46) with STAR (47). Expression level of the isoforms can also be estimated through

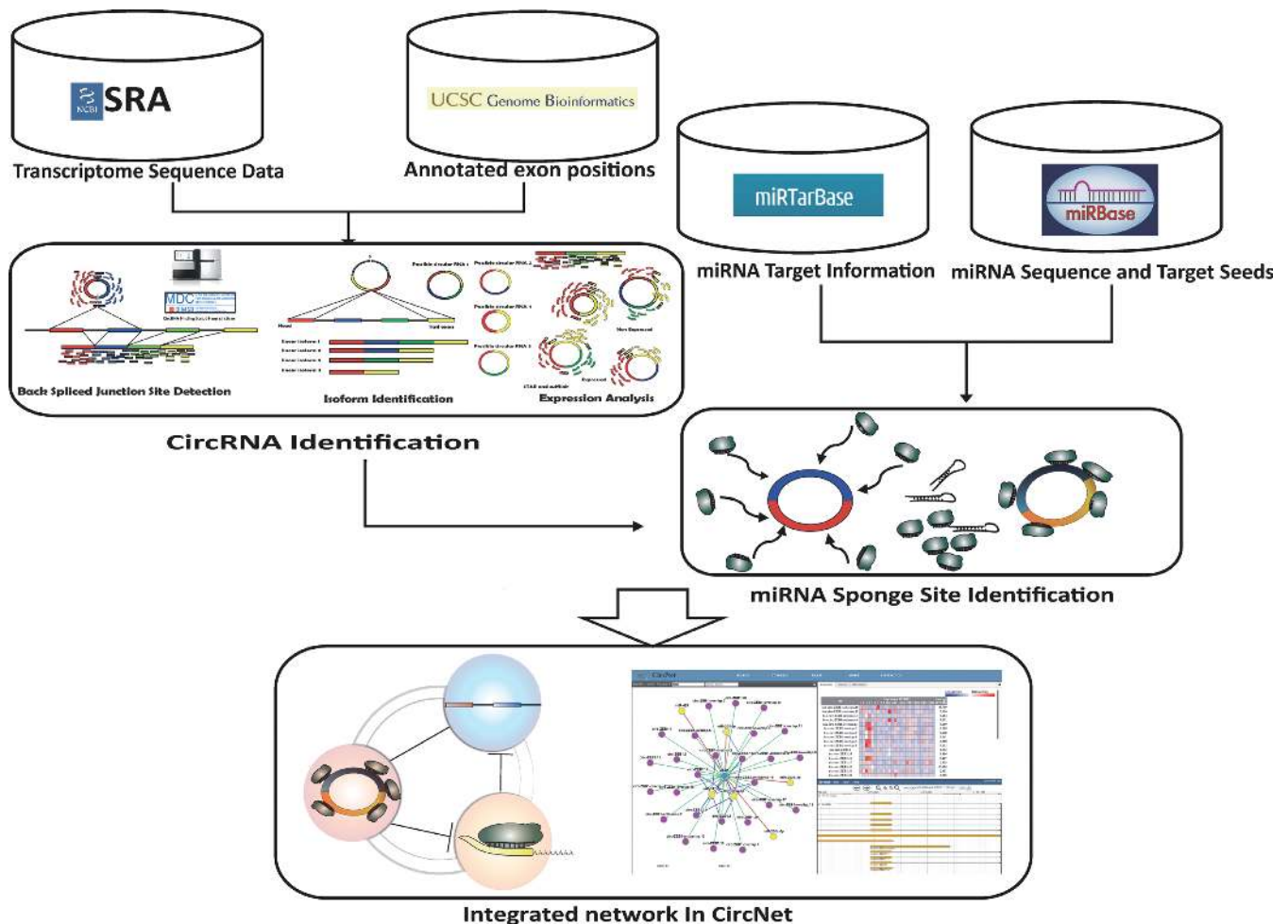


Figure 1. Framework of the database construction in CircNet. The graph illustrates how the network in CircNet was constructed. For circRNA identification, transcriptome sequencing data sets were obtained from the NCBI Sequence Read Archive (SRA). The back-spliced junction sites in each RNA-seq sample were identified using a circRNA discovery pipeline we developed (scripts provided on circBase). Detected back-spliced junction sites were further compared with hg19 human genome annotation to define circRNA isoforms. Only isoforms with an average expression level (those with an average FPKM in the 464 examined samples greater than 0.01) were included in the database. To predict circRNA–miRNA interactions, the occurrence of miRNA target seeds in circRNA isoforms were examined and normalized by isoform length. The significance of interactions was evaluated by referring to the background distribution of miRNA seeds in all transcripts and only circRNA–miRNA interactions with P -values < 0.001 were reported.

