




A Comparative Study of Algorithms Detecting Differential Rhythmicity in Transcriptomic Data

Lin Miao^{1,2} , Douglas E Weidemann^{1,2} , Katherine Ngo^{1,2}, Benjamin A Unruh^{1,2} and Shihoko Kojima^{1,2} 

¹Department of Biological Sciences, Virginia Tech, Blacksburg, VA, USA. ²Fralin Life Sciences Institute, Virginia Tech, Blacksburg, VA, USA.

Bioinformatics and Biology Insights
Volume 18: 1–11
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/11779322241281188



ABSTRACT: Rhythmic transcripts play pivotal roles in driving the daily oscillations of various biological processes. Genetic or environmental disruptions can lead to alterations in the rhythmicity of transcripts, ultimately impacting downstream circadian outputs, including metabolic processes and even behavior. To statistically compare the differences in transcript rhythms between 2 or more conditions, several algorithms have been developed to analyze circadian transcriptomic data, each with distinct features. In this study, we compared the performance of 7 algorithms that were specifically designed to detect differential rhythmicity (DODR, LimoRhyde, CircaCompare, compareRhythms, diffCircadian, dryR, and RepeatedCircadian). We found that even when applying the same statistical threshold, these algorithms yielded varying numbers of differentially rhythmic transcripts, most likely because each algorithm defines rhythmic and differentially rhythmic transcripts differently. Nevertheless, the output for the differential phase and amplitude were identical between dryR and compareRhythms, and diffCircadian and CircaCompare, while the output from LimoRhyde2 was highly correlated with that from diffCircadian and CircaCompare. Because each algorithm has unique requirements for input data and reports different information as an output, it is crucial to ensure the compatibility of input data with the chosen algorithm and assess whether the algorithm's output fits the user's needs when selecting an algorithm for analysis.

KEYWORDS: Circadian rhythm, transcriptome, differential rhythmicity, algorithm, rhythmic transcript

RECEIVED: January 24, 2024. **ACCEPTED:** August 19, 2024.

TYPE: Research Article

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Institute of Health [grant number: F31 AG071393 (to BAU) and R01 GM126223 (to SK)].

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Shihoko Kojima, Department of Biological Sciences, Virginia Tech, Blacksburg, VA 24061, USA. Email: skojima@vt.edu

Background

Most living organisms exhibit circadian rhythms, which are internal processes that drive biochemical and behavioral changes approximately every 24 hours.¹ The rhythmicity of these processes is driven ultimately by rhythmically expressed transcripts, which are regulated by the molecular clock machinery.^{2–8} Genetic or environmental disruptions can lead to alterations in the rhythmicity of these transcripts, ultimately impacting downstream circadian outputs, including metabolic processes and behavior.^{9–14}

To detect differences in rhythms between groups or conditions in transcriptomic data sets, different approaches have been used. One option is Venn diagram analysis (VDA). In this approach, transcripts are independently categorized as either rhythmic or nonrhythmic under each condition using a rhythmicity detection algorithm with an arbitrary statistical threshold.^{15–19} Subsequently, transcripts that are categorized differently between conditions are recognized as differentially rhythmic. This approach offers a straightforward means of distinguishing transcripts that are rhythmic under 1 condition but not under the other. However, it presents several challenges: (1) the analysis is indirect; no statistical test is performed to reject the null hypothesis that there are no differences in rhythmicity between conditions, (2) this approach can only identify transcripts that are rhythmic in 1 condition but not in the other and does not detect quantitative changes in their rhythms (eg,

changes in amplitude or phase), and (3) this approach generally overestimates the number of differentially rhythmic transcripts because of the uncontrolled false discoveries from 2 individual rhythmicity tests.^{20–22} In fact, when artificial data with benchmark transcripts were used, VDA failed to identify these transcripts as differentially rhythmic with high precision. The precision–recall performance of VDA did not improve with changes to the false discovery rate (FDR) or the amplitude thresholds.²⁰

More recently, several algorithms have been developed to statistically compare rhythmicity between 2 or more conditions. Even though these algorithms share the same goal, their approaches to detecting rhythmic transcripts and differentially rhythmic transcripts are different. In addition, the requirements for input data sets and the output information are also often different. Here, we focus on 7 of these algorithms published in or before 2022: DODR,²¹ LimoRhyde/LimoRhyde2,^{23,24} CircaCompare,²⁵ compareRhythms,²⁰ diffCircadian,²² dryR,²⁶ and RepeatedCircadian,²⁷ and perform a systematic comparison of their characteristics and performance. We highlight 3 features of these algorithms: (1) the requirements for input data, (2) the number of rhythmic and differentially rhythmic transcripts detected, and (3) the differences in their outputs. We also analyzed the same RNA-seq data set with each approach to directly compare their performance.²⁸ We uncovered both notable differences and similarities across



these algorithms and our study will help guide future users in choosing the algorithm that best suits both the user's data set and the desired output.

Materials and Methods

All analyses were performed in R (version 4.3.1), and all graphs were created with the ggplot2 (version 3.4.2), upsetR (version 1.4.0), and corrplot (version 0.92) R packages. Pearson correlation analysis was performed using the cor() function from Hmisc package (version 5.1-0). The circular version of Pearson correlation analysis was performed using the cor.circular() function from the circular package (version 0.4-95). The exact same settings were used for each algorithm throughout the paper, including MetaCycle, and a fixed period of 24 hours was used to detect both rhythmic and differentially rhythmic transcripts.

Input data set

The RNA-seq data set (GSE253826)²⁸ was used for all analyses. Reads were mapped to the mouse mm10 genome (GENECODE: GRCm38.p6.genome.fa) and transcripts per million (TPM) and reads per kilobase per million mapped (RPKM) were quantified with HOMER (V4.11.1)²⁹ using the gencode.vM25.annotation.gtf with the option of condenseGenes. Transcripts with mean TPM < 0.25 (averaged across all time points and conditions) were removed from the analysis, resulting in a total of 11 313 transcripts. We assessed the rhythmicity of these 11 313 transcripts with LimoRhyde, diffCircadian, CircaCompare, and dryR using the RPKM data. Rhythmicity of RPKM data was independently assessed by MetaCycle (v 1.2.0), which was run using the following parameters: minper = 24, maxper = 24, and cycMethod = c("ARS," "JTK," "LS"). Transcripts with Benjamini-Hochberg-adjusted q-values (B.H. q) < 0.25 were considered rhythmic. 866 transcripts that had B.H. q < 0.25 with MetaCycle in either condition were used as the input to identify differentially rhythmic transcripts in all 6 algorithms (DODR, LimoRhyde, CircaCompare, compareRhythms, diffCircadian, and dryR).

Differential rhythmicity algorithms

The DODR R package (version 0.99.2)²¹ was downloaded from <https://cran.r-project.org/src/contrib/Archive/DODR/> and manually installed in R from the tar file, since this package is no longer available from CRAN. DODR was run with the RPKM data and used the following parameters: norm = TRUE, period = 24, and method = "robust." The function p.adjust() from the stats R package with method = "BH" option was used to transform the *P* values into FDR-corrected q-values (B.H. q) using the Benjamini-Hochberg procedure.^{30,31}

The LimoRhyde R package (version 1.0.1)²³ was run using the RPKM data with period = 24. The rhythmic transcripts under the control condition were identified using the

limorhyde() function followed by limma R package (version 3.56.2)^{32,33} with the option of trend = TRUE for the function eBayes(), and the default options were used for other functions. The function p.adjust() from the stats R package with method = "BH" option was used to transform the *P* values into FDR corrected q-values (B.H. q) using the Benjamini-Hochberg procedure.³⁰ The LimoRhyde2 R package (version 0.1.0)²⁴ was run using the RPKM data with the option of period = 24, and sinusoid = TRUE for the function of getModelFit(), and the default options were used for other functions. The peak-to-trough amplitude was divided by 2 to convert it into the peak-to-mesor amplitude.

