



METHODS ARTICLE

A genome-wide circular RNA transcriptome in rat

Disha Sharma,^{1,3} Paras Sehgal,^{2,3} Sridhar Sivasubbu ^{2,3} and
Vinod Scaria ^{1,3,*}

¹GN Ramachandran Knowledge Center for Genome Informatics, CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), Mathura Road, Delhi 110025, India, ²Genomics and Molecular Medicine, CSIR Institute of Genomics and Integrative Biology, Mathura Road (CSIR-IGIB), Delhi 110025, India and ³Academy of Scientific & Innovative Research (AcSIR), CSIR-IGIB South Campus, Mathura Road, Delhi 110025, India

*Correspondence address. GN Ramachandran Knowledge Center for Genome Informatics, CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), Mathura Road, Delhi 110025, India. Tel: +91-9650466002; E-mail: vinods@igib.in

Abstract

Circular RNAs (circRNAs) are a novel class of noncoding RNAs that back-splice from 5' donor site and 3' acceptor sites to form a circular structure. A number of circRNAs have been discovered in model organisms including human, mouse, *Drosophila*, among other organisms. There are a few candidate-based studies on circRNAs in rat, a well-studied model organism as well. A number of pipelines have been published to identify the back splice junctions for the discovery of circRNAs but studies comparing these tools have suggested that a combination of tools would be a better approach to identify high-confidence circRNAs. The availability of a recent dataset of transcriptomes encompassing 11 tissues, 4 developmental stages, and 2 genders motivated us to explore the landscape of circRNAs in the organism in this context. In order to understand the difference among different pipelines, we employed five different combinations of tools to identify circular RNAs from the dataset. We compared the results of the different combination of tools/pipelines with respect to alignment, total number of circRNAs identified and read-coverage. In addition, we identified tissue-specific, development-stage specific and gender-specific circRNAs and further independently validated 16 circRNA junctions out of 24 selected candidates in 5 tissue samples and estimated the quantitative expression of five circRNA candidates using real-time polymerase chain reaction and our analysis suggests three candidates as tissue-enriched. This study is one of the most comprehensive studies which provides a map of circRNAs transcriptome as well as to understand the difference among different computational pipelines in rat.

Keywords: RNA-Seq; circular RNA; circRNA; rat; transcriptome; noncoding RNAs

Introduction

The last decade has witnessed tremendous advances in technology, which has enabled an unprecedented opportunity to understand genomes. RNA-sequencing (RNA-Seq) is one of the approaches which have tremendously improved our depth and breadth of understanding transcripts and transcript isoforms at resolutions previously few could fathom. This, consequently, has helped in the identification of a new isoform of RNA known as cir-

cular RNAs (circRNAs), which have been rediscovered recently and found to express ubiquitously in all cellular organisms [1, 2]. These isoforms are produced by unique back-splicing and therefore could be classified as a splice isoform. Although earlier considered as splicing errors, three major studies including Memczak *et al.* (1900 transcripts from the exonic, intronic, and intergenic region) [3], Jeck *et al.* (reported identified exonuclease-treated transcripts >25 000) [3, 4] and Salzman *et al.* (identified circRNA from

Received: 22 March 2021; Revised: 11 August 2021; Editorial Decision: 12 August 2021; Accepted: 12 August 2021

© The Author(s) 2021. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

different human cell types) [5] showed the abundance of circRNAs and established it as emerging class of noncoding RNAs. Till now, circ are identified in a number of organisms including model organisms like humans [3, 6–9], worms [10], *Drosophila* [10, 11], mouse [12–14], chicken [15, 16], and pig [17–19]. Circs are known to exist as cell-line specific and tissue-specific isoforms [20, 21] and are also known to have conserved splice sites. Recent studies have also suggested that circular RNAs might have a role in regulation of gene expression of protein-coding genes [22]. Circs have also been shown to have multiple miRNA binding sites, suggesting their role as miRNA sponges [22–25]. Different reports have also suggested the role of circ RNAs in disease and their potential as disease biomarkers [26–28].

Humans, mice, and rats share high levels of genetic conservation among each other. Rat, first mammalian species domesticated for research purposes [29], is the most widely studied experimental model so far in medical research and an excellent model for cardiovascular disease, stroke [30, 31], and hypertension [32–37]. In many cases, rats were preferred to the mouse to model learning, memory, and cognitive research [38]. Rat is also a primary model for mechanistic studies in the field of human reproduction and a standard model for physiological and toxicological studies [38, 39]. Yu et al. [40] have provided an rRNA depleted rat RNA-Seq transcriptome map for 11 tissues and 4 developmental stages (2-, 6-, 21-, and 104-week olds). Subsequent studies on rat circ RNA transcriptome have used this dataset to identify circ RNAs using CIRI [41].

Till now, several computational methods for the detection of back-splice events from RNA-Seq data have been developed, such as CIRCexplorer [42, 43], testrealign/segemehl [42, 44], circRNA_finder [11], find_circ [3, 11], CIRI [45], UROBORUS [45, 46], NCLscan [47], PTESFinder [48], KNIFE [48, 49], Pcirc_finder [50], and Acs [51]. Each of the algorithms for circRNA identification relies on different approaches including different read aligners, requirement of genome annotations, and also in many cases in the output formats. In a study by Hansen et al. [52], five circular RNA identification tools including CIRCexplorer [42, 52], circRNA_finder [11], CIRI [45], find_circ [3], and MapSplice [3, 53] were compared for the levels of false positives and sensitivity. The comparison suggested that the high number of circRNAs from a pipeline does not necessarily mean true positive candidates and one single method is not reliable. This study also concluded that it is perhaps better to use a combination of two or more methods to increase the robustness of circRNA detection, increase sensitivity, and reduce false negative and false-positive predictions. CirComPara is one such pipeline that helps in using four circRNA prediction tools including CIRCexplorer2, CIRI, find_circ, and test-realign with different combinations for different aligners including STAR, TopHat2, Bowtie2, HISAT2, and segemehl [30].

In this study, we took the opportunity to compare different combinations of aligners and the annotation tools/pipelines for circ RNA using publicly available dataset of total RNA sequencing from Yu et al. [40]. This has led to a comprehensive map of circ RNAs in rats. A subset of candidates were further independently validated using experimental approaches. This is by far the most comprehensive circ RNA transcriptome of the rat with respect to tissues, developmental stages, and gender using multiple combinations of tools.

Materials and methods

Datasets

A total of 320 datasets were downloaded from the Gene Expression Omnibus, a publicly available database with the

accession ID GSE53960. These datasets encompass study by Yu et al. report the transcriptome from different tissues and developmental stages of the rat [40]. We downloaded ribosomal RNA-depleted RNA-Seq datasets for different organs including liver, heart, kidney, brain, lung, muscle, spleen, thymus, adrenal gland, uterus, and testis. These samples were further categorized as male and female, the age of 2, 6, 21, and 104 weeks and 4 biological replicates each [GEO: GSE53960]. We have converted Sequence Read Archive (SRA) files to FASTQ format using SRA toolkit for further mapping and analysis. The SRA IDs for the datasets used in the present analysis have been summarized in [Supplementary Table S1](#).

Identification of circular RNAs from rat samples using CirComPara

We downloaded *Rattus norvegicus* reference genome (version m6) from UCSC genome browser and annotation file from Ensembl [54] database. CirComPara [30, 54] is an automated pipeline to detect, quantify, and annotate circ RNA junctions from RNA-Seq data and can be used for four different pipelines to identify back-splice junctions. This pipeline also helps in quantifying the expression of linear RNA and their gene expression which can be compared and correlated with circRNAs. CirComPara uses four different detection tools including CIRI, testrealign, CIRCexplorer, and find_circ. CirComPara can be used in any pair of pipelines or alignment tools with customized cutoffs for read coverage and filters for alignment scores. We used Segemehl, STAR, Tophat, and Bowtie2 aligners, creating a total of 4 combinations—Segemehl_CIRCexplorer2, Star_CIRCexplorer2, tophat_CIRCexplorer2, and Bowtie2_Find Circ. The first step to identify back-splice junctions is to align the RNA-Seq reads. CirComPara aligns the RNA-Seq reads over the reference genome using HISAT2 to discard reads aligning over linear transcripts and uses only the discarded reads to identify circ-junctions. After discarding the mapped reads that were possibly mapping with the linear transcriptome, we took the unmapped reads to identify circRNA junctions. We aligned our data using different aligners including HISAT2, Bowtie2, segemehl, STAR, and Tophat2. [Figure 1](#) shows a schematic representation for identification of circ RNAs.

