www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2021; 10(11): 939-942

www.thepharmajournal.com
Received: 02-08-2021
Accepted: 13-09-2021

Waseem Akram Malla
Division of Veterinary Biotechnology, ICAR-Indian Veterinary Research Institute, Izatnagar, Uttar Pradesh, India

Richa Arora
Division of Veterinary Biotechnology, ICAR-Indian Veterinary Research Institute, Izatnagar, Uttar Pradesh, India

Niha Ayman
Faculty of Veterinary Sciences and Animal Husbandry, Sher-e-Kashmir University of Agricultural Sciences and Technology-Kashmir, Jammu and Kashmir, India

Sheikh Firdous Ahmad
Division of Animal Genetics, ICAR-Indian Veterinary Research Institute, Izatnagar, Uttar Pradesh, India

Sajad Ahmad Beigh
Faculty of Veterinary Sciences and Animal Husbandry, Sher-e-Kashmir University of Agricultural Sciences and Technology-Kashmir, Jammu and Kashmir, India

Sonalika Mahajan
Division of Veterinary Biotechnology, ICAR-Indian Veterinary Research Institute, Izatnagar, Uttar Pradesh, India

Basavaraj Sajjanar
Division of Veterinary Biotechnology, ICAR-Indian Veterinary Research Institute, Izatnagar, Uttar Pradesh, India

Ravikumar Gandham
Division of Veterinary Biotechnology, ICAR-Indian Veterinary Research Institute, Izatnagar, Uttar Pradesh, India

Ashok Kumar Tiwari
ICAR-Central Avian Research Institute, Izatnagar Bareilly, Uttar Pradesh, India

Corresponding Author:
Ashok Kumar Tiwari
ICAR-Central Avian Research Institute, Izatnagar Bareilly, Uttar Pradesh, India

Circular RNA prediction from transcriptomic data of 4T1 mouse mammary tumor cell line

Waseem Akram Malla, Richa Arora, Niha Ayman, Sheikh Firdous Ahmad, Sajad Ahmad Beigh, Sonalika Mahajan, Basavaraj Sajjanar, Ravikumar Gandham and Ashok Kumar Tiwari

Abstract
Circular RNAs (circRNAs) are non-canonical forms of RNA that lack a poly-A tail, are highly stable, and form circular structures due to covalently linked 3' and 5' ends. 4T1 cell line is a mouse mammary tumor cell line used as a model for studying triple-negative breast cancer. RNASeq datasets from 4T1 mouse mammary tumor cell line were obtained from the ENA database and analyzed for the expression of circRNAs. The CIRI-full pipeline was used for circRNA prediction and full structure reconstruction. Three circRNAs (1:173740380|173810307, 1:173698944| 173733197, and 10:22161590|22183197) were found to be highly expressed, out of which one was found to be expressed in all datasets. Several other circRNAs detected in the study were expressed at low levels. Analysis of the host genes of these circRNAs revealed that the host genes are involved in several cellular processes that regulate cancer cell proliferation, enhance their survival and affect apoptosis.

Keywords: circRNA, breast cancer, MALAT1, biomarkers, circular RNA, non-coding RNA

1. Introduction
Circular RNAs (circRNAs) have been recently discovered as non-canonical forms of RNA in the cells which differ from mRNA as they lack a poly-A tail, are highly stable and resistant to exonuclease digestion, and form circular structures due to a phenomenon called back-splicing that covalently links their 3' and 5' ends [1, 2]. They are expressed in a tissue- and cell-type-specific manner, playing a role in the normal physiology as well as in diseases like cancer, by regulating cell proliferation, cellular differentiation, pluripotency, and apoptosis [1, 3-6].
4T1 cell line is a mouse mammary tumor cell line that resembles triple-negative breast cancer and is thus used as a model system for breast cancer research, from studying the effects of drugs to biomarker discovery for cancer diagnosis and prognostic evaluation. We hypothesized that the 4T1 cell line being a cancer cell line could express circRNAs that have a role in cancer biology. In this study, publicly available transcriptome data of 4T1 cells were analyzed to gain insights into the expression of circRNAs and the function of their host genes.