FPKM (Fragments Per Kilobase of exon per Million fragments mapped) values during the process.

The newly identified circRNAs were named by following the nomenclature suggested by Jeck *et al.* (48). The ID of a circRNA specifies the gene symbol of the source gene. If the back-spliced junction site is not exactly aligned with annotated exons, an ‘overlap’ tag is included in the ID. Similar tags suggesting that the circRNA is antisense to the origin gene are included in the ID for the applicable circRNAs.

MicroRNA sponge detection

Potential miRNA binding sites on circRNAs were identified through iteratively searching the circRNA isoform sequences for miRNA target sequences deemed typical: 6mer, 7mer-A1, 7mer-m8 and 8mer sequences (35). Perfect complementarity was required of these sequences for a circRNA–miRNA target relationship to be identified. To normalize the number of occurrences of these sites, the be-

low formula was used:

$$\text{Frequency of Nmer} = \frac{\text{Number of target seeds} \times 1000}{N \times \text{Length of CircRNA}}$$

With this formula, four frequency numbers can be acquired from each pair of circRNA and miRNA. To distinguish circRNA from linear isoforms, frequency values were also calculated for linear mRNA and miRNA pairs. P -values for each circRNA and miRNA relationship were acquired by one minus the calculated the culminated distribution function of the frequency Z score. The circRNA–miRNA pair with P -value < 0.001 represents high regulatory potential between the circRNA and miRNA. Information of these regulatory relationships was combined with target genes of miRNA from miRTarBase (34) to form the networks demonstrated on CircNet.

WEB INTERFACE

The web interface of CircNet is summarized in Figure 2.

(i) Reference sequence: the Hg19 human genome sequence, (ii) HG19_repeats: any repeat sequences on Hg19 human genome sequence for the searched subject; and (iii) CircRNAs: all circRNAs as well as overlapped linear transcripts around the selected area.

Sequence search is also available on CircNet (Figure 2C). Through integrated BLAST, users can search for all circRNAs that align with input nucleotide sequences. Back-spliced junction sites overlapped with only intergenic regions can also be accessed through this function.

EXAMPLE APPLICATIONS

We use the ZEB1-derived network as an example to illustrate the information provided by CircNet (Supplementary Figure S2). When ZEB1 is inputted as the keyword, CircNet collects available information about miRNAs targeting ZEB1, circRNAs originating from the ZEB1 sequence, and the regulatory relationship between the circRNAs and miRNAs to create an integrated mRNA-miRNA-circRNA regulatory network. The network suggests that the circRNAs circ-ZEB1.5, circ-ZEB1.19, circ-ZEB1.17 and circ-ZEB1.33 have the potential to sponge the miRNA miR-200a-3p. Since miR200a has been reported to target ZEB1 (34), this suggests the existence of a negative feedback control loop. The integrated genome browser provides the visualization of these circRNAs: all these circRNAs contain the exon in position chr10:31791276–31791437, which is enriched with miR-200a targeting seeds. From the flanking area of this specific exon, we found the flanking MIR repeat sequences MIR_dup1543 and MIR3_dup768 and MIR3_dup769, providing a lead to the biogenesis of the circRNA. As illustrated in Supplementary Figure S3, expression patterns provided in CircNet indicate that all four identified circRNAs — circ-ZEB1.5, circ-ZEB1.19, circ-ZEB1.17 and circ-ZEB1.33 — were up-regulated in available normal lung tissue samples compared to the lung cancer samples.

DISCUSSION AND FUTURE DEVELOPMENT

Regarding other circRNA databases, Glazar *et al.* constructed circBase (<http://circbase.org/>) (43), and Li *et al.* constructed starBase (<http://starbase.sysu.edu.cn/>) (49). Supplementary Table S1 lists the comparison between CircNet, starBase and circBase and it suggests some strong advantages of CircNet in comparison to these other works. One of the advanced features that makes CircNet distinctive is the novel design of circRNA naming in the catalog. Emerging evidence suggests that multiple circular isoforms originate from the same back splice junction site (11,26). Given this, we catalog circRNAs based on distinctive isoforms instead of regarding one observed back-spliced junction position as a single circRNA. Following the Jeck *et al.* suggestions in 2014 Nature Reviews (48), we designed a circRNA naming system which provides information regarding source genes, whether the circRNA is antisense or intronic, and whether the back splice junction site is exactly located on well annotated exons. This annotation and incorporation into an integrated genome browser should provide a clear method for future curation Expression profile and