Identification of circular RNAs from rat samples using FindCirc

In addition to CirComPara, we have also added find-circ with Bowtie2 as the aligner (the standard pipeline mentioned in Memzak et al. [3]) as an additional pipeline without using HISAT2 to discard unmapped reads which is default feature in CirComPara. The datasets downloaded were aligned over the indexed reference genome m6 using Bowtie2. In this approach, after alignment using Bowtie2, unmapped reads were segmented and mapped from the terminals head and tails to form anchor sequences. We mapped the FASTQ file over the reference genome and discarded the mapped reads. The mapped reads were discarded to avoid any ambiguity with linear transcripts. Next, using the customized scripts from the reference study, we took 20 mers from the head and tail of unmapped reads for unique alignment and extended the read to form an anchor sequence. Certain filters were applied including unique anchor alignment, maximum of two mismatches in extended sequences and alignment score above 35, a splice site of GU/AG to avoid false-positives candidates as mentioned in the reference study. The reads that fulfilled the

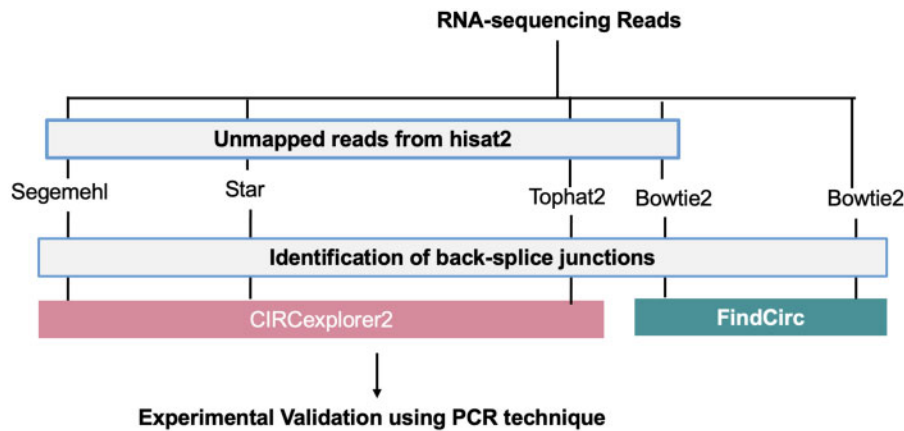


Figure 1: The schematic representation of the pipeline followed for the identification of circular RNAs. In this study five different combinations of tools including Segemehl_CIRCEplorer2, STAR_CIRCEplorer2, Tophat2_CIRCEplorer2, Bowtie2_FindCirc, and Bowtie2_FindCirc (not prealigned on HISAT2).

criteria to form putative circ RNAs were saved as output bed files.

Genome-wide annotation and distribution of circular RNA junctions

We downloaded the annotation GTF file for *Rattus norvegicus* from ensemble (Rnor_6.0.93). The annotation file has 32 624 genes, 27 780 5'UTR, 22 322 3'UTR, 40 807 transcripts, 26 224 start codon and 25 926 stop codon coordinates. We overlapped the circRNA junctions identified from each sample over the annotation file to understand the genome-wide distribution pattern, if any, for circRNA coordinates.

Tissue-specific, gender-specific, and developmental-stage-specific circular RNAs

We further analyzed the identified circRNA junctions for tissue specificity. For each tissue, we also analyzed the data with respect to developmental stage and gender. We also identified candidates with read coverage >2 , >2 to <10 , >10 to <100 , 100 – 200 , and >200 to study the relation of function with their expression level in the tissue.

Splice-site identification

We extracted the splice site nucleotides for each circRNA junction to analyze any pattern in the splice site of circ RNA and to spot any difference from the linear counterpart. We have studied three conditions as shown in [Supplementary Fig. S1](#).

Experimental validation

In order to validate the predicted circRNA junctions from Bowtie2_FindCirc pipeline, we used a polymerase chain reaction (PCR)-based approach. We randomly selected candidates based on the significance of genes in the tissue and expression of circular RNA junction that should be expressed in at least 30 out of 32 samples for each tissue. In our study we have obtained RNA of different rat tissues (Clontech, USA). In order to validate the circRNA junctions, PCR amplification was performed on complementary DNAs (cDNAs) of corresponding tissues and genomic DNA of the rat using divergent primers. We designed divergent primers of ~20 bp length across the predicted back-spliced junction for different circRNAs in six rat tissues (brain, heart, lungs, liver, kidney, and thymus). First, stranded cDNA

was prepared from 500 ng of RNA from individual tissues using random hexamers and superscript II reverse transcriptase (Invitrogen, USA). PCR amplification was performed using divergent primers designed for a total 24 candidates across 6 tissues using respective cDNAs.

Expression analysis using quantitative real-time PCR

Expression analysis was performed for selected circular RNA candidates, predicted for their tissue-specific expression patterns. RNAs from six different tissues (brain, heart, lungs, liver, kidney, and thymus) were used to synthesize cDNA as previously described. Circ RNA levels were quantified by quantitative real-time PCR (qRT-PCR), using Sybr Green mix (dssTakara, Japan) and detected by Lightcycler LC 480 (Roche). Primer sequences for qRT-PCR have been summarized in [Supplementary Table S2](#). The data in the form of cycle of threshold (Ct) were obtained and analyzed using the $\Delta\Delta Ct$ method [55].

RNAse R treatment to validate circular RNAs

RNA from each of the six tissues was treated with RNAse R as described previously [56]. In order to validate circ RNAs candidates, 5 μg of RNA from individual tissues were treated with 15 units of RNAse R (Epicentre, Illumina, USA) for 15 min at 37°C. LiCl was used to precipitate the RNAse R treated (RNAseR+) and untreated RNA (RNAseR-). After the purification, cDNA was synthesized from 1 μg of the untreated RNA and from the same amount of RNAse R treated RNA using random hexamer and Superscript II reverse transcriptase (Invitrogen, USA). PCR amplification was attempted using divergent primers for selected circular RNA candidates. *Beta-actin* was used as a control, for which convergent primers were used in the experiment. Primer sequences for PCR have been summarized in [Supplementary Table S2](#).

Results

Summary of RNA-Seq data

The RNA-Seq dataset used in the study was obtained from a previous publication which sequenced 11 tissues. The data for 320 samples encompassing 11 tissues, 4 developmental stages, both gender (male and female) were retrieved from NCBI SRA. Each of the datasets had four replicates each and had an

average of 40 million reads each. The samples and the read counts are summarized in [Supplementary Table S1](#). The average read count is 41 million reads for which Lvr_21_M_1 had the minimum number of reads, 16038547 and Kdn_104_M_2 had the maximum number of reads, that is, 82590785. [Figure 2](#) shows the total number of reads for each sample.

Combination of tools used to identify circular RNA junctions

After alignment of RNA-Seq reads over HISAT2, the average alignment percentage was 87.78%. [Supplementary Table S1](#) shows the alignment details for each sample using HISAT2. The minimum alignment was 68.75% in the liver tissue and maximum alignment was 90.53% in the kidney tissue. The unmapped reads were processed through Segemehl, Tophat, STAR, and Bowtie2 aligners to identify back-splice junctions. The alignment details for each aligner have been summarized in [Supplementary Table S1](#). The mean alignment in case of Bowtie2_findcirc (noHisat2) was 85%. Similarly, in other cases including Bowtie2, STAR, Segemehl, and Tophat2, the mean alignment percentage was 66.7, 38.7, 73.37, and 61.5%, respectively. From the unmapped reads, we observed an alignment of 30–75% from Bowtie2, 42–85% from Segemehl, 20–57% from STAR, 32–75% from Tophat, and 15–76% from Bowtie2_FindCirc. [Figure 3](#) shows the alignment percentage for each sample for each pipeline.

Identification of circular RNAs

Our analysis identified a total of 57 022 unique back-splice junctions in the rat genome from Segemehl_CIRCexplorer2, 21 375 from STAR_CIRCexplorer2, 30 943 from Tophat_CIRCexplorer2, 244 511 from Bowtie2_FindCirc and 9109 from Bowtie2_FindCirc (no HISAT2). Out of the total 5 358 249 unique cirRNA junctions from these 5 combinations, we found 490 were common among all the five combinations ([Supplementary Table S3](#)).