2. Materials and Methods
2.1 RNASeq datasets
A total of 3 publicly available Next Seq 500 paired-end RNA-Seq datasets of 4T1 mouse mammary tumor cell line samples were downloaded from European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/home; Accession IDs: SRR9274397 (sample 1), SRR9274398 (sample 2), SRR9274399 (sample 3)). The datasets were filtered using prinseq-lite software to remove singleton reads and those reads that were shorter than 76 nts.

2.2 circRNA detection pipeline
The filtered reads were used for circRNA detection and assembly using the CIRI-full pipeline [7], which integrates data from CIRI2 [8] and CIRI-AS [9]. The latest build of the mouse reference genome and annotation files were used and obtained from the Ensembl genome browser. Firstly, the mouse genome was indexed using the BWA index option, after which each of the paired datasets was aligned to the indexed genome using the BWA-MEM algorithm [10]. Afterward, the alignment map (SAM) file generated from each sample in the previous step was analyzed in a step-wise manner using the protocol given by Zheng and Zhao [11].

CIRI2 is capable of detecting circRNAs from large datasets of variable read length and outputs strand of predicted circRNAs. CIRI-AS detects circRNA internal components and alternative splicing events but cannot process reads of variable length and processed data need to be trimmed to equal lengths before running this algorithm. CIRI-full uses Back Splice Junction (BSJ – for circexon detection and circRNAs boundary demarcation) and reverse overlap (RO – to generate detailed landscape within boundary sites) features to reconstruct and quantify full-length circular RNAs from RNA-Seq data sets.

2.3 Visualization of detected circRNAs

The predicted circRNAs from the CIRI-full pipeline were visualized using the CIRI-vis tool [12], which renders the alignments of BSJ & RO merged reads and estimates the related abundance of isoforms according to the output of CIRI-full or CIRI-AS. All algorithms were run with recommended parameters.

3. Results

The read lengths of each dataset ranged from 36 to 76 nucleotides (nts). Since CIRI-AS cannot work with datasets having variable read lengths, reads shorter than 76 nts long were removed from all the datasets. A total of 19591148, 20899272, and 18886348 were retained for further analysis. A total of 41, 34, and 37 candidate circRNAs were found by CIRI2 in the first scanning, however, the number of circRNAs reduced to 5, 5, and 6 in samples 1, 2, and 3 respectively after the second scanning. CIRI-AS then was run to detect the alternative splicing events and a detailed list of mapping positions of BSJ reads, coverage, and circexons was obtained for all the samples which aided in the reconstruction of full-length circRNAs in the subsequent steps as follows.

CIRI-full looks for reverse overlaps (RO) in paired-end sequencing data and when ROs of sufficient quality were found, the read pairs found in the dataset were assembled into long reads and then aligned to the reference genome. Next, full-length circRNA were reconstructed from the identified BSJ boundaries and RO-merged reads. In sample 1, a total of 11 circRNA transcripts were detected, out of which 8 were full circRNAs (fully reconstructed) whereas 3 were broken (could not be fully reconstructed). Their length ranged from 60 nts to 243 nts. In sample 2, a total of 5 full and 2 broken circRNAs were detected, with lengths ranging from 75 to 139 nts. Similarly, in sample 3, a total of 6 full and 3 broken circRNA transcripts were detected with lengths ranging from 75 to 175 nucleotides. The full list of circRNAs detected and the host genes from which they arise is given in Table 1. While some of the circRNAs were expressed at comparatively high levels, most of the circRNAs were expressed at very low levels in all the samples.