accessibility to the examined sample condition provided in CircNet would be a useful tool to study circRNA tissue specific function as well as correlation to disease. One observation worth mention is the potential wide existence of double negative feedback control loops in the network. As seen in the ZEB1-derived network example, the circRNA originating from within ZEB1 was putatively found to sponge a miRNA targeting ZEB1. The double negative control relationship on ZEB1 gene and mir-200 family was previously reported by Bracken *et al.* in 2008 (50). The correlation between mir-200 ZEB1 regulation and lung adenocarcinoma was recently reported by Yang *et al.* (51). Data provided in CircNet indicates that circRNAs spanning the chr10:31791276–31791437 exon of ZEB1 play significant roles in this regulation. With the newly discovered ZEB1 circRNAs enriched with mir-200 targeting seeds, a hypothetical double negative feedback control loop of circRNA is demonstrated on the default network on homepage of CircNet. Further inquiries using the regulatory networks identified using CircNet may discover additional novel feedback loops with applicability to human disease.

AVAILABILITY

The CircNet database will be continuously maintained and updated. The database is now publicly accessible at <http://circnet.mbc.nctu.edu.tw/>.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank the Ministry of Science and Technology of the Republic of China for financially supporting this research. We also thank the UST-UCSD International Center of Excellence in Advanced Bio-engineering sponsored by the Taiwan Ministry of Science and Technology I-RiCE Program, Veterans General Hospitals and University System of Taiwan (VGHUST) Joint Research Program, and MOE ATU. We would like to thank The Cancer Genome Atlas for use of data.

FUNDING

Ministry of Science and Technology of the Republic of China [MOST 101-2311-B-009-005-MY3, MOST 103-2628-B-009-001-MY3, MOST 103-2320-B-005-004, MOST 104-2621-M-005-005-MY3 and MOST 104-2320-B-005-005]; UST-UCSD International Center of Excellence in Advanced Bioengineering sponsored by the Ministry of Science and Technology I-RiCE Program [MOST 103-2911-I-009-101]; Veterans General Hospitals and University System of Taiwan (VGHUST) Joint Research Program [VGHUST103-G5-11-2]. MOE ATU. Funding for open access charge: Ministry of Science and Technology of the Republic of China, Taiwan [103-2628-B-009-001-MY3]. This work was supported by a research grant from Health and welfare surcharge of tobacco products, Ministry of Health and Welfare (MOHW103-TD-B-111-08; MOHW104-TDU-B-212-124-005).

Conflict of interest statement. None declared.