The CircaCompare R package (version 0.1.1)²⁵ was run using the RPKM data with the option of period = 24 and other default options. The output *P* values of amplitude and phase differences were converted into FDR corrected q-values (B.H. q) using the Benjamini-Hochberg procedure using the function p.adjust() from the stats R package with the method = "BH" option.³⁰ Transcripts with either significantly different amplitudes or phases were defined as differentially rhythmic. The circa_single() function was used to detect the rhythmicity of transcripts from the control condition.

The compareRhythms R package (version 0.99.3)²⁰ was run using the RPKM data with the options of method = "mod_sel," "just_classify = FALSE," period = 24 and default parameters. Transcripts classified as having "gain," "loss," or "change" of rhythmicity were defined as differentially rhythmic. The peak-to-trough amplitude was divided by 2 to convert it into the peak-to-mesor amplitude.

The diffCircadian R package (version 0.0.0)²² was run using the RPKM data with the options of period = 24, and method = "LR." The function p.adjust() from the stats R package with method = "BH" option was used to transform the *P* values into FDR corrected q-values (B.H. q) using the Benjamini-Hochberg procedure.³⁰

The dryR R package (version 1.0.0)²⁶ was run with the RPKM data using the drylm() function with the parameter of period = 24. Transcripts whose best fitting model was either 2 (rhythmic only in control), 3 (rhythmic only in *Bmal1* knock-down), or 5 (rhythmicity changed between the 2 conditions) were defined as differentially rhythmic. The peak-to-trough amplitude was divided by 2 to convert it into the peak-to-mesor amplitude. The f_24() function with the parameter of period = 24 was applied to detect the rhythmicity of transcripts from the control condition.

Results

Characteristics of algorithms for detecting differential rhythmicity

We analyzed the 7 algorithms that are the focus of this study and summarized their characteristics in Table 1. We first compared the input data requirements, as they differ in the types of data they can analyze (Table 1). Our focus centered on several

Table 1. Characteristics of 7 algorithms used in this study.

APPROACH	DODR		LIMORHYDE/ LIMORHYDE2		CIRCACOMPARE		COMPARERHYTHMS		DIFFCIRCADIAN		DRYR		REPEATEDCIRCADIAN	
	HYPOTHESIS TESTING	No	HYPOTHESIS TESTING/NO STATISTICAL TEST	No/yes	HYPOTHESIS TESTING	No	HYPOTHESIS TESTING OR MODEL SELECTION	Yes	HYPOTHESIS TESTING	No	MODEL SELECTION	Yes	HYPOTHESIS TESTING	No
Input	Read count option	No	No/yes	No	HYPOTHESIS TESTING	Yes	HYPOTHESIS TESTING OR MODEL SELECTION	Yes	HYPOTHESIS TESTING	No	MODEL SELECTION	Yes	HYPOTHESIS TESTING	No
	Uneven sampling intervals	Yes	Yes/yes	Yes	HYPOTHESIS TESTING	Yes	HYPOTHESIS TESTING OR MODEL SELECTION	Yes	HYPOTHESIS TESTING	Yes	MODEL SELECTION	Yes	HYPOTHESIS TESTING	Yes
	Missing sample	Yes	Yes/yes	Yes	HYPOTHESIS TESTING	Yes	HYPOTHESIS TESTING OR MODEL SELECTION	Yes	HYPOTHESIS TESTING	Yes	MODEL SELECTION	Yes	HYPOTHESIS TESTING	Yes
	Missing value	Yes	Yes/yes	Yes	HYPOTHESIS TESTING	Yes	HYPOTHESIS TESTING OR MODEL SELECTION	No	HYPOTHESIS TESTING	No	MODEL SELECTION	Yes	HYPOTHESIS TESTING	Yes
	Nonsinusoidal rhythm option	No	Yes/yes	No	HYPOTHESIS TESTING	No	HYPOTHESIS TESTING OR MODEL SELECTION	No	HYPOTHESIS TESTING	No	MODEL SELECTION	No	HYPOTHESIS TESTING	No
	Period	User selected	User selected/ user selected	User selected or estimated by model	HYPOTHESIS TESTING	User selected	HYPOTHESIS TESTING OR MODEL SELECTION	User selected	HYPOTHESIS TESTING	User selected	MODEL SELECTION	User selected	HYPOTHESIS TESTING	User selected
	Repeated measurement required	No	No/no	No	HYPOTHESIS TESTING	No	HYPOTHESIS TESTING OR MODEL SELECTION	No	HYPOTHESIS TESTING	No	MODEL SELECTION	No	HYPOTHESIS TESTING	Yes
Output	Parameters tested for significant differences	Overall rhythmicity	Overall rhythmicity/ none	Amplitude, phase	HYPOTHESIS TESTING	Overall rhythmicity	HYPOTHESIS TESTING OR MODEL SELECTION	Overall rhythmicity	HYPOTHESIS TESTING	Overall rhythmicity, amplitude, phase	MODEL SELECTION	Overall rhythmicity	HYPOTHESIS TESTING	Overall rhythmicity
	Statistical test result provided	P value	P value/none	P value	HYPOTHESIS TESTING	P value or model weight	HYPOTHESIS TESTING OR MODEL SELECTION	P value	HYPOTHESIS TESTING	P value	MODEL SELECTION	model weight	HYPOTHESIS TESTING	P value
	Circadian parameters	No	No/yes	Yes	HYPOTHESIS TESTING	Yes	HYPOTHESIS TESTING OR MODEL SELECTION	Yes	HYPOTHESIS TESTING	Yes	MODEL SELECTION	Yes	HYPOTHESIS TESTING	Yes
	Circadian parameter differences	No	No/yes	Yes	HYPOTHESIS TESTING	No	HYPOTHESIS TESTING OR MODEL SELECTION	No	HYPOTHESIS TESTING	No	MODEL SELECTION	No	HYPOTHESIS TESTING	No
Basic info	Rhythm detection	No	Yes/no	Yes	HYPOTHESIS TESTING	Yes	HYPOTHESIS TESTING OR MODEL SELECTION	Yes	HYPOTHESIS TESTING	Yes	MODEL SELECTION	Yes	HYPOTHESIS TESTING	Yes
	Publication year	2016	2019/2023	2020	HYPOTHESIS TESTING	2022	HYPOTHESIS TESTING OR MODEL SELECTION	2021	HYPOTHESIS TESTING	2021	MODEL SELECTION	2021	HYPOTHESIS TESTING	2022 bioRxiv

critical characteristics including the algorithm's ability to use untransformed count data (eg, read counts), deal with uneven sampling intervals, handle missing samples or missing values for individual transcripts, and fit nonsinusoidal rhythm patterns. We also identified options to adjust the rhythm's period and whether the algorithm required data with repeated measurements. We found that LimoRhyde2, compareRhythms, and dryR can accept untransformed count data, while all the other algorithms require normalized data (eg, Reads Per Kilobase of transcript per Million mapped reads (RPKM), Counts per Million (CPM), or log-transformed expression values). All algorithms can accept data sets with uneven sampling intervals (ie, time points not evenly distributed) and missing 1 or more entire samples (eg, missing 1 or more replicas) or time points (eg, the data for specific time points are absent). Except for compareRhythms and diffCircadian, the algorithms also accept missing values for individual transcripts in individual samples (eg, 1 or more RNA expression data points can be blank or NA). Only LimoRhyde and LimoRhyde2 have the option to fit nonsinusoidal curves since these 2 algorithms have the option to fit both cosinor and periodic spline regression models. RepeatedCircadian is only compatible with data sets with repeated measurement experimental designs that have multiple measurements from the same sample over a time period (eg, measurement of blood pressure at different time points within a circadian cycle of the same individual), and not measurements from different independent samples at each time point (eg, data sets from liver samples collected from different mice at different times).