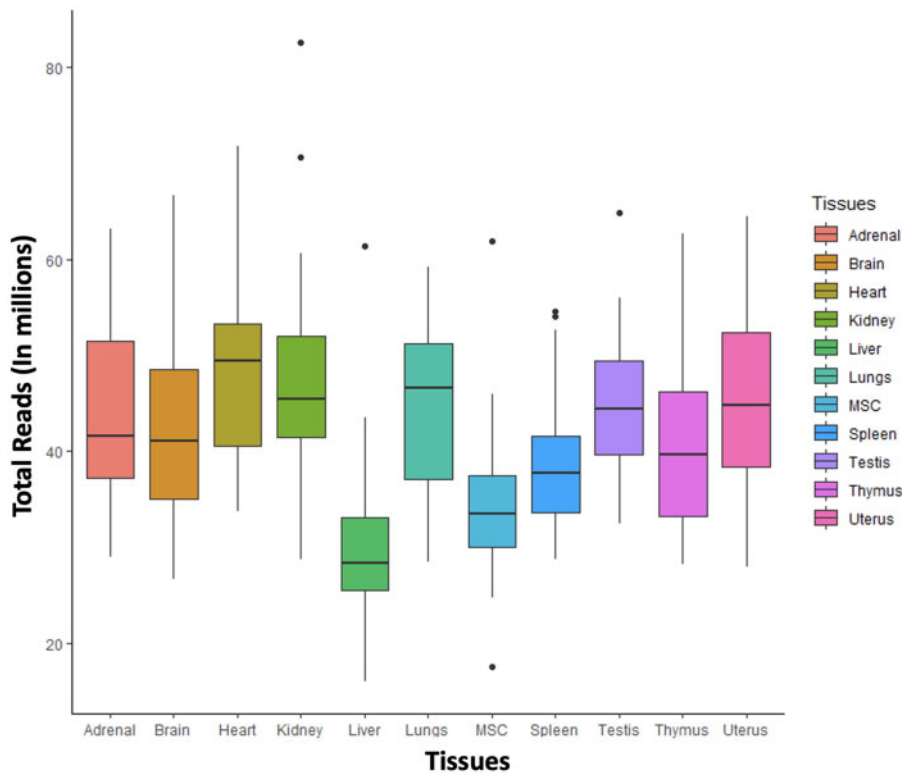


Figure 2: The total read count statistics in the different tissue sample datasets. Each tissue had 32 samples for 4 developmental stages each with replicates.

Genome-wide distribution of circ-junctions

Different studies have shown that circ RNAs loci in the genome could map to exons, introns, 5'UTR as well as 3'UTR. We mapped the circular RNAs identified from each combination to the refseq annotations to analyze the overall genome-wide

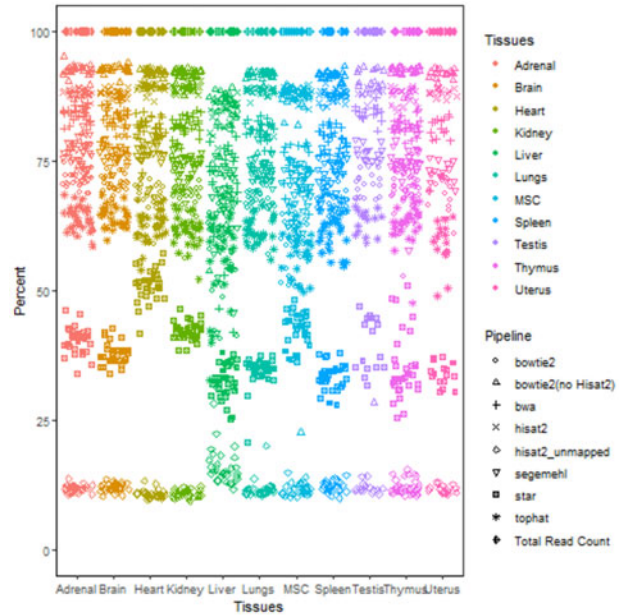


Figure 3: The alignment details of the samples in the different combinations of tools. X-axis of the plot represents different tissues and Y-axis represents percent of total reads. Different aligner conditions are shown in different shapes and colors to represent the tissues. Hisat2: linear aligned reads alignment, Hisat2 unmapped represents the alignment percentage of unmapped reads, Bowtie2 is the alignment percentage for unmapped reads from HISAT2 aligned to Bowtie2, Bowtie2 (no Hisat2) is alignment percentage for reads not prealigned using HISAT2.

distribution of circ-junctions. For the circ RNAs identified by Bowtie2_FindCirc (no. HISAT2), our analysis revealed 1111 circRNAs mapping (from start to stop) to the 5'UTR, 506 circRNAs to the 3'UTR, 1361 overlapped with the start codon, and 730 overlapped with the stop codon. In addition to this, we found 86 and 83 circRNAs that had start and stop boundaries upstream 1000 from gene start and downstream 1000 bps gene end, respectively. In the case of Bowtie2_FindCirc (after HISAT2), we found 3093 and 3062 circRNA junctions from upstream and downstream 1000 base pairs from gene boundaries. Similarly, the circular RNAs identified by the STAR+CIRCexplorer2 approach, 1994 mapped to the 5'UTR and 369 to the 3'UTR. We could not find intergenic regions in CIRCexplorer2 as it uses existing gene models for annotation by default and we would need to select the “low-confidence” option. We did not want to increase false positives in our analysis. The numbers overlapping the genome features have been shown in [Supplementary Table S4](#).

Tissue-specific circRNA junctions

Since the dataset encompassed 11 tissues, we further explored the tissue specific circ RNA junctions. Out of the total circ RNAs identified from 11 tissues, the brain was found to have the maximum number of circRNA junctions, and the liver was found to have the least number of circRNA junctions. [Supplementary Table S5](#) shows the total number of circ RNAs identified from

each tissue. Out of the total circRNA-junctions, the unique number of circRNAs specific to each tissue is also mentioned in [Supplementary Table S6](#). The data clearly show tissue-enriched and tissue-specific circular RNAs. We have shown the number of circ-junctions identified from each pipeline specific to tissues in [Supplementary Table S1](#). [Figure 4](#) summarizes the number of circRNA junctions with respect to tissues and pipelines used. While comparing all the combinations, we found that 130 from adrenal, 806 from brain, 208 from heart, 326 from kidney, 308 from lungs, 80 from liver, 111 from MSC, 312 from thymus, 244 from spleen, 299 from testis, and 135 from uterus were common in all 5 combinations. From these total circRNAs among 5 pipelines, we found 11 from adrenal, 407 from brain, 37 from heart, 10 from liver, 49 from lungs, 83 from kidney, 61 from thymus, 26 from spleen, 13 from MSC, 97 from testis, and 13 from uterus tissue-specific circ-junctions [Supplementary Fig. S2](#).

We performed gene-ontology (GO) analysis for coordinates associated with these common circular RNAs using DAVID. From our analysis, we found that the genes from 18 circ RNAs common between all the tissues were involved in general cellular processes, such as the gamma-signaling pathway, extracellular matrix organization, and transcription. In the case of adrenal tissue, we found two genes (*Cyp11a1* and *Pcsk5*) associated with response to gonadotropin with significant P-value. We found the highest number of common circ RNAs in brain tissue and GO analysis of these genes associated with these junctions showed gamma-aminobutyric acid signaling pathway,

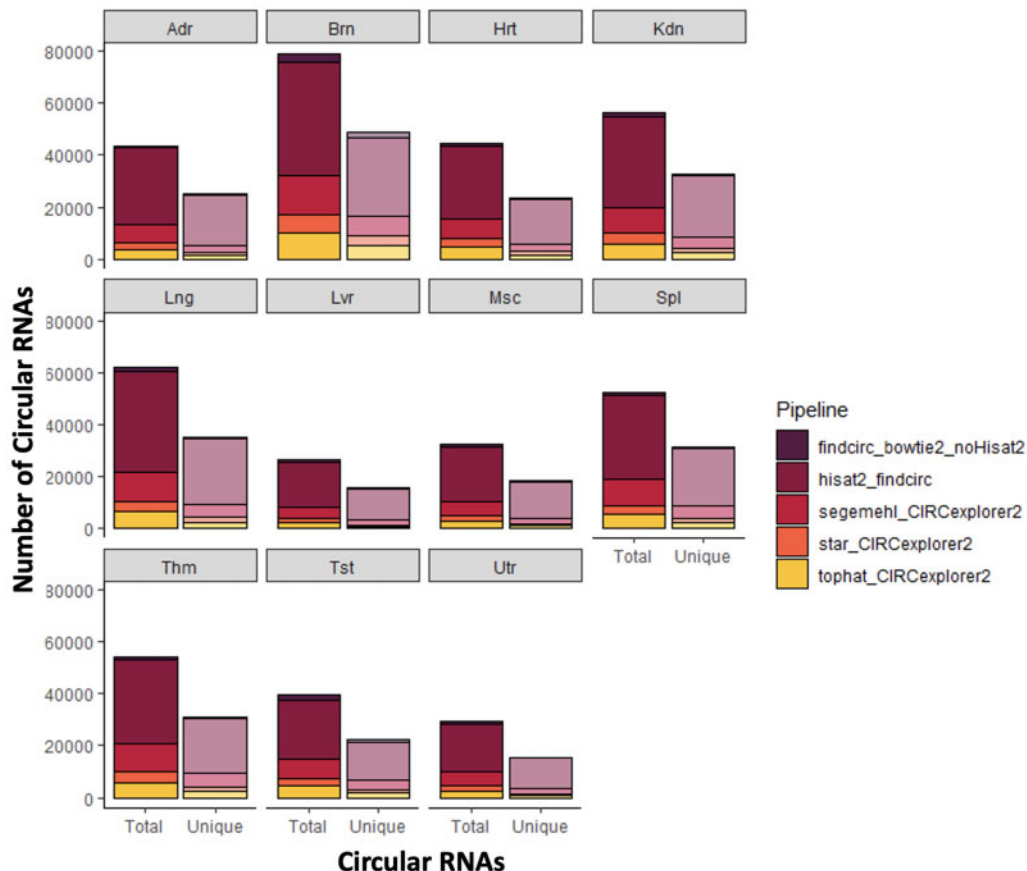


Figure 4: The plot showing the total versus unique tissue-specific number of circular RNAs for each pipeline (MSC = mesenchymal stem cells). The colors show total circular RNAs for five different combinations of tools. The darker color shows total circular RNAs and lighter color shows unique circular RNAs. Hisat2_FindCirc is a combination of unmapped reads from HISAT2 mapped on Bowtie2 and annotated with FindCirc. FindCirc_Bowtie2_noHisat2 is the combination where total reads are aligned directly on Bowtie2 followed by FindCirc.

