LimoRhyde2: genomic analysis of biological rhythms based on effect sizes

Dora Obodo[1,2#], Elliot H. Outland[1#], and Jacob J. Hughey[1,2,3]*

[1]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee, USA
[2]Program in Chemical and Physical Biology, Vanderbilt University School of Medicine, Nashville, Tennessee, USA
[3]Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee, USA

[#]These authors contributed equally.
*Author to whom correspondence should be addressed. Email: jakejhughey@gmail.com

Abstract

Genome-scale data have revealed daily rhythms in various species and tissues. However, current methods to assess rhythmicity largely restrict their focus to quantifying statistical significance, which may not reflect biological relevance. To address this limitation, we developed a method called LimoRhyde2 (the successor to our method LimoRhyde), which focuses instead on rhythm-related effect sizes and their uncertainty. For each genomic feature, LimoRhyde2 fits a curve using a series of linear models based on periodic splines, moderates the fits using an Empirical Bayes approach called multivariate adaptive shrinkage (Mash), then uses the moderated fits to calculate rhythm statistics such as peak-to-trough amplitude. The periodic splines capture non-sinusoidal rhythmicity, while Mash uses patterns in the data to account for different fits having different levels of noise. To demonstrate LimoRhyde2's utility, we applied it to multiple circadian transcriptome datasets. Overall, LimoRhyde2 prioritized genes having high-amplitude rhythms in expression, whereas a prior method (BooteJTK) prioritized "statistically significant" genes whose amplitudes could be relatively small. Thus, quantifying effect sizes using approaches such as LimoRhyde2 has the potential to transform interpretation of genomic data related to biological rhythms.

Introduction

Much of life on Earth shows rhythms that follow the ~24-hour cycle of day and night. To produce these daily rhythms, each organism has a system of cell-autonomous oscillators, or circadian clocks, that senses environmental cues and drives cellular, physiological, and behavioral outputs [1]. In mammals, these clocks "tick" in nearly every tissue [2], although their tissue-specific mechanisms and inter-tissue interactions are only partially understood [3–6]. To study circadian systems and their relevance to human health, genome-scale approaches are invaluable, e.g., revealing mechanisms of intercellular communication in the circadian response to feeding [7] and highlighting drug targets for circadian medicine [8].

Nonetheless, current methods for analyzing rhythms in genome-scale data have multiple limitations. Some methods require that timepoints be equally spaced, or assume that rhythms are sinusoidal. Most importantly, almost all current methods—including our method LimoRhyde [9]—focus on hypothesis testing (i.e., p-values and statistical significance), raising at least two issues [10,11]. First, the null hypothesis, e.g., that a gene has a log fold-change of exactly zero, is seldom true. Rather, observed biological effects typically vary from small to large. Second, statistical significance does not necessarily imply that a result is biologically relevant, since the p-value depends on both the estimated effect size and its uncertainty.

An alternative to calculating p-values is to estimate effect sizes such as log fold-change or rhythm amplitude directly. The broader field of genomics has developed multiple methods for estimating effects [12–14], which has become an important part of differential expression analysis. Previous work has applied effect size estimation to biological rhythms [15], but not to genome-scale data.

Therefore, we developed LimoRhyde2, a new approach to quantify rhythmicity in genomic data. LimoRhyde2 builds on previous work by our group and others to rigorously analyze data from genomic experiments [9,16,17], capture non-sinusoidal rhythms [18], and accurately estimate effect sizes [14,19]. Whereas existing methods seek to answer the question "Is there an effect?", LimoRhyde2 seeks to answer the often more relevant question "How strong is the effect?". To illustrate LimoRhyde2's utility, we applied the method to multiple circadian transcriptome datasets, comparing its output to that of a prior method. Our findings suggest that LimoRhyde2 can enable new insights into biological rhythms and circadian systems.

Methods

The LimoRhyde2 R package is available at https://limorhyde2.hugheylab.org. Code, data, and results for this study are available on Figshare (https://doi.org/10.6084/m9.figshare.22001519).

LimoRhyde2 algorithm

Fit linear models

Similarly to LimoRhyde, LimoRhyde2 starts by fitting a linear model to the measurements of each genomic feature in the dataset (e.g., expression of each gene). By default in LimoRhyde2, the model terms for time (e.g., zeitgeber or circadian time) are based on a periodic cubic spline with three internal knots. Alternatively, the terms for time can be based on sine and cosine components (i.e., cosinor, the default in LimoRhyde). The model can include additional terms for covariates, e.g., to account for batch effects, but assumes all samples come from the same condition. Thus, the model could be

$$E(y_{g,i}) = \beta_{g,0} + \sum_{j=1}^{n} \beta_{g,j} B_j(\theta_i)$$

where $E(y_{g,i})$ is the expected (log-transformed) measurement for feature g in sample i, $\beta_{g,j}$ are the unknown coefficients for feature g, n is the number of spline knots, B_j are the periodic spline basis functions with period τ, $\theta_i = \dfrac{2\pi t_i}{\tau}$, and t_i is the time for sample i. LimoRhyde2 fits the models, i.e., estimates the coefficients, using limma-trend, limma-voom [20,21], or DESeq2 [12], all state-of-the-art approaches for analyzing genomic data.