Table 1: Sample-wise details of circRNAs with chromosome number, isoform length, host gene name, and expression (NA: Host gene name not available)

SRR9274397				
Circle_ID	**Chromosome**	**Length**	**Gene name**	**Counts**
17:24034793|24034855	17	63	Srrm2	1
4:88329315|88329475	4	161	Hacd4	1
1:173740380|173810307	1	75	Ifi211	27
5:105754930|105754989	5	60	Lrrc8c	1
11:106677105|106677364	11	260	Ddx5	1
1:173698944|173733197	1	175	Mndal	8
8:124153515|124153757	8	243	NA	2
7:78431148|78431259	7	112	Mrpl46	1
3:144303206|144303269	3	64	Selenof	1
5:115539391|115539531	5	141	NA	1
10:22161590|22183197	10	139	H60b	12
SRR9274398				
Circle_ID	**Chromosome**	**Length**	**Gene name**	**Counts**
10:62581147|62581249	10	103	Ccar1	1
19:5851913|5851994	19	82	Malat1	1
8:23428009|23428085	8	77	Kat6a	1
1:173740380|173810307	1	75	Ifi211	23
5:33817774|33817849	5	76	NA	1
5:135940194|135940271	5	78	Ywhag	1
10:22161590|22183197	10	139	H60b	12
SRR9274399				
Circle_ID	**Chromosome**	**Length**	**Gene name**	**Counts**
6:140370103|140372281	6	138	Plekha5	1
1:173740380|173810307	1	75	Ifi211	24
7:130586940|130587041	7	102	Htra1	1
11:84401148|84401243	11	96	Aatf	1
17:40157395|40157554	17	160	Rn18s	1
18:31766843|31769152	18	125	Sap130	2
1:173698944|173733197	1	175	Mndal	4
15:58327360|58327466	15	107	Fam91a1	1
19:23653372|23653478	19	107	Gm6563	1

4. Discussion

The analysis revealed that one circRNA 1:173740380|173810307 showed consistently high expression across all three samples. This circRNA is contained with the host gene Ifi211 (interferon activated gene 211). Ifi211 (also known as Mnda) is said to play an important role in cell

death, and its expression is critical for DNA damage response (DDR) and cell survival in embryonic stem cells [13]. Similarly, another circRNA 1:173698944|173733197 was found to be highly expressed in 2 of the samples and is contained within the host genes Mndal (also called Ifi212). Myeloid nuclear differentiation antigen-like (Mndal) protein belongs to the same family of proteins as Ifi211, that is, interferon inducible-200 (Ifi200) family associated with various diseases, including solid tumors [14]. Another highly expressed circRNA 10:22161590|22183197 that was common between two of the samples arises from the host gene H60b which belongs to the NKG2D (natural-killer group 2, member D) ligand family. Members of this family of proteins are involved in tumor immunity [15]. All the three circRNA arise from genes that have a role to play in tumor biology, which suggests that these circRNAs might regulate some of the processes involved in breast cancer progression in mammary glands in mice. The structures of the highly expressed 3 circRNAs generated with CIRI-vis are shown in Figure 1 (A-C). All three were broken circles, which means that the full sequence of these circRNAs could not be generated from the given data despite their high expression.