synapse assembly, and nervous development pathways as most significant. Genes associated with memory and learning, axonogenesis, GTPase activity, and endocytosis were also found significant. In the case of heart, we found actin filament network formation and TOR signaling pathways associated with unique circ RNAs. Genes from circRNA junctions of lungs were significantly involved in cellular response to hypoxia and blood vessel modeling. *Runx2* and *Picalm* are hematopoietic genes that are associated with spleen circular RNAs. We have shown the GO analysis for these tissue-enriched circ RNAs in [Supplementary Table S16](#).

Development-stage specific circRNA junctions

The RNA-Seq dataset utilized in this analysis encompassed four developmental stages of rats that are 2-, 6-, 21-, and 108-week olds. Out of the total circ-junctions identified from Segemehl_CIRCexplorer2, we found 10 060 circ-junctions uniquely belonged to 2-week-old stage, 10 852 to 6week-old stage, 10 420 to 21-week-old stage and 10 753 to 108-week-old stage of the rat. In the case of Bowtie2_FindCirc, we found 1093 uniquely from week 2, 1241 from week 6, 1580 from week 21,

and 1674 from week 104. [Supplementary Fig. S3](#) shows the development-stage-specific circRNAs. With respect to tissues, we found a maximum number of 2-week-old stage-specific circRNAs in brain tissue and least in uterus tissue. Similarly, in 6-week-old stage, we found maximum circRNAs in brain tissue and least in liver tissue followed by 21-week-old stage where we found circRNAs in brain tissue and least from liver tissue and 108-week-old stage where we found circRNAs in brain tissue and least in mesenchymal cell tissue. [Supplementary Fig. S4](#) shows the tissue-wise development-stage specific circRNAs in rat. Our analysis suggests the maximum number of circular RNAs are identified in 104 weeks stage and least from 2 weeks stage ([Supplementary Table S7–S12](#)). Among all five combinations, we found 537 circRNA junctions common at 2 weeks stage, 655 circRNA-junctions at 6 weeks stage, 776 at 21 weeks stage, and 833 circ-junctions at 104 weeks stage. Out of these common junctions expressing among all pipeline combinations, we found 102 uniquely expressing at 2 weeks stage, 130 at 6 weeks stage, 193 at 21 weeks stage, and 247 at 104 weeks stage, and 258 were common among these stages. [Figure 5](#) shows the number of circular RNAs identified from each pipeline for each development stage. The Venn diagram is shown in

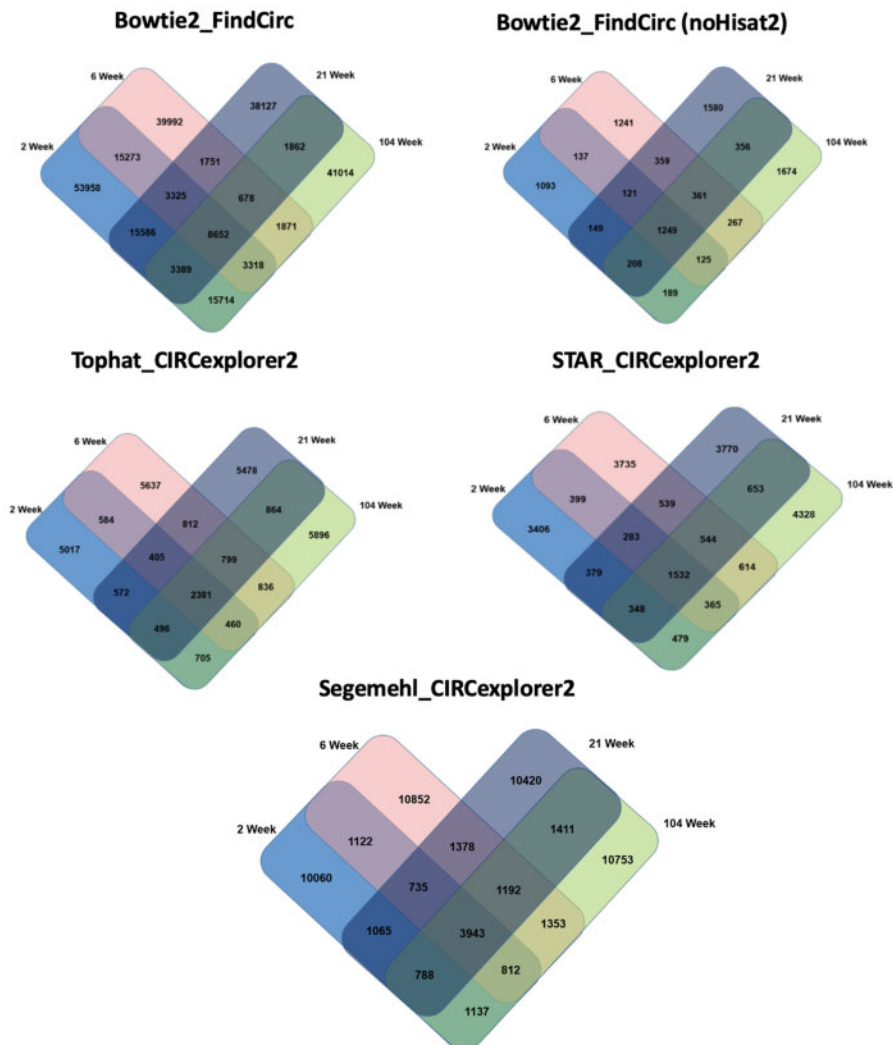


Figure 5: The figure shows a Venn diagram comparing total circular RNAs count for each pipeline in all development stages.

[Supplementary Fig. S5](#) which shows the common circ RNAs from all the combinations of tools.

Genes associated with the development stages were involved in pathways, such as GTPase activity, intracellular signal transduction, and activation of MAPK activity. GO analysis for different developmental stages circular RNAs did not give any specific pathways but we saw few interesting patterns. For example, with age, we observed the highest number of circular RNAs, and we found the highest number of circular RNAs at the most active phase of testis. We have also added GO analysis for each development-stage specific circ RNAs in [Supplementary Table S17](#).

Gender-specific circRNA junctions

We further categorized the circ-junctions based on gender. We identified 5517 circRNA junctions from female samples of Bowtie2_FindCirc (no HISAT2) and 6524 circ-junctions from male samples. Out of these 9109 total unique candidates, 32.18% junctions were common among both genders. We found 46.86 and 55.06% unique circ-junctions from females and males, respectively. Similarly, in the case of Bowtie2_FindCirc, we found 79.68% unique circ-junctions from females and 80.23% from males. In the case of Segemehl_CIRCexplorer2, we found 21% unique common junctions. We identified 23.6% common circ-junctions, 40% unique male and 37.72% unique female junctions from Star_CIRCexplorer2. Out of total circRNAs junctions from male, we found ~89% unique circRNA junctions from male and female both with 11.9% common circ-junctions. [Figure 6](#) shows the distribution of circular RNAs in both genders. Among the common circ-junctions from five combinations, we found 683 circRNA junctions common among all 5 pipelines. We found 59 circ-junctions unique to the male gender and 25 circ-junctions unique to the female gender common in all 5 pipelines (data shown in [Supplementary Tables S13–S14](#)).

GO analysis of female-specific genes showed circ RNAs were involved in memory gland duct morphogenesis and mitotic cytokinesis. GO-analysis of male-specific circRNA genes showed pathways such as sperm axoneme assembly, mRNA splicing, and signal transduction. We have also added GO analysis for gender-specific circular RNAs in [Supplementary Table S18](#).