We also compared the output from each algorithm (Table 1). Most algorithms use hypothesis testing approach and provide *P* values based on the differences in the overall shape of the rhythms (DODR, LimoRhyde, compareRhythms except model selection option, diffCircadian, and RepeatedCircadian), or the difference in specific rhythm parameters, such as amplitude and phase (CircaCompare and diffCircadian). Another approach uses model selection approach (compareRhythms with the model selection option and dryR) and compares how well models with different patterns of rhythmicity (no rhythms, same rhythms, different rhythms) fit the data. Then, it reports the model that best fits the data by using Bayesian Information Criterion (BIC) or Akaike Information Criterion (AIC) which indicates the degree of certainty that the selected model best fits the data. If the best model includes differences in rhythms, the transcript is defined as differentially rhythmic. The compareRhythms package is a collection of separate algorithms, rather than a single approach, and includes options for both model selection and hypothesis testing approaches. Unlike other algorithms, LimoRhyde2 does not perform any statistical tests. Rather, it focuses on quantifying rhythm parameters and the magnitude of differences of the parameters between conditions.³⁴ Thus, this algorithm emphasizes biological relevance over statistical significance. In addition to their statistical test

results, most algorithms also return calculated circadian parameters (amplitude and phase) for each condition, while LimoRhyde2 and CircaCompare additionally report calculated differences in parameters between the 2 conditions.

Testing algorithm performance with RNA-seq data

To compare these algorithms' performance side-by-side, we first analyzed the same transcriptomic data set from mouse fibroblast NIH3T3 cells.²⁸ This data set consists of 2 conditions, control and the knockdown of the core circadian clock gene, *brain* and *muscle* *Arnt*-like protein 1 (*Bmal1* knockdown). *Bmal1* is one of the core circadian clock genes, and knockdown or knockout of *Bmal1* dramatically reduces the number of rhythmic transcripts in various mouse tissues and cells.^{35,36} For each condition, samples were collected every 2 hours throughout 1 circadian cycle (ie, 24 hours) after serum shock. To analyze these data sets, we selected the model selection option for compareRhythms to directly compare with dryR, which is another algorithm that uses the model selection approach. We did not include RepeatedCircadian in the downstream analysis, as it requires data with repeated measurement experimental design.

Comparison of the performance in detecting rhythmic transcripts

To compare the performance of each algorithm, we used a fixed period of 24 hours in all downstream analyses to ensure that the rhythmicity and differential rhythmicity detected by each algorithm would be comparable. However, all the algorithms allow users to select the range of periods (Table 1). To detect differences in rhythmicity under different conditions, transcripts have to be rhythmic in at least 1 condition a priori. Four algorithms (LimoRhyde, CircaCompare, diffCircadian, and dryR) perform a statistical test to define the rhythmicity of each transcript under each condition (Table 1: Basic info), however, the specifics of the statistical analyses are not fully described.^{22,23,25,26} The hypothesis testing approaches ("dodr," "limma," "voom," "deseq2," "edger," and "cosinor") report which genes are rhythmic under each condition, although only "dodr" includes a *P* value for each condition. For the model selection option, compareRhythms does not provide a separate rhythmicity test, although rhythmicity in each condition can be inferred from which model best fits the data. In contrast, DODR does not include a rhythmicity test, and it is recommended to use other packages, such as MetaCycle¹⁸ and RAIN,¹⁹ to preidentify rhythmic transcripts.

We therefore compared the performance of the 4 algorithms that provide rhythmicity tests for individual conditions first. Using the control condition samples, we found that CircaCompare detected the highest number of rhythmic transcripts, with 5474, 3630, and 1321 transcripts detected using

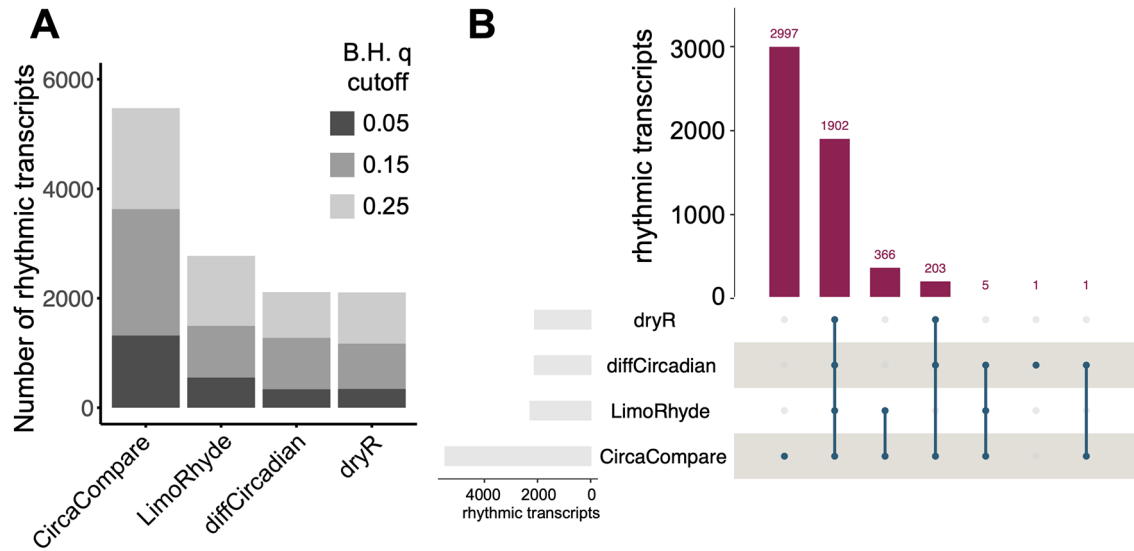


Figure 1. Rhythmic transcripts detected by CircaCompare, LimoRhyde, dryR, and diffCircadian in the control samples. A total of 11 313 transcripts (see Materials and Methods for more details) were tested for rhythmicity. Normalized (RPKM) data were used for CircaCompare, LimoRhyde, diffCircadian, and dryR. (A) Number of rhythmic transcripts with B.H. q cutoffs of 0.05, 0.15, and 0.25. (B) Overlap in transcripts identified as rhythmic by the 4 algorithms. Gray bars on the left represent the number of rhythmic transcripts detected by each algorithm. Dark red bars represent the number of transcripts commonly identified as rhythmic by the algorithms connected by the blue line, but not identified as rhythmic by the other algorithms. Transcripts were considered rhythmic if B.H. $q < 0.25$.

cutoff values of Benjamini-Hochberg-adjusted P -value (B.H. q) < 0.25 , 0.15, and 0.05, respectively (Figure 1A). LimoRhyde detected the second highest number of rhythmic transcripts with the least stringent cutoff (B.H. $q < 0.25$), followed by diffCircadian and dryR, all of which still detected over 2000 rhythmic transcripts (Figure 1A). With the most stringent cutoff (B.H. $q < 0.05$), 1902 transcripts were commonly detected as rhythmic by all 4 algorithms (Figure 1B, Table S1). As a reference, MetaCycle,¹⁸ another algorithm that detects rhythmicity, detected 794, 631, and 344 rhythmic transcripts with B.H. q cutoffs < 0.25 , 0.15, and 0.05, respectively. The number of rhythmic transcripts detected by MetaCycle was lower than the number detected by the 4 algorithms from this study at all 3 significance thresholds.