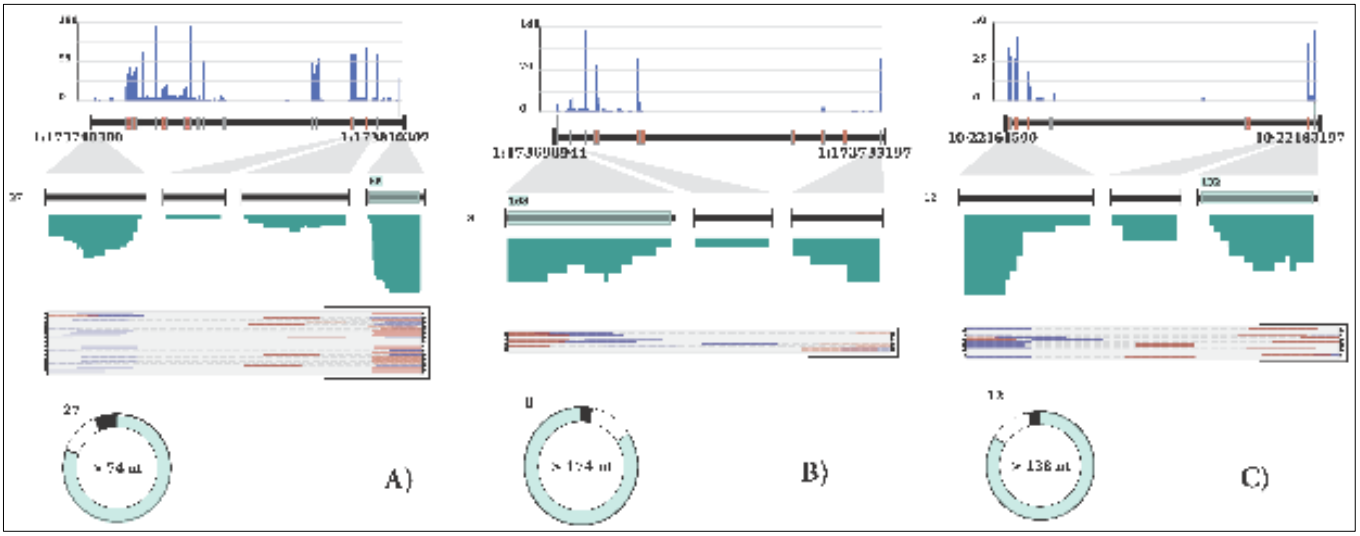

Fig 1: Circular structure of the three highly expressed circRNAs: A) 1:173740380|173810307, B) 1:173698944|173733197, and C) 10:22161590|22183197.

The topmost panel shows the RNASeq read coverage, the second panel shows BSJ read coverage and forward splice events, the third panel shows BSJ read alignments without forward splice events and the last (bottom) panel shows the reconstructed circRNA (broken circles).

Figure 2 shows the circular structure of some of the full circRNA transcripts detected in the study, however, they were expressed at a lower level as circRNAs are generally expressed at very low levels when compared to mRNAs. The host genes of other circRNAs detected in this study were also found to have a role in cancer progression, cell proliferation, and survival. Cell division cycle and apoptosis regulator protein 1 (Ccar1), from which one circRNA arises, is a β-Catenin co-activator and contributes to cancer cell proliferation and survival in gastric cancer [16]. Similarly, one more circRNA arises from MALAT1 (metastasis-associated lung adenocarcinoma transcript 1). MALAT1 is a long non-coding RNA and has been implicated in various cancers [17]. It has been previously shown that the circRNA Circ-MALAT1 arising from this lncRNA is highly expressed in cancer stem cells (CSCs) arising from hepatocellular carcinoma where it promotes cancer progression by self-renewal of CSCs [18]. Similarly, other host genes such as PLEKHA5, HtrA1, and AATF are involved in cancer progression and can act as biomarkers for cancer diagnosis [19-21].

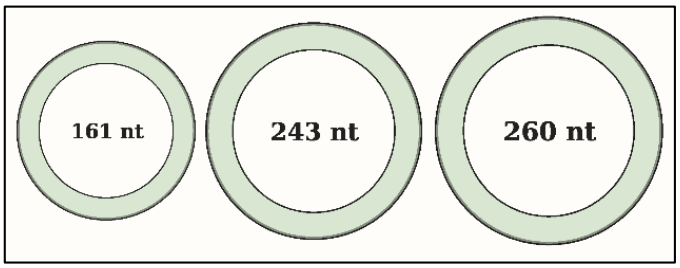

Fig 2: Circular structure of the three low-expressed circRNAs. The circRNAs are 161, 243, and 260 nucleotides long and form complete circles.

5. Conclusion

RNASeq datasets from 4T1 mouse mammary tumor cell line were obtained from the ENA database and analyzed for the expression of circRNAs. Three circRNAs were found to be highly expressed, out circRNA of which one was found to be expressed in all datasets. Several other circRNAs detected in the study were expressed at low levels. Analysis of the host genes of these circRNAs revealed that the host genes are involved in various cellular processes that regulate cancer cell proliferation, enhance their survival and affect apoptosis. These findings could further extend our understanding of circRNAs and their role in cancer and lead to the discovery of novel prognostic and diagnostic biomarkers.