Circ-junctions based on read coverage

We have shortlisted circ-junctions with cutoff of read coverage at splice junctions with cutoff >2 , >2 , and <10 ; >10 and <100 ; 100 – 200 and >200 to study the relation of function with their expression level in the tissue. Among five different combinations, we found that maximum number of circRNAs with circ-count 26 094 fall with >200 cutoff is from Segemehl_CIRCexplorer2 and minimum circRNAs with 3 from >200 cutoff is from Bowtie2_FindCirc combination. Segemehl_CIRCexplorer2 has a maximum number of reads in the range 100–200 reads. [Supplementary Table S15](#) shows the read counts from each range for each combination and [Fig. 7](#) shows the overall distribution of reads by coverage cutoff.

Splice-sites involve in the formation of circRNA junction

We identified the splice sites for all the three cases from start and stop sites for each pipeline combination and interestingly we found that splice sites used to identify back-splice junctions for each pipeline were conserved. We found conserved AG and AC splice sites at the start position (case 2) and GG, CC splice

sites at stop position. We also found that there is conserved T/C at one site before the splice site (TA and CA at start sites in case 1) and G/A at one site after stop position (CC, GG, GA at stop site in case 3). Similarly, at the stop position, we found conserved C/G right before the splice site (case 1 with CT/GT splice sites at stop site) and A nucleotide one position after the splice site.

Experimental validation

We selected 24 circRNA candidates based on literature significance using divergent primers approach with PCR in tissues including brain, thymus, liver, lung, and kidney that were expressed in minimum 30 samples out of 32 samples with more than 3 reads at the junction. We could validate 16 circ-junctions out of 24 candidates. The list of selected candidates is given in [Supplementary Table S2](#) with the primer sequences. Divergent sets of primers of length ~20 nucleotides were designed overlapping the back-splice junction to obtain the amplicon product of around 200–300 base pairs. The selected candidates were amplified from cDNA of respective tissues and genomic DNA as control. The genomic DNA was used to negate the possibility that the sequence of circRNA junctions could be a part of DNA itself as cases of repetitive exons and trans-splicing are reported. The results of experimental validation have been summarized in [Fig. 8](#).

Quantitative expression using qRT-PCR

The experimentally validated circRNA candidates were further analyzed for their quantitative expression analysis using qRT-PCR based approach in these tissues. Interestingly the expression of circRNA from brain specific gene ‘Anks1b’ has shown the tissue-enriched expression in case of brain and the expression of circRNA originating from thymus specific gene ‘Themis’ has shown the tissue-enriched expression in case of thymus. In the case of the liver, *Efemp1* showed tissue-enriched expression. We also performed RT-PCR for 2 circRNAs originating from genes ‘Ubr5’ and ‘Ralgapa1’. These circRNAs did not show tissue-enriched expression but showed differential expression among different tissues. We have shown the quantitative expression of the circRNAs in [Supplementary Fig. S7](#). The figure clearly shows the tissue-enriched expression of *circ-Anks1b*, *circ-Efemp1*, and *circ-Themis* and differential expression of *circ-Ubr5* and *circ-Ralgapa1*. The validation of circular RNAs was also done using RNaseR (shown in [Fig. 9](#)). We have also compared the expression with the bioinformatics data from Bowtie2_FindCirc (no HISAT2) read-coverage and the RT-PCR expression was shown to be in coherence with the bioinformatics data. The bioinformatics data for these candidates have been shown in [Supplementary Fig. S6a and b](#). *circ-themis* shows tissue-specific expression in the thymus. Host *themis* gene plays a regulatory role in T-cell positive and negative selection during thymocyte development [57, 58]. Another tissue-specific circular RNA *circ-anks1b* is predominantly found in the brain and we also found expression of *Anks1b* in brain tissue only. The *Anks1b* gene protein interacts with amyloid beta protein precursors and is involved in brain development [59]. Role of *Efemp1* has also been reported in liver cancer and methylation of the promoter causes decrease in expression of *Efemp1* in hepatocellular carcinoma (HCC) [60–62]. The gene is also shown to be involved in pathogenesis of Alzheimer’s disease. Genes *Ubr5* and *Ralgapa1* are involved in signaling pathways and we have observed circular RNAs originating from these genes expressing in multiple tissues but differentially expressed.

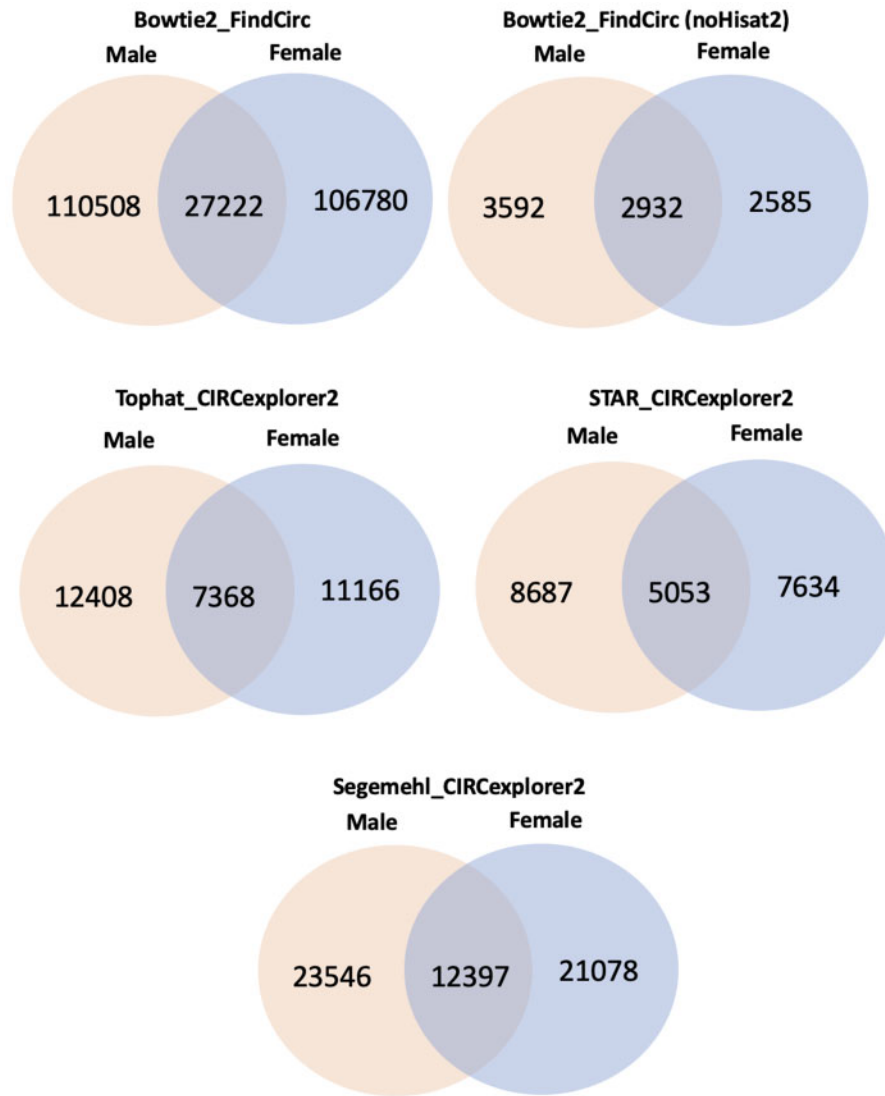


Figure 6: The Venn diagram for gender-specific circRNAs in each pipeline. Bowtie2_findCirc (no HISAT2) includes circular RNAs identified from Memczak et al. [3] pipeline without aligning prior with HISAT2.

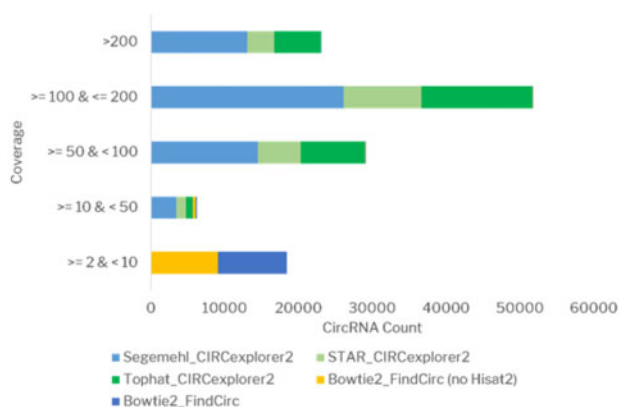


Figure 7: The classification of circular RNAs based on read coverage from each of the pipelines considered. X-axis represents different combinations of algorithms used and Y-axis represents different categories of count of circular RNAs for different read-coverage categories.

Discussion

In this study, we have performed a comprehensive analysis of circ RNAs in rat utilizing transcriptome profiling dataset made available by Yu et al [40]. The dataset encompassed 320 samples including 11 tissues, 4 developmental stages, and 2 genders each with 4 replicates.