Comparison of algorithm performance in detecting differentially rhythmic transcripts

Next, we compared the performance of all the algorithms in detecting differentially rhythmic transcripts (except for LimoRhyde2 which does not perform any statistical tests). Since a transcript must be rhythmic in at least 1 condition as a prerequisite for testing for differential rhythmicity, and different algorithms yielded different numbers of rhythmic transcripts (Figure 1), we used the same set of rhythmic transcripts as an input to directly compare their performance. To this end, we preselected 866 transcripts that were detected as rhythmic (B.H. $q < 0.25$) by MetaCycle in either control or *Bmal1* knockdown cells (781 in control only, 73 in knockdown only, and 13 in both). To make comparisons as direct as possible, we

also used RPKM values, specified a 24-hour period, and fit sinusoidal curves for all the algorithms.

Among the algorithms using the hypothesis testing approach and testing for changes in overall rhythmicity, we found that DODR identified more differentially rhythmic transcripts between control and *Bmal1* knockdown cells than LimoRhyde and diffCircadian with B.H. q cutoffs < 0.25 and 0.15 (Figure 2A). However, LimoRhyde and diffCircadian detected slightly more transcripts than DODR with the most stringent cutoff at B.H. $q < 0.05$ (Figure 2A). For the model selection approaches that also assess the changes in overall rhythmicity, the recommended BICW threshold to compare 2 conditions is 0.6.^{20,26} With BICW cutoffs of 0.6, 0.75, and 0.9, dryR identified 354, 231, and 103 transcripts, while compareRhythms identified 338, 224, and 100 transcripts that are differentially rhythmic (Figure 2B). For CircaCompare, we defined differentially rhythmic transcripts as those that show statistically significant differences (B.H. $q < 0.25$, 0.15, or 0.05) in either amplitude or phase. Under this condition, CircaCompare identified 177, 144, and 79 differentially rhythmic transcripts, respectively (Figure 2C). Because of differences in methodology and the necessity of choosing arbitrary cutoff values to define rhythmicity differences, the results cannot be directly compared among these 3 groups of algorithms.

Overlaps of differentially rhythmic transcripts

Next, we analyzed how many transcripts were commonly detected as differentially rhythmic across different algorithms. Because they perform statistical tests differently, we first

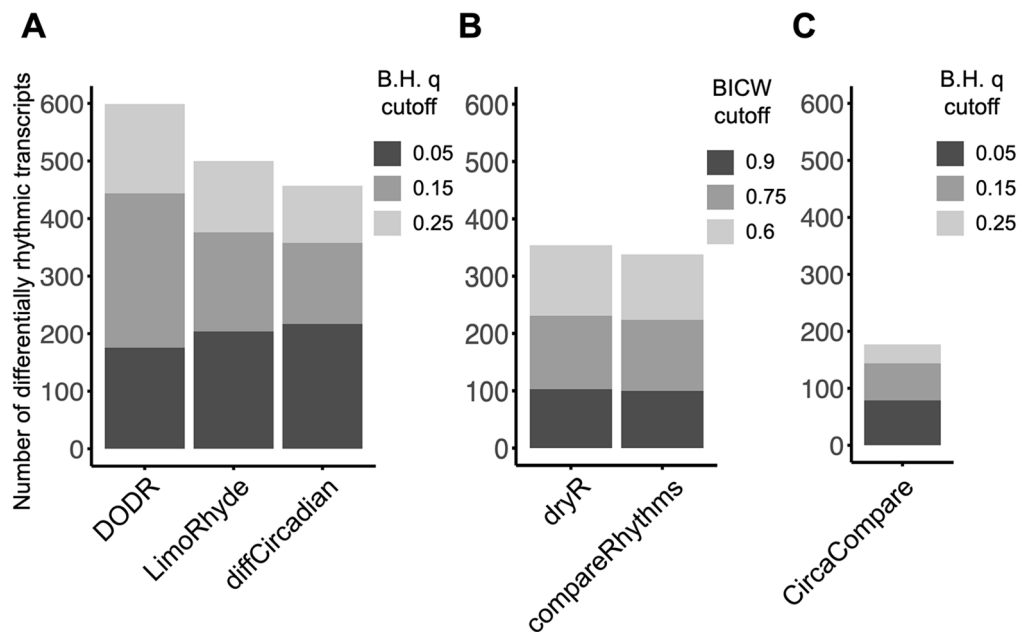


Figure 2. Number of differentially rhythmic transcripts detected by (A) DODR, LimoRhyde, and diffCircadian, (B) dryR and compareRhythms, and (C) CircaCompare. A total of 866 transcripts (rhythmic in either the control or *Bmal1* knockdown condition defined by MetaCycle B.H. $q < 0.25$) were used as input. Normalized (RPKM) data were used for all the algorithms. B.H. q cutoffs of 0.05, 0.15, and 0.25 (A, C) or BICW cutoffs of 0.9, 0.75, and 0.6 (B) were applied to determine the number of differentially rhythmic transcripts (see Materials and Methods for more details).

focused on the overlaps among the algorithms that are comparable with each other (Figure 2). Among the algorithms that use hypothesis testing approach and detect overall differential rhythmicity (DODR, LimoRhyde, and diffCircadian), a total of 209 transcripts were commonly detected as differentially rhythmic in all 3 when using B.H. $q < 0.25$ (Figure 3A). More transcripts overlapped between DODR and LimoRhyde, compared with those between DODR and diffCircadian or LimoRhyde and diffCircadian (Figure 3A). We also compared the algorithms based on the model selection approach (dryR and compareRhythms) with BICW > 0.6 as a cutoff. We found that 338 differentially rhythmic transcripts identified by compareRhythms were also detected by dryR, while dryR detected an additional 16 transcripts as differentially rhythmic. We also compared the overlaps among the 6 algorithms, even though these results are not technically comparable. We found that 32 transcripts were commonly identified as differentially rhythmic by all 6 algorithms (Figure 3B).

We also compared their performance without preidentifying rhythmic transcripts. This would test each algorithm's internal functions for both rhythmicity and differential rhythmicity, except for DODR which can directly test differential rhythmicity without preidentifying rhythmic transcripts (DODR recommends using preidentified rhythmic transcripts as an input) (Table 1). When the 11 313 transcripts were used as the input (the same as Figure 1), a lower number of differentially rhythmic transcripts was commonly detected by all 6 algorithms (27 in Figure 4 vs 32 in Figure 3B). While the overlap unique to LimoRhyde and DODR was much higher (951

in Figure 4 vs 8 in Figure 3B) and more transcripts were identified as differentially rhythmic by diffCircadian only (465 in Figure 4 vs 128 in Figure 3B). These differences derive from the method of how each algorithm defines rhythmic transcript (Figure 1), as the only difference between these 2 analyses is whether transcripts were preidentified as rhythmic by MetaCycle or not. This further emphasizes that any difference in overall performance can come from both how to define the rhythmicity and differential rhythmicity.

Are differences in amplitude and phase detected by each algorithm similar?