REFERENCES

- Diener, T.O. (1989) Circular RNAs: relics of precellular evolution? *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 9370–9374.
- Nigro, J.M., Cho, K.R., Fearon, E.R., Kern, S.E., Ruppert, J.M., Oliner, J.D., Kinzler, K.W. and Vogelstein, B. (1991) Scrambled exons. *Cell*, **64**, 607–613.
- Diener, T.O. (1979) Viroids: structure and function. *Science*, **205**, 859–866.
- Sanger, H.L., Klotz, G., Riesner, D., Gross, H.J. and Kleinschmidt, A.K. (1976) Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc. Natl. Acad. Sci. U.S.A.*, **73**, 3852–3856.
- Capel, B., Swain, A., Nicolis, S., Hacker, A., Walter, M., Koopman, P., Goodfellow, P. and Lovellbadge, R. (1993) Circular transcripts of the testis-determining gene *Sry* in adult-mouse testis. *Cell*, **73**, 1019–1030.
- Pasman, Z., Been, M.D. and Garcia-Blanco, M.A. (1996) Exon circularization in mammalian nuclear extracts. *Rna*, **2**, 603–610.
- Zaphiropoulos, P.G. (1996) Circular RNAs from transcripts of the rat cytochrome P450 2C24 gene: correlation with exon skipping. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 6536–6541.
- Chao, C.W., Chan, D.C., Kuo, A. and Leder, P. (1998) The mouse formin (*Fmn*) gene: abundant circular RNA transcripts and gene-targeted deletion analysis. *Mol. Med.*, **4**, 614–628.
- Wang, Z., Gerstein, M. and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, **10**, 57–63.
- Salzman, J., Gawad, C., Wang, P.L., Lacayo, N. and Brown, P.O. (2012) Circular RNAs Are the Predominant Transcript Isoform from Hundreds of Human Genes in Diverse Cell Types. *PLoS One*, **7**, e30733.
- Salzman, J., Chen, R.E., Olsen, M.N., Wang, P.L. and Brown, P.O. (2013) Cell-type specific features of circular RNA expression. *PLoS Genet.*, **9**, e1003777.
- Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M. et al. (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, **495**, 333–338.
- Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., Marzluff, W.F. and Sharpless, N.E. (2013) Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA*, **19**, 141–157.
- Suzuki, H., Zuo, Y., Wang, J., Zhang, M.Q., Malhotra, A. and Mayeda, A. (2006) Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. *Nucleic Acids Res.*, **34**, e63–e63.
- Hansen, T.B., Wiklund, E.D., Bramsen, J.B., Villadsen, S.B., Statham, A.L., Clark, S.J. and Kjems, J. (2011) miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J.*, **30**, 4414–4422.
- Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K. and Kjems, J. (2013) Natural RNA circles function as efficient microRNA sponges. *Nature*, **495**, 384–388.
- Ashwal-Fluss, R., Meyer, M., Pamudurti, N.R., Ivanov, A., Bartok, O., Hanan, M., Evantal, N., Memczak, S., Rajewsky, N. and Kadener, S. (2014) circRNA biogenesis competes with pre-mRNA splicing. *Mol. Cell*, **56**, 55–66.
- Westholm, J.O., Miura, P., Olson, S., Shenker, S., Joseph, B., Sanfilippo, P., Celniker, S.E., Graveley, B.R. and Lai, E.C. (2014) Genome-wide analysis of *Drosophila* circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. *Cell Rep.*, **9**, 1966–1980.
- Wang, P.L., Bao, Y., Yee, M.-C., Barrett, S.P., Hogan, G.J., Olsen, M.N., Dinneny, J.R., Brown, P.O. and Salzman, J. (2014) Circular RNA is expressed across the eukaryotic tree of life. *PLoS One*, **9**, e90859.
- Gao, Y., Wang, J. and Zhao, F. (2015) CIRI: An efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biol.*, **16**, 4.
- Kelly, S., Greenman, C., Cook, P.R. and Papanonis, A. (2015) Exon skipping is correlated with exon circularization. *J. Mol. Biol.*, **427**, 2414–2417.
- Bachmayr-Heyda, A., Reiner, A.T., Auer, K., Sukhbaatar, N., Aust, S., Bachleitner-Hofmann, T., Mesteri, I., Grunt, T.W., Zeillinger, R. and Pils, D. (2015) Correlation of circular RNA abundance with proliferation-exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis, and normal human tissues. *Sci. Rep.*, **5**, 8057.
- Zhang, X.-O., Wang, H.-B., Zhang, Y., Lu, X., Chen, L.-L. and Yang, L. (2014) Complementary sequence-mediated exon circularization. *Cell*, **159**, 134–147.
- Guo, J.U., Agarwal, V., Guo, H. and Bartel, D.P. (2014) Expanded identification and characterization of mammalian circular RNAs. *Genome Biol.*, **15**, 409.
- Caiment, F., Gaj, S., Claessen, S. and Kleinjans, J. (2015) High-throughput data integration of RNA–miRNA–circRNA reveals novel insights into mechanisms of benzo [a] pyrene-induced carcinogenicity. *Nucleic Acids Res.*, **43**, 2525–2534.
- You, X., Vlatkovic, I., Babic, A., Will, T., Epstein, I., Tushev, G., Akbalik, G., Wang, M., Glock, C. and Quedenau, C. (2015) Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nat. Neurosci.*, **18**, 603–610.
- Rybak-Wolf, A., Stottmeister, C., Glazár, P., Jens, M., Pino, N., Giusti, S., Hanan, M., Behm, M., Bartok, O. and Ashwal-Fluss, R. (2015) Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. *Mol. Cell*, **58**, 870–885.
- Szabo, L., Morey, R., Palpant, N.J., Wang, P.L., Afari, N., Jiang, C., Parast, M.M., Murry, C.E., Laurent, L.C. and Salzman, J. (2015) Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human fetal development. *Genome Biol.*, **16**, 126–126.
- Bahn, J.H., Zhang, Q., Li, F., Chan, T.-M., Lin, X., Kim, Y., Wong, D.T. and Xiao, X. (2015) The landscape of microRNA, Piwi-interacting RNA, and circular RNA in human saliva. *Clin. Chem.*, **61**, 221–230.
- Li, Y., Zheng, Q., Bao, C., Li, S., Guo, W., Zhao, J., Chen, D., Gu, J., He, X. and Huang, S. (2015) Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. *Cell Res.*, **25**, 981–984.
- Thomas, L.F. and Sætrom, P. (2014) Circular RNAs are depleted of polymorphisms at microRNA binding sites. *Bioinformatics*, **30**, 2243–2246.
- Ivanov, A., Memczak, S., Wyler, E., Torti, F., Porath, H.T., Orejuela, M.R., Piechotta, M., Levanon, E.Y., Landthaler, M. and Dieterich, C. (2015) Analysis of intron splicing reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep.*, **10**, 170–177.
- Griffiths-Jones, S., Grocock, R.J., Van Dongen, S., Bateman, A. and Enright, A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–D144.
- Hsu, S.-D., Tseng, Y.-T., Shrestha, S., Lin, Y.-L., Khaleel, A., Chou, C.-H., Chu, C.-F., Huang, H.-Y., Lin, C.-M. and Ho, S.-Y. (2014) miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Res.*, **42**, D78–D85.
- Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
- Leinonen, R., Sugawara, H. and Shumway, M. (2010) The sequence read archive. *Nucleic Acids Res.*, **39**, D19–D21.
- Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A. and Kircher, M. (2011) The evolution of gene expression levels in mammalian organs. *Nature*, **478**, 343–348.
- Brennan, E.P., Morine, M.J., Walsh, D.W., Roxburgh, S.A., Lindenmeyer, M.T., Brazil, D.P., Gaora, P.O., Roche, H.M., Sadlier, D.M. and Cohen, C.D. (2012) Next-generation sequencing identifies TGF- β 1-associated gene expression profiles in renal epithelial cells reiterated in human diabetic nephropathy. *Biochim. Biophys. Acta (BBA)-Mol. Basis Dis.*, **1822**, 589–599.
- Lovén, J., Orlando, D.A., Sigova, A.A., Lin, C.Y., Rahl, P.B., Burge, C.B., Levens, D.L., Lee, T.I. and Young, R.A. (2012) Revisiting global gene expression analysis. *Cell*, **151**, 476–482.
- Ngondo-Mbongo, R.P., Myslinski, E., Aster, J.C. and Carbon, P. (2013) Modulation of gene expression via overlapping binding sites exerted by ZNF143, Notch1 and THAP11. *Nucleic Acids Res.*, **41**, 4000–4014.
- Pickrell, J.K., Marioni, J.C., Pai, A.A., Degner, J.F., Engelhardt, B.E., Nkadori, E., Veyrieras, J.-B., Stephens, M., Gilad, Y. and Pritchard, J.K. (2010) Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*, **464**, 768–772.