Different short-read aligners and circ RNA annotation tools have used different algorithms which create variability in the number of identified circular RNAs when used in combination [63]. In this study, we took the opportunity to compare different aligners and annotation tools and compared the differences in the outputs. We have also compared genome-wide distribution of circ RNAs as well as read-coverage preferences using different aligners as some aligners are good for exonic region alignment whereas others perform better for introns and intergenic regions and this could clearly impact the number of circ RNAs identified. In addition, we provide a comprehensive circ RNA transcriptome map of Rat using tools for both reference-based as well as de novo-based strategies. The overlapping candidates are close to 50% even for the popular tools, such as FindCirc and

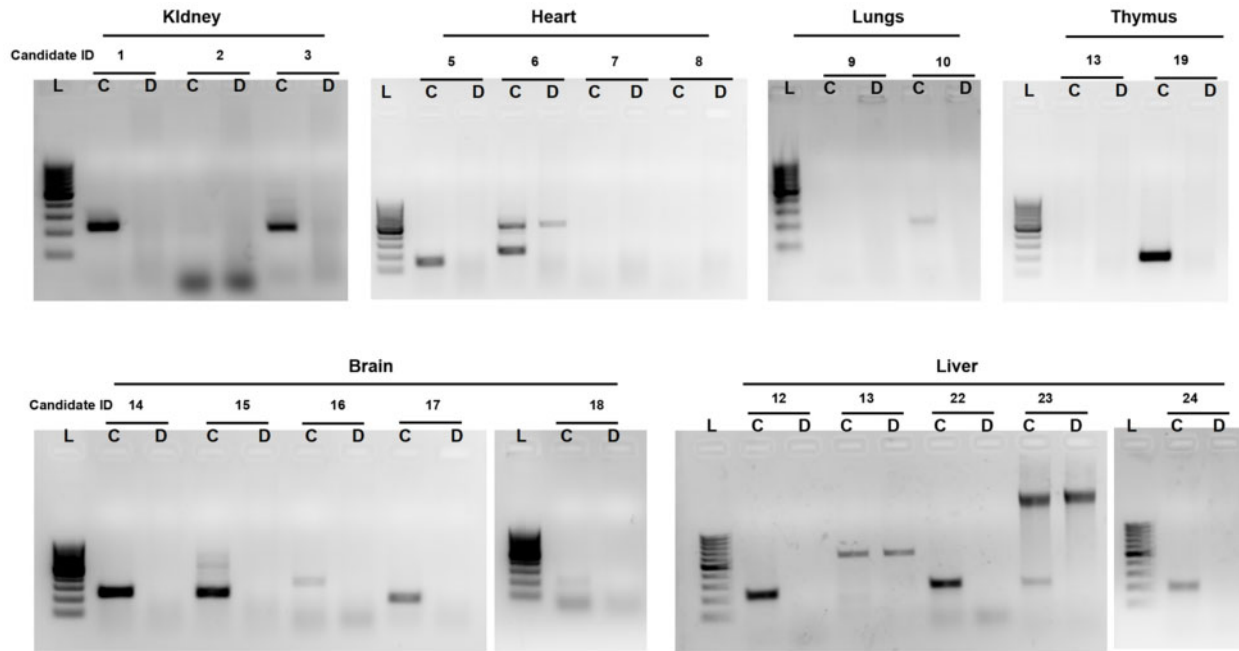


Figure 8: Validation of circular RNAs using RT-PCR for six tissues panel. Agarose gel represents the amplification products obtained from cDNA synthesized from different rat tissues and control rat DNA. Total 24 candidates were analyzed in 6 different rat tissues.

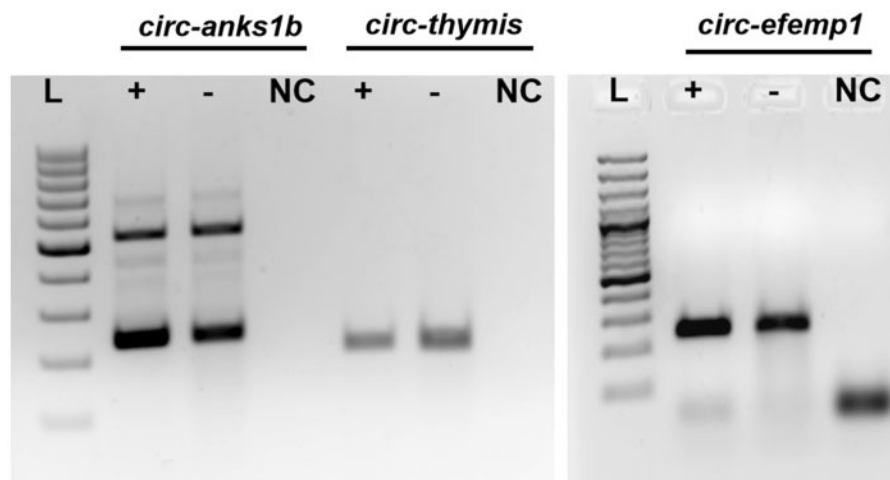


Figure 9: Validation of tissue-enriched circRNAs with RNaseR treatment using PCR based approach. Agarose gel represents PCR amplified products in RNaseR treated (+) and RNaseR not treated (-) and no template control (NC).

CIRI [45, 56], which have motivated us to use multiple combinations of tools to identify “missed” circ RNAs.

Since no single computational approach is perfect to detect circRNAs at high sensitivity and specificity, we used a combination of tools to identify putative back-splice junctions [52]. Our study focused on using five different pipeline combinations to identify circRNA junctions. These pipelines include Tophat2_CIRCexplorer2, STAR_CIRCexplorer2, Segemehl_CIRCexplorer2, Bowtie2_FindCirc, and Bowtie2_FindCirc (no HISAT2). These pipelines use aligners including segemehl, STAR, Bowtie2, and Tophat2 to identify the back-splice junctions and then different algorithms including CIRCexplorer2, Find_Circ, and segemehl to filter back-splice junctions. Our study shows the (i) tissue-specific, (ii) development stage-specific, and (iii) gender-specific circRNAs from each pipeline combination.

The RNA-Seq dataset was varying for read count from 16 to 82 millions. Liver samples had the lowest read count and kidney, brain, and heart were among the samples with the most number of reads with very good quality. The first step, in CirComPara, is to trim the input RNA-Seq reads with phred cut-off 30. Our analysis showed that the quality of reads was good enough so no reads were removed. Next step is to align the data over alignment tool to remove the reads aligning the linear transcripts. This step is very critical as it removes the ambiguous reads matching with both linear and back-splice junctions and improves the sensitivity of the circRNA detection by reducing false positives. CirComPara uses HISAT2 (hierarchical indexing for spliced alignment of transcripts) as the alignment tool to discard the reads aligning to the linear transcriptome [30]. HISAT2 uses two-way indexing, one is whole genome

Ferragina–Manzini (FM) index to anchor each alignment and multiple local FM indexes for rapid extension of the alignments. HISAT2 uses only 4 GB memory to run and is one of the fastest alignment tools available for large datasets. After alignment with HISAT2, with average alignment of 87%, alignment percentage ranged from 68.75 to 90%. The lowest alignment percentage was for liver samples and kidney and heart were samples that showed very high alignment.

CirComPara uses HISAT2 as default aligner to discard the linear mapped reads but in case of Memczak et al. [3], the complete pipeline includes Bowtie2 from the step to remove reads mapping over the linear transcripts to the identification of circRNA junctions using the anchor alignment method. So, we added one more variation to the default procedure that is find-circ with Bowtie2 aligner without aligning before with HISAT2. In this case, Bowtie2 uses options like, `-very-sensitivity`, `-score-min=C,-15,0`, to use very stringent conditions to filter out reads mapping contiguously and full length to the genomes. We observed the minimum number of circRNAs from Bowtie2_FindCirc (no HISAT2). The study describing CirComPara also explained that testrealign predicts 86% of the total circRNAs predicted and the difference with other methods is at least ~5 times more which is due to the reason that testrealign does not perform any postprocessing specific for circRNAs [30]. So not combining it with other methods can reduce the loose predictions so we have not used testrealign in our pipeline. Hansen et al. compared five prediction tools including circRNA_finder, find_circ, CIRCexplorer, CIRI, and MapSplice to check the sensitivity and number of false positives [52]. The data from the study clearly showed the problems with each pipeline. In the case of top candidates, Find_circ performs badly as it is very distracted by the highly expressed linear RNA species. CIRCexplorer is the most reliable to predict top 100 candidates. So, we did a further analysis using the individual pipeline as well as the combination of pipelines. Comparing analysis between Bowtie2_FindCirc and Bowtie2_FindCirc (no HISAT2), we found that default features of CirComPara do not use two read cutoff and splice-site cutoff of 100 kb because of which we have identified very high number of circular RNAs in Bowtie2_FindCirc. Moreover, FindCirc is not dependent on existing models of genome whereas CIRCexplorer2 uses an annotated GTF file to annotate circular RNAs. Because of these reasons, Bowtie2_FindCirc has a relatively higher number of circular RNAs than CIRCexplorer2.