Five algorithms (LimoRhyde2, CircaCompare, compareRhythms, diffCircadian, and dryR) also report circadian parameters (ie, phase and amplitude) for each transcript (Table 1), allowing us to compare the differences in these parameters between conditions. We used peak-to-mesor as amplitude hereafter as this is the definition of amplitude,³⁷ even though LimoRhyde2, compareRhythms, and dryR use peak-to-trough as amplitude. Because compareRhythms, diffCircadian, and dryR only calculate each parameter for each condition, we manually calculated the difference in phase or amplitude by subtracting the value for the control condition from that of the *Bmal1* knockdown condition. We then compared these values with the differences calculated directly by CircaCompare and LimoRhyde2. We found 21 transcripts that were commonly identified as differentially rhythmic by all 5 algorithms with the least stringent cutoff (ie, B.H. $q < 0.25$ or BICW > 0.6 ;

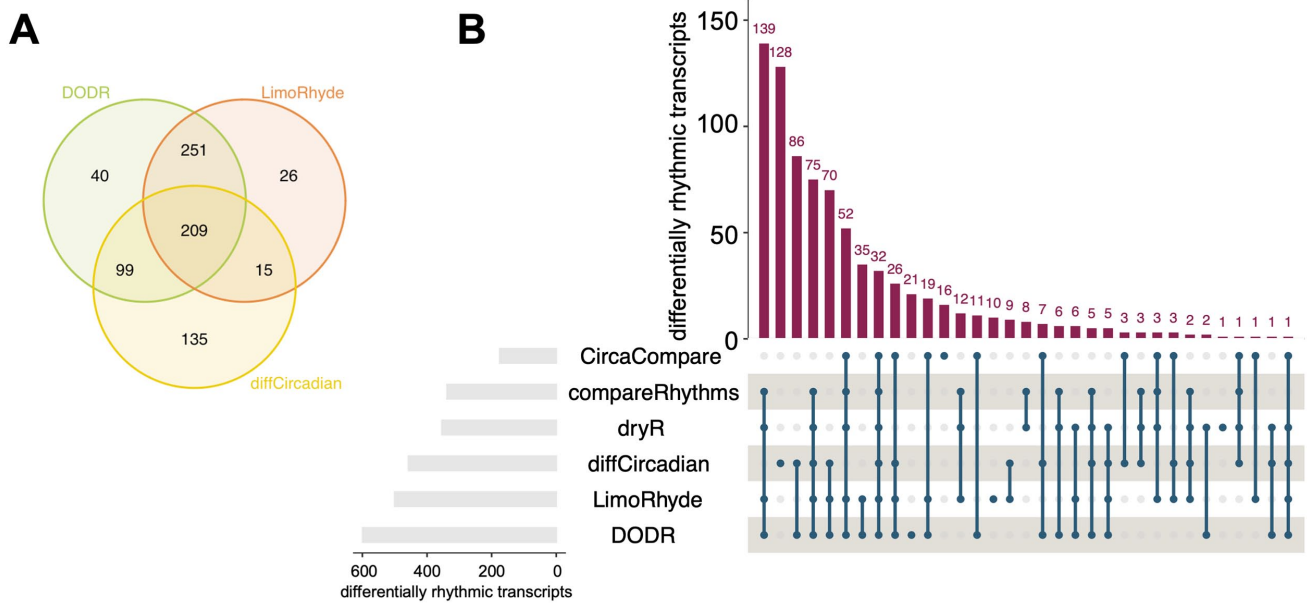


Figure 3. The number of overlapping transcripts identified as differentially rhythmic. A total of 866 transcripts (rhythmic in either the control or *Bmal1* knockdown condition defined by MetaCycle B.H. $q < 0.25$) were used as input. (A) The number of overlapping transcripts identified as differentially rhythmic among diffCircadian, LimoRhyde, and DODR with B.H. q cutoff of 0.25. (B) The number of overlapping transcripts identified as differentially rhythmic among the 6 algorithms (B.H. $q < 0.25$ or BICW > 0.6). Gray bars on the left represent the number of differential rhythmic transcripts detected by each algorithm. Dark red bars represent the number of transcripts commonly identified as differentially rhythmic by the algorithm(s) connected by the blue lines.

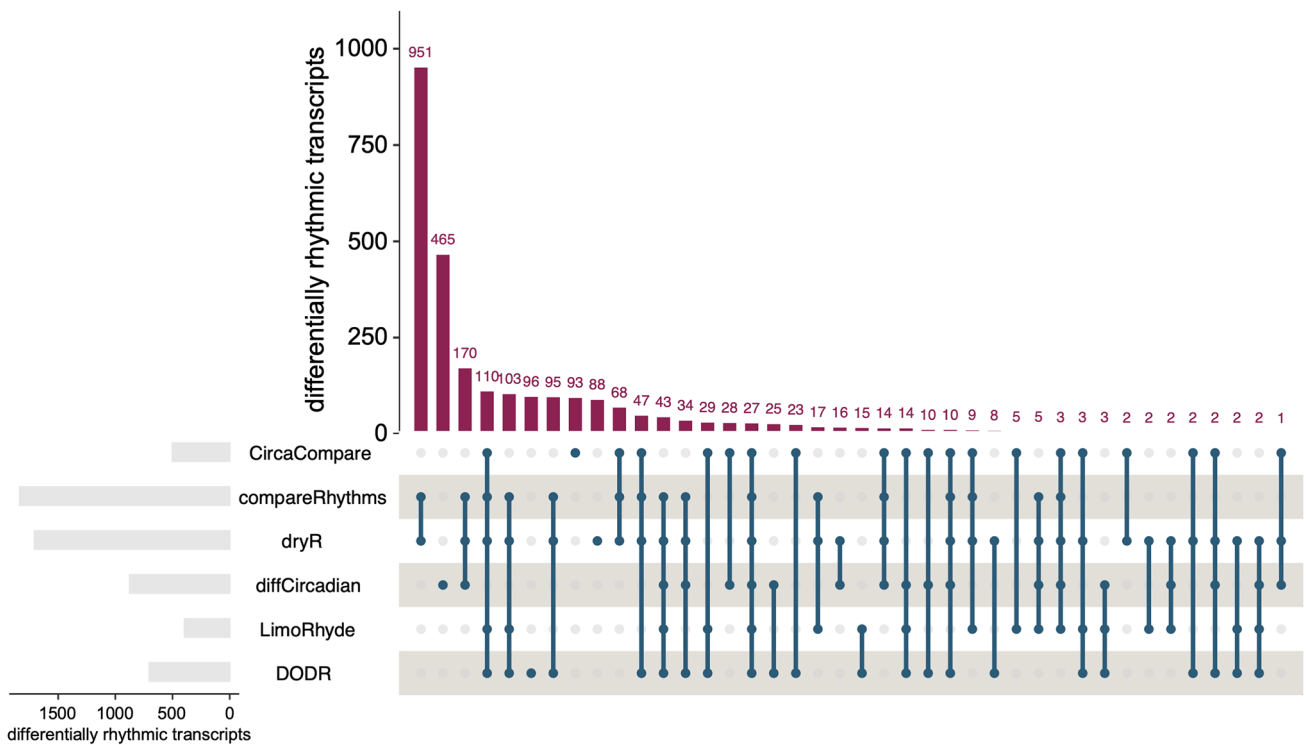


Figure 4. The number of overlapping transcripts identified as differentially rhythmic among the 6 algorithms. A total of 11 313 transcripts (see Materials and Methods for more details) were used as input. Gray bars on the left represent the number of differentially rhythmic transcripts detected by each algorithm. Differentially rhythmic transcripts were defined as B.H. $q < 0.25$ (DODR, LimoRhyde, CircaCompare, diffCircadian) or BICW > 0.6 (compareRhythms, dryR). Dark red bars represent the number of transcripts commonly identified as differentially rhythmic by the algorithm(s) connected by the blue lines.