42. Derrien,T., Johnson,R., Bussotti,G., Tanzer,A., Djebali,S., Tilgner,H., Guernec,G., Martin,D., Merkel,A. and Knowles,D.G. (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.*, **22**, 1775–1789.
43. Glažar,P., Papavasileiou,P. and Rajewsky,N. (2014) circBase: a database for circular RNAs. *RNA*, **20**, 1666–1670.
44. Karolchik,D., Barber,G. P., Casper,J., Clawson,H., Cline,M. S., Diekhans,M. *et al.* (2014) The UCSC Genome Browserdatabase: 2014 update. *Nucleic Acids Res.*, **42**, D764–D770.
45. Trapnell,C., Williams,B.A., Pertea,G., Mortazavi,A., Kwan,G., van Baren,M.J., Salzberg,S.L., Wold,B.J. and Pachter,L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, **28**, 511–515.
46. Roberts,A., Pimentel,H., Trapnell,C. and Pachter,L. (2011) Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics*, **27**, 2325–2329.
47. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
48. Jeck,W.R. and Sharpless,N.E. (2014) Detecting and characterizing circular RNAs. *Nat. Biotechnol.*, **32**, 453–461.
49. Yang,J.-H., Li,J.-H., Shao,P., Zhou,H., Chen,Y.-Q. and Qu,L.-H. (2011) starBase: a database for exploring microRNA–mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Res.*, **39**, D202–D209.
50. Bracken,C.P., Gregory,P.A., Kolesnikoff,N., Bert,A.G., Wang,J., Shannon,M.F. and Goodall,G.J. (2008) A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition. *Cancer Res.*, **68**, 7846–7854.
51. Yang,Y., Ahn,Y.-H., Chen,Y., Tan,X., Guo,L., Gibbons,D.L., Ungewiss,C., Peng,D.H., Liu,X. and Lin,S.H. (2014) ZEB1 sensitizes lung adenocarcinoma to metastasis suppression by PI3K antagonism. *J. Clin. Invest.*, **124**, 2696.