Next, we were interested in observing how every pipeline differs on the level of read coverage and we found contrasting patterns between CIRCexplorer2 and FindCirc as CIRCexplorer2 had most of the circRNAs fall in the range of <200 reads whereas FindCirc pipeline had most circular RNAs in range of 2–10 reads. Studies have shown that circular RNAs are originating from exons, introns, 5'UTR, and 3'UTRs [64]. We were also interested to observe the pattern in rat and found that there are many circular RNAs originating from only 5'UTR and from 3'UTRs which could explain the regulatory role of circRNAs in transcription. Even though we did not use arguments to identify novel circ RNAs from CIRCexplorer2, we identified many circ RNAs originating from intergenic regions from FindCirc and CIRI.

While comparing tissue-specific circular RNAs, we found the maximum number of circRNAs was expressed in the brain and least in the liver. The high number of circRNAs in the brain is quite expected from the previous studies and though the least number of circRNAs in the liver could be because of the sample

read count also. We found ~22–30% of unique candidates from each tissue. Interestingly, out of number of circRNAs candidates identified from each pipeline, our analysis showed that very small of circ-junctions were common in all the five combinations in each tissue ranging from 80 to 900 (maximum in brain), out of which, 15–400 were unique to each tissue with maximum from brain, that is, 400 which increase the depth of the problem. In comparison to literature and existing databases such as circAtlas [65] and circfunbase [66], we found only ~7000 circRNAs common. Rest of the circular RNA junctions can be considered novel.

Hansen recently compared 11 computational algorithms for identification of circ RNAs and found that combination of two or more predicted pipelines can eliminate the false positives and increase the confidence for the identified candidates [67]. In this study, the author observed that CIRCexplorer and CIRCexplorer2 predicted least RNaseR sensitive species of circ RNAs. Among the top 100 expressed circRNAs, CIRI predicted the most RNaseR sensitive candidates which were from distinct loci of the genome and most of the circRNA candidates were exclusively identified by only one algorithm. Even though CIRCexplorer2 can identify de novo circ RNAs, the probability of these candidates to be true positives will be very less but it behaves well in case of exonic circular RNAs as it uses annotated gene model as reference. We also observed comparable performance of CIRCexplorer2 with different aligners but interestingly, we also found differences in read coverage of the identified circ RNAs between the outputs of different algorithms as FindCirc included circ RNAs with read count toward the lower end and CIRCexplorer2 had very high circ RNAs with read count >200 with almost no contribution from Bowtie2-FindCirc combination output in that category. The author also suggested using FindCirc with high mapping quality cutoff which also decreases RNaseR sensitive species and falsely annotated candidates ultimately reducing overall numbers of circ RNAs. CIRCexplorer2, overall, shows good performance in comparison with the rest of the tools but it is highly recommended to use with any other tool FindCirc or CIRI to increase stringency.

Our comparison of the development stage specific circ RNAs revealed most of the circRNAs from 104-week-old stage followed by 2-week-old stage. We found an interesting pattern for each tissue over the development stage as the number of circRNAs increases over the age in the brain which is also mentioned by Zhou et al. [41]. Adrenal circRNA count increases from 2- to 21-week-old stage then decreases with aging. Heart and thymus have the most number of circRNAs at puberty, that is, 6-week-old stage followed by 2-week-old stage then reduce with adult stage and aging. Lungs circRNAs are highest at immature stage and aged stage. Interestingly the circRNAs in mesenchymal stem cells are reduced with aging and in testis, circRNAs are highest in the adult stage and uterus circRNAs increase with age (highest at 104-week-old stage). So clearly there is pattern and tissue-specific behavior in circ RNAs. We also identified gender-specific circRNAs and found only 59 circRNAs that were common in all the circRNAs pipeline combinations and yet unique to the male gender. Similarly, we found 25 circRNAs specific to females that were common in all 5 combinations of tools.

Recent studies have also pointed at the coding potential of circular RNAs as a large number of circ RNAs can be translated into small peptides. CircPro [68] assigns translation potential score for each circRNA using coding potential calculator based on identification of open reading frame in the sequence but

circular RNAs might not have a start codon so CircCode [68, 69] uses a machine learning based approach to classify coding versus noncoding circRNAs. In our study, we identified three tissue-enriched circ RNAs and two differentially expressed candidates and analyzed the protein-coding potential of these circ RNAs but could not find any such property (data not shown).

Conclusions

We have created a comprehensive map of rat circular RNA transcriptome from 320 samples from 11 tissues, 4 developmental stages, and 2 genders. We have also validated a few circular RNAs candidates. Out of these, three circular RNAs showed tissue-enriched patterns and two are differentially expressed. In the current scenario, a large number of circular RNAs have been identified in most popular model organisms. It would be interesting to understand the functional role of circular RNAs in much more detail using experimental approaches.

Supplementary data

Supplementary data is available at *Biology Methods and Protocols* online.

Acknowledgments

D.S. would like to thank Intel India fellowship for providing fellowship. P.S. would like to acknowledge CSIR for his fellowship. The authors would like to acknowledge Samatha Mathew for her help in performing experiments.

Conflict of interest statement. None declared.

Data availability

Not applicable. Any required links or identifiers are present in the manuscript.

Funding

We have used the funding from CSIR-India grant MLP2001.

References

- Zhao W, Chu S, Jiao Y. Present scenario of circular RNAs (circRNAs) in plants. *Front Plant Sci* 2019;**10**:379.
- Salzman J. Circular RNA expression: its potential regulation and function. *Trends Genet* 2016;**32**:309–16.
- Memczak S, Jens M, Elefsinioti A et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 2013;**495**:333–8.
- Jeck WR, Sorrentino JA, Wang K et al. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 2013;**19**:141–57.
- Salzman J, Gawad C, Wang PL et al. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* 2012;**7**:e30733.
- Song L, Xiao Y. Downregulation of hsa_circ_0007534 suppresses breast cancer cell proliferation and invasion by targeting miR-593/MUC19 signal pathway. *Biochem Biophys Res Commun* 2018;**503**:2603–10.
- Nicolet BP, Engels S, Agliatoro F et al. Circular RNA expression in human hematopoietic cells is widespread and cell-type specific. *Nucleic Acids Res* 2018;**46**:8168–80.
- Long Z, Xie J, Liu Y-P et al. [Differentially expressed circular RNAs in human gastric cancer cells]. *Sheng Li Xue Bao* 2018;**70**:384–90.
- Zhang Z, Yang T, Xiao J. Circular RNAs: promising biomarkers for human diseases. *EBioMedicine* 2018;**34**:267–74.
- Cortés-López M, Gruner MR, Cooper DA et al. Global accumulation of circRNAs during aging in *Caenorhabditis elegans*. *BMC Genomics* 2018;**19**:8.
- Westholm JO, Miura P, Olson S et al. Genome-wide analysis of drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. *Cell Rep* 2014;**9**:1966–80.
- Feng D, Li Z, Wang G et al. Microarray analysis of differentially expressed profiles of circular RNAs in a mouse model of intestinal ischemia/reperfusion injury with and without ischemic postconditioning. *Cell Physiol Biochem* 2018;**48**:1579–94.
- Zhang J, Jiang J, Huang R et al. Circular RNA expression profiles are significantly altered in mice bone marrow stromal cells after total body irradiation. *Leuk Res* 2018;**70**:67–73.
- Singh M, George AK, Homme RP et al. Circular RNAs profiling in the cystathionine- β -synthase mutant mouse reveals novel gene targets for hyperhomocysteinemia induced ocular disorders. *Exp Eye Res* 2018;**174**:80–92.
- Wang Y, Guo Z, Zi C et al. CircRNA expression in chicken granulosa cells illuminated with red light.
- Shen M, Li T, Zhang G et al. Dynamic expression and functional analysis of circRNA in granulosa cells during follicular development in chicken. *BMC Genomics* 2019;**20**:96.
- Mester-Tonczar J, Winkler J, Einzinger P et al. Association between circular RNA CDR1as and post-infarction cardiac function in pig ischemic heart failure: influence of the anti-fibrotic natural compounds bufalin and lycorine. *Biomolecules* 2020;**10**:1180.
- Guo T, Zhang J, Yao W et al. CircINHA resists granulosa cell apoptosis by upregulating CTGF as a ceRNA of miR-10a-5p in pig ovarian follicles. *Biochim Biophys Acta Gene Regul Mech* 2019;**1862**:194420.
- Liang G, Yan J, Guo J et al. Identification of ovarian circular RNAs and differential expression analysis between MeiShan and large white pigs. *Animals (Basel)* 2020;**10**:1114.
- Coscujuela Tarrero L, Ferrero G, Miano V et al. Luminal breast cancer-specific circular RNAs uncovered by a novel tool for data analysis. *Oncotarget* 2018;**9**:14580–96.
- Zhou C, Molinie B, Daneshvar K et al. Genome-wide maps of m6A circRNAs identify widespread and cell-type-specific methylation patterns that are distinct from mRNAs. *Cell Rep* 2017;**20**:2262–76.
- Li X, Yang L, Chen L-L. The biogenesis, functions, and challenges of circular RNAs. *Mol Cell* 2018;**71**:428–42.
- Wu Z, Huang W, Wang X et al. Circular RNA CEP128 acts as a sponge of miR-145-5p in promoting the bladder cancer progression via regulating SOX11. *Mol Med* 2018;**24**:40.
- Lin X, Chen Y. Identification of potentially functional CircRNA-miRNA-mRNA regulatory network in hepatocellular