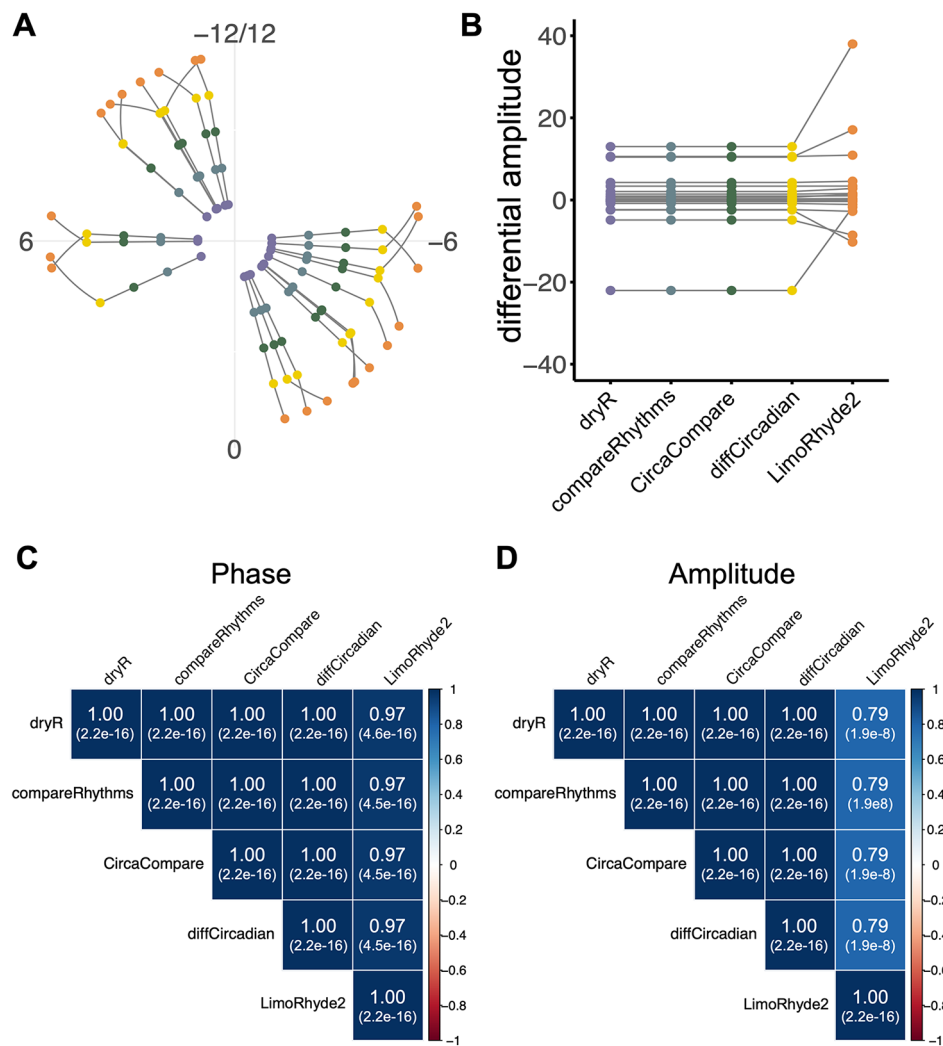


Figure 5. Comparison of the changes in phase or amplitude of differentially rhythmic transcripts commonly detected by 5 algorithms. Differences in (A) phase or (B) peak-to-mesor amplitude. Orange: LimoRhyde2, Green: CircaCompare, Purple: dryR, Yellow: diffCircadian, and Blue: compareRhythms. Each dot represents a measurement for a transcript from one of the algorithms, and each line represents each transcript. (C, D) Heatmap displaying pairwise Pearson correlation analysis of the differences in (C) phase or (D) peak-to-mesor amplitude. Blue and red denoted positive and negative correlation, respectively. The correlation coefficient and P value (number in brackets) are shown in each box.

LimoRhyde significance test results were used in place of LimoRhyde2) (Figure 5A and B). All 5 algorithms also identified these transcripts as rhythmic under both conditions ($P < 0.05$). To assess the similarities in the changes of phase and amplitude, we conducted a pairwise correlation analysis of their respective outputs (Figure 5C and D). We found that the phase and amplitude differences measured by all 5 algorithms are positively correlated with each other (Pearson: $P < 1.9 \times 10^{-8}$) (Figure 5C and D). The correlation is particularly strong (correlation coefficient=1) among CircaCompare, compareRhythms, diffCircadian, and dryR for both phase and amplitude. These indicate that the changes in phase and amplitude measured by these 5 algorithms were remarkably similar for these 21 transcripts.

We also compared the differential phase and amplitude output from the 5 algorithms, regardless of whether each transcript was defined as differentially rhythmic. We used the 866 rhythmic transcripts as the input (MetaCycle B.H. $q < 0.25$, either in Control or *Bmal1* knockdown samples, the same input used in Figures 2 and 3). 606 transcripts were first removed from the downstream analysis, because they were considered arrhythmic in one of the conditions at least by 1 algorithm and the phase and amplitude could not be calculated and compared. Note that these transcripts can be “differentially rhythmic” especially when the algorithm defines them as arrhythmic. From the 260 remaining transcripts, we removed the 21 transcripts commonly detected as differentially rhythmic (Figure 5). When we compared the differential phase or amplitude of

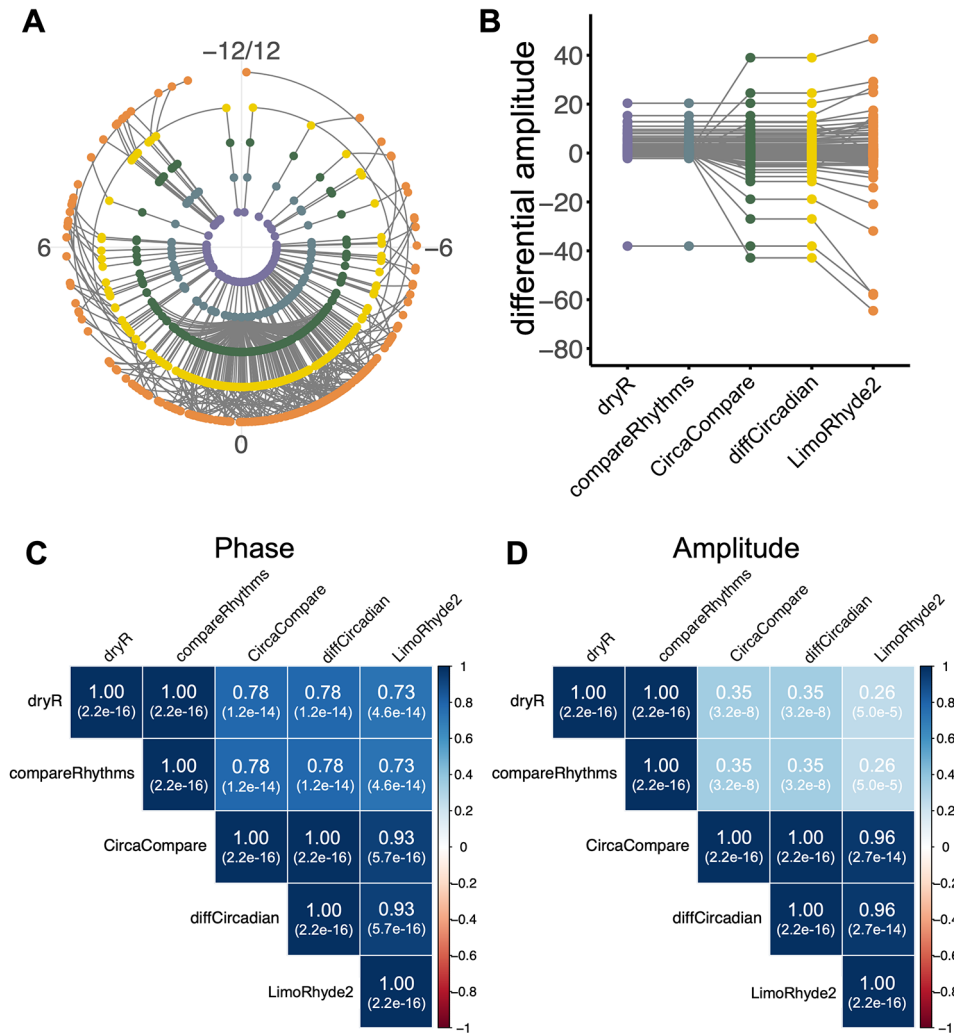


Figure 6. Comparison of the changes in phase or amplitude of transcripts across 5 algorithms. Differences in (A) phase or (B) peak-to-mesor amplitude detected by the 5 algorithms. Orange: LimoRhyde2, Green: CircaCompare, Purple: dryR, Yellow: diffCircadian, and Blue: compareRhythms. Each dot represents a measurement for a transcript from one of the algorithms, and each line represents each transcript. (C, D) Heatmap displaying pairwise Pearson correlation analysis of the differences in (C) phase or (D) peak-to-mesor amplitude calculated by the 5 algorithms. Blue and red denoted positive and negative correlation, respectively. The correlation coefficient and *P* value (number in brackets) are shown in each box.

the remaining 239 transcripts, we found that the output was again identical between CircaCompare and diffCircadian and between compareRhythms and dryR for both phase and amplitude (Figure 6A and B). The pairwise correlation analysis supported this, and the correlation coefficients were 1.00 for both phase and amplitude between CircaCompare and compareRhythms, and between diffCircadian and dryR (Figure 6C and D). The correlation between these pairs was not as high, especially for the amplitude, whereas the correlation between CircaCompare/diffCircadian and LimoRhyde2 was very high for both phase and amplitude even though they use different methods to calculate the differential phase and amplitude (CircaCompare/diffCircadian: estimated from modified cosine curve,^{22,25} LimoRhyde 2: multivariate adaptive shrinkage (Mash))³⁸ (Figure 6C and D).