6. Acknowledgment
We thank Director, IVRI for providing the facilities and support to conduct this study.

7. References

1. Greene J, Baird AM, Brady L, Lim M, Gray SG, McDermott R *et al*. Circular RNAs: Biogenesis, Function and Role in Human Diseases 2017, 4(38).
2. Yu CY, Kuo HC. The emerging roles and functions of circular RNAs and their generation. Journal of biomedical science 2019;26(1):29.
3. Barrett SP, Salzman J. Circular RNAs: analysis, expression and potential functions. Development. 2016;143(11):1838-47.
4. Chen L, Wang C, Sun H, Wang J, Liang Y, Wang Y *et al*. The bioinformatics toolbox for circRNA discovery and analysis. Briefings in Bioinformatics 2020;22(2):1706-28.
5. He JH, Han ZP, Luo JG, Jiang JW, Zhou JB, Chen WM, *et al*. Hsa_Circ_0007843 Acts as a mIR-518c-5p Sponge to Regulate the Migration and Invasion of Colon Cancer SW480 Cells 2020, 11(9).
6. Hsu MT, Coca-Prados M. Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells. Nature 1979;280(5720):339-40.
7. Zheng Y, Ji P, Chen S, Hou L, Zhao F. Reconstruction of full-length circular RNAs enables isoform-level quantification. Genome Medicine 2019;11(1):2.
8. Gao Y, Zhang J, Zhao F. Circular RNA identification based on multiple seed matching. Brief Bioinform 2018;19(5):803-10.
9. Gao Y, Wang J, Zheng Y, Zhang J, Chen S, Zhao F. Comprehensive identification of internal structure and alternative splicing events in circular RNAs. Nature Communications 2016;7(1):12060.
10. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 2009;25(14):1754-60.
11. Zheng Y, Zhao F. Detection and Reconstruction of Circular RNAs from Transcriptomic Data. Methods Mol Biol. 2018;1724:1-8.
12. Zheng Y, Zhao F. Visualization of circular RNAs and their internal splicing events from transcriptomic data. Bioinformatics 2020;36(9):2934-5.
13. Khromov T, Dressel R, Siamishi I, Nolte J, Opitz L, Engel W *et al*. Apoptosis-related gene expression profiles of mouse ESCs and maGSCs: role of Fgf4 and Mnda in pluripotent cell responses to genotoxicity. PloS one 2012;7(11):e48869.
14. Zhang K, Kagan D, DuBois W, Robinson R, Bliskovsky V, Vass WC *et al*. Mndal, a new interferon-inducible family member, is highly polymorphic, suppresses cell growth, and may modify plasmacytoma susceptibility. Blood 2009;114(14):2952-60.
15. Nausch N, Cerwenka A. NKG2D ligands in tumor immunity. Oncogene 2008;27(45):5944-58.
16. Chang TS, Wei KL, Lu CK, Chen YH, Cheng YT, Tung SY *et al*. Inhibition of CCAR1, a Coactivator of β-Catenin, Suppresses the Proliferation and Migration of Gastric Cancer Cells 2017;18(2):460.
17. Amodio N, Raimondi L, Juli G, Stamato MA, Caracciolo D, Tagliaferri P *et al*. MALAT1: a druggable long non-coding RNA for targeted anti-cancer approaches. Journal of Hematology & Oncology 2018;11(1):63.
18. Chen L, Kong R, Wu C, Wang S, Liu Z, Liu S *et al*. Circ-MALAT1 Functions as Both an mRNA Translation Brake and a microRNA Sponge to Promote Self-Renewal of Hepatocellular Cancer Stem Cells 2020;7(4):1900949.
19. Iezzi S, Fanciulli M. Discovering Che-1/AATF: a new attractive target for cancer therapy 2015;6(141).
20. Zhu F, Duan YF, Bao WY, Liu WS, Yang Y, Cai HH. HtrA1 regulates epithelial-mesenchymal transition in hepatocellular carcinoma. Biochemical and biophysical research communications 2015;467(3):589-94.
21. Nagamura Y, Miyazaki M, Nagano Y, Yuki M, Fukami K, Yanagihara K, *et al*. PLEKHA5 regulates the survival and peritoneal dissemination of diffuse-type gastric carcinoma cells with Met gene amplification. Oncogenesis 2021;10(3):25.