- carcinoma by integrated microarray analysis. *Med Sci Monit Basic Res* 2018;**24**:70–8.
25. Wang X, Zhu X, Zhang H et al. Increased circular RNA hsa_circ_0012673 acts as a sponge of miR-22 to promote lung adenocarcinoma proliferation. *Biochem Biophys Res Commun* 2018;**496**:1069–75.
 26. Li Y, Zeng X, He J et al. Circular RNA as a biomarker for cancer: a systematic meta-analysis. *Oncol Lett* 2018;**16**:4078–84.
 27. Weng Q, Chen M, Li M et al. Global microarray profiling identified as a potential immune-associated prognosis biomarker for hepatocellular carcinoma. *J Med Genet* 2019;**56**:32–8.
 28. Sheng J-Q, Liu L, Wang M-R et al. Circular RNAs in digestive system cancer: potential biomarkers and therapeutic targets. *Am J Cancer Res* 2018;**8**:1142–56.
 29. Kuramoto T, Nakanishi S, Ochiai M et al. Origins of albino and hooded rats: implications from molecular genetic analysis across modern laboratory rat strains. *PLoS One* 2012;**7**:e43059.
 30. Gaffo E, Bonizzato A, Kronnie GT et al. CirComPara: a multi-method comparative bioinformatics pipeline to detect and study circRNAs from RNA-seq data. *Noncoding RNA* 2017;**3**:
 31. Rubattu S, Volpe M, Kreutz R et al. Chromosomal mapping of quantitative trait loci contributing to stroke in a rat model of complex human disease. *Nat Genet* 1996;**13**:429–34.
 32. Sung Y-L, Wu C-E, Syu J-Y et al. Effects of long-term exercise on arrhythmogenesis in aged hypertensive rats. *Comput Biol Med* 2018;**102**:390–5.
 33. Bier A, Braun T, Khasbab R et al. A high salt diet modulates the gut microbiota and short chain fatty acids production in a salt-sensitive hypertension rat model. *Nutrients* 2018;**10**:1154.
 34. Redina OE, Markel AL. Stress, genes, and hypertension. contribution of the ISIAH rat strain study. *Curr Hypertens Rep* 2018;**20**:66.
 35. Serikawa T, Mashimo T, Kuramoro T et al. Advances on genetic rat models of epilepsy. *Exp Anim* 2015;**64**:1–7.
 36. Alam I, Koller DL, Cañete T et al. High-resolution genome screen for bone mineral density in heterogeneous stock rat. *J Bone Miner Res* 2014;**29**:1619–26.
 37. Alam I, Koller DL, Cañete T et al. Fine mapping of bone structure and strength QTLs in heterogeneous stock rat. *Bone* 2015;**81**:417–26.
 38. Vorhees CV, Williams MT. Assessing spatial learning and memory in rodents. *Ilar J* 2014;**55**:310–32.
 39. Martin MT, Knudsen TB, Reif DM et al. Predictive model of rat reproductive toxicity from ToxCast high throughput screening. *Biol Reprod* 2011;**85**:327–39.
 40. Yu Y, Fuscoe JC, Zhao C et al. A rat RNA-Seq transcriptomic BodyMap across 11 organs and 4 developmental stages. *Nat Commun* 2014;**5**:3230.
 41. Zhou T, Xie X, Li M et al. Rat BodyMap transcriptomes reveal unique circular RNA features across tissue types and developmental stages. *RNA* 2018;**24**:1443–56.
 42. Zhang X-O, Wang H-B, Zhang Y et al. Complementary sequence-mediated exon circularization. *Cell* 2014;**159**:134–47.
 43. Zhang X, Ma X, Ning L et al. Genome-wide identification of circular RNAs in peanut (*Arachis hypogaea* L.). *BMC Genomics* 2019;**20**:653.
 44. Hoffmann S, Otto C, Doose G et al. A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. *Genome Biol* 2014;**15**:R34.
 45. Gao Y, Wang J, Zhao F. CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biol* 2015;**16**:4.
 46. Song X, Zhang N, Han P et al. Circular RNA profile in gliomas revealed by identification tool UROBORUS. *Nucleic Acids Res* 2016;**44**:e87.
 47. Chuang T-J, Wu C-S, Chen C-Y et al. NCLscan: accurate identification of non-co-linear transcripts (fusion, trans-splicing and circular RNA) with a good balance between sensitivity and precision. *Nucleic Acids Res* 2016;**44**:e29.
 48. Izuogu OG, Alhasan AA, Alafghani HM et al. PTESFinder: a computational method to identify post-transcriptional exon shuffling (PTES) events. *BMC Bioinformatics* 2016;**17**:31.
 49. Szabo L, Morey R, Palpant NJ et al. Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human fetal development. *Genome Biol* 2015;**16**:126.
 50. Chen L, Yu Y, Zhang X et al. PcircRNA_finder: a software for circRNA prediction in plants. *Bioinformatics* 2016;**32**:3528–9.
 51. You X, Conrad TO. Acfs: accurate circRNA identification and quantification from RNA-Seq data. *Sci Rep* 2016;**6**:38820.
 52. Hansen TB, Venø MT, Damgaard CK et al. Comparison of circular RNA prediction tools. *Nucleic Acids Res* 2016;**44**:e58.
 53. Wang K, Singh D, Zeng Z et al. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res* 2010;**38**:e178.
 54. Rosenbloom KR, Armstrong J, Barber GP et al. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res* 2015;**43**:D670–81.
 55. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{(-Delta Delta C(T))} method. *Methods* 2001;**25**:402–8.
 56. Sharma D, Sehgal P, Mathew S et al. A genome-wide map of circular RNAs in adult zebrafish. *Sci Rep* 2019;**9**:3432.
 57. Chabod M, Pedros C, Lamouroux L et al. A spontaneous mutation of the rat Themis gene leads to impaired function of regulatory T cells linked to inflammatory bowel disease. *PLoS Genet* 2012;**8**:e1002461.
 58. Iwata R, Sasaki N, Agui T. Contiguous gene deletion of Ptpkr and Themis causes T-helper immunodeficiency (thid) in the LEC rat. *Biomed Res* 2010;**31**:83–7.
 59. Tindi JO, Chávez AE, Cvejic S et al. ANKS1B gene product AIDA-1 controls hippocampal synaptic transmission by regulating GluN2B subunit localization. *J Neurosci* 2015;**35**:8986–96.
 60. Chen Y, Gilbert MA, Grochowski CM et al. A genome-wide association study identifies a susceptibility locus for biliary atresia on 2p16.1 within the gene EFEMP1. *PLoS Genet* 2018;**14**:e1007532.
 61. Dou C-Y, Cao C-J, Wang Z et al. EFEMP1 inhibits migration of hepatocellular carcinoma by regulating MMP2 and MMP9 via ERK1/2 activity. *Oncol Rep* 2016;**35**:3489–95.
 62. Hu J, Duan B, Jiang W et al. Epidermal growth factor-containing fibulin-like extracellular matrix protein 1 (EFEMP1) suppressed the growth of hepatocellular carcinoma cells by promoting Semaphorin 3B(SEMA3B). *Cancer Med* 2019;**8**:3152–66.

63. Szabo L, Salzman J. Detecting circular RNAs: bioinformatic and experimental challenges. *Nat Rev Genet* 2016; **17**:679–92.
64. Santer L, Bär C, Thum T. Circular RNAs: a novel class of functional RNA molecules with a therapeutic perspective. *Mol Ther* 2019; **27**:1350–63.
65. Wu W, Ji P, Zhao F. CircAtlas: an integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes. *Genome Biol* 2020; **21**:101.
66. Meng X, Hu D, Zhang P. CircFunBase: a database for functional circular RNAs. *Database* 2019; **2019**:baz003.
67. Hansen TB. Improved circRNA identification by combining prediction algorithms. *Front Cell Dev Biol* 2018; **6**:20.
68. Meng X, Chen Q, Zhang P et al. CircPro: an integrated tool for the identification of circRNAs with protein-coding potential. *Bioinformatics* 2017; **33**:3314–6.
69. Sun P, Li G. CircCode: a powerful tool for identifying circRNA coding ability. *Front Genet* 2019; **10**:981.