Discussion

In this study, we compared 7 algorithms designed to detect differential rhythmicity between transcriptomic data sets. These algorithms show distinct features and use different approaches in defining the rhythmicity of transcripts as well as in identifying differences in rhythmicity between different conditions (Table 1). They also report different numbers of rhythmic transcripts and differentially rhythmic transcripts even with the same statistical threshold (Figures 1 and 2). These differences most likely derive from (1) how each algorithm defines rhythmicity (Figures 1 and 4), (2) how each algorithm calculates the phase and amplitude of each transcript (Figures 5 and 6), (3) how each algorithm compares rhythmicity (eg, overall rhythmicity vs changes in phase or amplitude) (Figure 2), (4) the type of statistical test (or lack thereof) used to define differential

rhythmicity (eg, model selection vs hypothesis testing) (Figure 2, Table 1), and (5) how the statistical threshold is chosen for both rhythmicity and differential rhythmicity. Nevertheless, we still detected large overlaps in the sets of differentially rhythmic transcripts detected by each algorithm (Figures 3 and 4), and the phase and amplitude differences measured by each algorithm for the set of commonly detected transcripts also showed high similarities (Figure 5).

These algorithms provide important and substantial advancements over VDA.²⁰⁻²² However, they also face several key limitations. Successfully detecting a change in rhythmicity depends heavily on statistical power, and this is affected by the amount of noise in the data, the magnitude of the change, and the number of data points (eg, number of sampling time points, number of replicates).²⁶ Different studies with different amounts of data noise or different experiment designs can have large differences in statistical power, potentially resulting in large differences in the number of differentially rhythmic transcripts that are detected.^{39,40} Another important consideration is that statistical significance does not always indicate biological significance. *LimoRhyde2* attempts to address this by focusing on estimating rhythm parameters and effect sizes, and the magnitude of rhythm differences. However, establishing a direct link between the amplitude (or change in amplitude) of a rhythm and its biological significance remains challenging. Follow-up work will likely be necessary to uncover if differences in rhythms are biologically relevant. Finally, these algorithms are only designed to compare rhythms between data sets that share the same scale. Data with different units or scales, for instance, comparing rhythms between transcriptomic data and proteomics data, or between transcriptomic data and RNA synthesis or degradation rates,²⁸ cannot be effectively analyzed using these algorithms. Venn diagram analysis, in this case, can be used as it evaluates rhythmicity under each condition independently. However, it still has the issues of indirect analysis, uncontrolled FDRs, and a lack of sensitivity to quantitative circadian changes. The development of algorithms capable of detecting differential rhythmicity in such data sets would represent a valuable and much-needed contribution to the field.

As users, we found several features helpful when analyzing data and recommend that developers include them in their future algorithms and R packages. First, we prefer algorithms that exhibit wide applicability and have the flexibility to accept a wide range of data types and experiment designs, as this would benefit a diverse range of users. Second, we advocate for a complete and detailed description of the algorithm, presented in a way that nonstatisticians can understand what the algorithm does and interpret the results correctly. For instance, clearly describe the hypotheses being tested by a statistical test, or the steps used for model fitting. Thorough code documentation and “vignettes” with worked examples are also important for users learning how to run the analysis. Third, it is important that algorithms offer

explicit and comprehensive information regarding the input data requirements, including formatting requirements and any other characteristics (eg, do the data need to have Gaussian-distributed residuals?). As for output, it is preferable for algorithms to provide calculated statistical values (eg, *P* values or BICW) rather than applying statistical threshold filtering within the code and only returning qualitative results. Finally, algorithms that allow users to specify the output they require have broader applicability. This flexibility is particularly useful when dealing with large data sets or computationally intensive tasks, as it can significantly reduce both time and computational load or memory requirements.

In addition to transcriptomic data, a wide variety of circadian time series data should be analyzable by one or more of these approaches as long as the data sufficiently meets the assumptions of the underlying regression model and other algorithm components. In addition to the requirements listed in Table 1, nontranscriptomic data can be analyzed with any of the algorithms in this study if they have residuals (the difference between each data point’s actual value and the regression model’s predicted value) with a Gaussian distribution and constant variance and lack problematic outlier data points. The only exception is *RepeatedCircadian*, which should only be used for repeated measurement experiments. For input data featuring non-Gaussian distributions or outliers, there are several options. For example, *DODR* can accommodate data with non-Gaussian distributions as long as the distribution is the same between the 2 groups being compared.²¹ Positive integer count data that have similar characteristics to RNA-seq read counts (negative binomial residual distribution and variance that changes with the mean of the data) can be analyzed with *LimoRhyde2*, *compareRhythms*, and *dryR*. However, in this case, users should select an option designed for RNA-seq read count data (eg, *voom*, *DESeq2*, or *edgeR* based method).^{20,24,26} Alternatively, data can potentially be transformed before analysis so that the residuals have a Gaussian distribution with constant variance, then analyzed using one of the Gaussian-assuming approaches listed above.²² If there are outliers in the data, *DODR* can be used.²¹ RNA-seq data sets contain data on thousands of features (eg, thousands of transcripts), yet it is common for other types of circadian studies to measure only one or a few features. In this case, *DODR*, *CircaCompare*, *dryR* (*drylm*), and *RepeatedCircadian* are the most straightforward to use.

Overall, we hope our study will help future users in selecting the algorithm that is most suitable for their data and specific research goals. Understanding the characteristics of these algorithms is a critical first step toward this, and thus our analysis will contribute to improved analyses of differential rhythmicity in transcriptomic data.

Acknowledgements

We thank the members of the Kojima lab for manuscript proofreading.

Author Contributions

LM contributed conceptualization, data curation, formal analysis, investigation, project administration, supervision, validation, visualization, writing—original draft preparation.

DEW contributed formal analysis, investigation, validation, writing—review & editing.

KN contributed formal analysis, investigation.

BAU contributed conceptualization, funding acquisition.

SK contributed conceptualization, data curation, funding acquisition, project administration, resources, supervision, validation, writing—review & editing.

Data Availability

The data set used and analyzed in the study is available in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE253826).

ORCID iDs

Lin Miao  <https://orcid.org/0000-0002-3866-3104>

Douglas E Weidemann  <https://orcid.org/0000-0002-6751-4327>

Shihoko Kojima  <https://orcid.org/0000-0001-9579-7830>

SUPPLEMENTAL MATERIAL

Supplemental material for this article is available online.

REFERENCES

- Romeijn N, Raymann RJ, Møst E, et al. Sleep, vigilance, and thermosensitivity. *Pflugers Arch*. 2012;463:169-176. doi:10.1007/s00424-011-1042-2
- Akhtar RA, Reddy AB, Maywood ES, et al. Circadian cycling of the mouse liver transcriptome, as revealed by cDNA microarray, is driven by the suprachiasmatic nucleus. *Curr Biol*. 2002;12:540-550. doi:10.1016/s0960-9822(02)00759-5
- Miller BH, McDearmon EL, Panda S, et al. Circadian and CLOCK-controlled regulation of the mouse transcriptome and cell proliferation. *Proc Natl Acad Sci USA*. 2007;104:3342-3347. doi:10.1073/pnas.0611724104
- Panda S, Antoch MP, Miller BH, et al. Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell*. 2002;109:307-320. doi:10.1016/s0092-8674(02)00722-5
- Storch KF, Lipan O, Leykin I, et al. Extensive and divergent circadian gene expression in liver and heart. *Nature*. 2002;417:78-83. doi:10.1038/nature744
- Koike N, Yoo SH, Huang HC, et al. Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science*. 2012;338:349-354. doi:10.1126/science.1226339
- Unruh BA, Kojima S. The making and breaking of RNAs: dynamics of rhythmic RNA expression in mammals. *J Biol Rhythms*. 2020;35:519-529. doi:10.1177/0748730420957498
- Li JZ, Bunney BG, Meng F, et al. Circadian patterns of gene expression in the human brain and disruption in major depressive disorder. *Proc Natl Acad Sci USA*. 2013;110:9950-9955. doi:10.1073/pnas.1305814110
- Miao L, Batty KR, Jackson AN, Pieno HA, Rhoades MW, Kojima S. Genetic and environmental perturbations alter the rhythmic expression pattern of a circadian long non-coding RNA, Per2AS, in mouse liver. *F1000Res*. 2022;11:1073. doi:10.12688/f1000research.125628.2
- Rudic RD, McNamara P, Curtis AM, et al. BMAL1 and CLOCK, two essential components of the circadian clock, are involved in glucose homeostasis. *PLoS Biol*. 2004;2:e377. doi:10.1371/journal.pbio.0020377
- Asher G, Schibler U. Crosstalk between components of circadian and metabolic cycles in mammals. *Cell Metab*. 2011;13:125-137. doi:10.1016/j.cmet.2011.01.006
- Antoch MP, Song EJ, Chang AM, et al. Functional identification of the mouse circadian Clock gene by transgenic BAC rescue. *Cell*. 1997;89:655-667. doi:10.1016/s0092-8674(00)80246-9
- Pickel L, Sung HK. Feeding rhythms and the circadian regulation of metabolism. *Front Nutr*. 2020;7:39. doi:10.3389/fnut.2020.00039
- James SM, Honn KA, Gaddameedhi S, Van Dongen HPA. Shift work: disrupted circadian rhythms and sleep—implications for health and well-being. *Curr Sleep Med Rep*. 2017;3:104-112. doi:10.1007/s40675-017-0071-6
- Hughes ME, Hogenesch JB, Kornacker K. JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *J Biol Rhythms*. 2010;25:372-380. doi:10.1177/0748730410379711
- Lomb NR. Least-squares frequency analysis of unequally spaced data. *Astrophys Space Sci*. 1976;39:447-462. doi:10.1007/BF00648343
- Yang R, Su Z. Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation. *Bioinformatics*. 2010;26:i168-i174. doi:10.1093/bioinformatics/btq189
- Wu G, Anafi RC, Hughes ME, Kornacker K, Hogenesch JB. MetaCycle: an integrated R package to evaluate periodicity in large scale data. *Bioinformatics*. 2016;32:3351-3353. doi:10.1093/bioinformatics/btw405
- Thaben PF, Westermark PO. Detecting rhythms in time series with RAIN. *J Biol Rhythms*. 2014;29:391-400. doi:10.1177/0748730414553029
- Pelikan A, Herzel H, Kramer A, Ananthasubramaniam B. Venn diagram analysis overestimates the extent of circadian rhythm reprogramming. *FEBS J*. 2022;289:6605-6621. doi:10.1111/febs.16095
- Thaben PF, Westermark PO. Differential rhythmicity: detecting altered rhythmicity in biological data. *Bioinformatics*. 2016;32:2800-2808. doi:10.1093/bioinformatics/btw309
- Ding H, Meng L, Liu AC, et al. Likelihood-based tests for detecting circadian rhythmicity and differential circadian rhythms in transcriptomic applications. *Brief Bioinform*. 2021;22:bbab224. doi:10.1093/bib/bbab224
- Singer JM, Hughey JJ. LimoRhyde: a flexible approach for differential analysis of rhythmic transcriptome data. *J Biol Rhythms*. 2019;34:5-18. doi:10.1177/0748730418813785
- Obodo D, Outland EH, Hughey JJ. LimoRhyde2: genomic analysis of biological rhythms based on effect sizes. *PLoS ONE*. 2023;18:e0292089. doi:10.1371/journal.pone.0292089
- Parsons R, Parsons R, Garner N, Oster H, Rawashdeh O. CircaCompare: a method to estimate and statistically support differences in mesor, amplitude and phase, between circadian rhythms. *Bioinformatics*. 2020;36:1208-1212. doi:10.1093/bioinformatics/btz730
- Weger BD, Gobet C, David FPA, et al. Systematic analysis of differential rhythmic liver gene expression mediated by the circadian clock and feeding rhythms. *Proc Natl Acad Sci USA*. 2021;118:e2015803118. doi:10.1073/pnas.2015803118
- Ding H, Meng L, Xing C, Esser KA, Huo Z. Statistical methods for detecting circadian rhythmicity and differential circadian patterns with repeated measurement in transcriptomic applications. *bioRxiv*. 2022. doi:10.1101/2022.06.05.494875
- Unruh BA, Weidemann DE, Miao L, Kojima S. Coordination of rhythmic RNA synthesis and degradation orchestrates 24- and 12-h RNA expression patterns in mouse fibroblasts. *Proc Natl Acad Sci USA*. 2024;121:e2314690121. doi:10.1073/pnas.2314690121
- Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38:576-589. doi:10.1016/j.molcel.2010.05.004
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B*. 2018;57:289-300. doi:10.1111/j.2517-6161.1995.tb02031.x
- Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res*. 2001;125:279-284. doi:10.1016/s0166-4328(01)00297-2
- Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res*. 2015;43:e47. doi:10.1093/nar/gkv007
- Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004;3:3. doi:10.2202/1544-6115.1027
- Sullivan GM, Feinn R. Using effect size—or why the P value is not enough. *J Grad Med Educ*. 2012;4:279-282. doi:10.4300/jgme-d-12-00156.1
- Ray S, Valekunja UK, Stangherlin A, et al. Circadian rhythms in the absence of the clock gene Bmal1. *Science*. 2020;367:800-806. doi:10.1126/science.aaw7365
- Stokes K, Nunes M, Trombley C, et al. The circadian clock gene, Bmal1, regulates intestinal stem cell signaling and represses tumor initiation. *Cell Mol Gastroenterol Hepatol*. 2021;12:1847-1872. doi:10.1016/j.jcmgh.2021.08.001
- Cornelissen G. Cosinor-based rhythmometry. *Theor Biol Med Model*. 2014;11:16. doi:10.1186/1742-4682-11-16
- Urbut SM, Wang G, Carbonetto P, Stephens M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat Genet*. 2019;51:187-195. doi:10.1038/s41588-018-0268-8
- Brooks TG, Manjrekar A, Mrc'ela A, Grant GR. Meta-analysis of diurnal transcriptomics in mouse liver reveals low repeatability of rhythm analyses. *J Biol Rhythms*. 2023;38:556-570. doi:10.1177/07487304231179600
- Ananthasubramaniam B. Are we finding functional or merely statistically significant rhythms? *J Biol Rhythms*. 2023;38:535-536. doi:10.1177/07487304231194663