To reduce the dependence of the fitted curves on the locations of the spline knots if the linear model is based on periodic splines, LimoRhyde2 repeats the above procedure multiple times, fitting a series of models for each feature such that the locations of the knots in each model are shifted by a different amount. For m shifted models (default 3), the value of the shift d_k for model k is set to

$$d_k = \frac{(k-1) \cdot \tau}{m \cdot (n+1)}$$

The overall raw fit $f_g(t, \dots)$ for feature g is calculated as

$$f_g(t, \dots) = \frac{1}{m} \sum_{k=1}^{m} f_{g,k}(t, \dots)$$

where $f_{g,k}(t, \dots)$ indicates the expected measurement of feature g according to model k, as a function of time t and any covariates.

Moderate model coefficients

LimoRhyde2 then moderates the model coefficients to obtain posterior fits using multivariate adaptive shrinkage (Mash) [14], which uses Empirical Bayes to learn patterns of similarity between coefficients and to improve estimates of effect sizes. LimoRhyde2 runs Mash on the coefficients for time for all shifted models. LimoRhyde2 does not moderate the intercept coefficients, as the relatively large number of samples in typical circadian transcriptome experiments makes these coefficients' standard errors (and the effect of Mash) quite small. By default, LimoRhyde2 runs Mash with data-driven covariance matrices, computed based on principal component analysis of strong signals in the data, with the number of principal components set to the number of spline knots. Given the raw estimates and standard errors for each coefficient for each feature, Mash computes corresponding posterior distributions, including posterior means and standard deviations. Mash's approach to estimating posteriors takes the place of the usual multiple testing adjustment, e.g., estimation of false discovery rates.

Calculate rhythm statistics

LimoRhyde2 then uses the moderated coefficients (or optionally, the raw coefficients) to calculate the following rhythm statistics, i.e., properties of the fitted curve with respect to time between 0 and τ of each feature:

- mesor (mean value)
- peak or maximum value
- peak phase (time at which the peak value occurs)
- trough or minimum value
- trough phase (time at which the trough value occurs)
- peak-to-trough amplitude (peak value minus trough value)
- root mean square (RMS) amplitude, calculated as $\sqrt{\dfrac{1}{\tau}\int_0^\tau \left(f(t) - E[f(t)]\right)^2 dt}$, where $E[f(t)] = \dfrac{1}{\tau}\int_0^\tau f(t)dt$

Quantify uncertainty

To quantify uncertainty in the fits, LimoRhyde2 can draw samples from the posterior distributions computed by Mash. For each posterior sample, which corresponds to a set of possible model coefficients for each feature, LimoRhyde2 can calculate the expected measurement of each feature as a function of time (and possibly any covariates), as well as the corresponding rhythm statistics. LimoRhyde2 then uses the resulting posterior distributions to calculate quantities such as the 90% (equal-tailed or highest-density) credible interval, the Bayesian analogue of a confidence interval.

Because amplitude estimates are non-negative, a credible interval based only on the posterior samples' amplitudes would nearly always be strictly positive (implying that that feature is rhythmic), even if the rhythm were quite weak and the posterior samples' phases were highly variable. To account for this uncertainty, LimoRhyde2 constructs credible intervals for peak-to-trough amplitude and RMS amplitude by first changing the sign (from positive to negative) of amplitudes for posterior samples whose peak phase is greater than $\tau/4$ away in either direction from the circular mean peak phase (weighted by amplitude). In this way, the credible interval for a rhythm whose amplitude is highly uncertain can span zero.

Processing circadian transcriptome data from mice

For microarray data from mouse lung (GSE59396), we used the seeker R package [22] to download the sample metadata and processed (Illumina) expression data from NCBI GEO, map probes to Entrez Gene IDs [23], and return $\log_2$-transformed expression values. For RNA-seq data from mouse liver (GSE67305) and suprachiasmatic nucleus (SCN) (GSE72095), we used seeker to download the sample metadata and to download and process the raw reads. We processed the data using Trim Galore for adapter and quality trimming [24], FastQC [25] and

MultiQC [26] for quality control, and salmon [27] and tximport [28] for quantifying gene-level counts and abundances based on Ensembl Gene IDs. We obtained the transcriptome index for salmon using refgenie [29]. Based on the plotMDS function from the limma package, we removed from analysis one extreme outlier sample from GSE59396. For GSE67305 and GSE72095, we kept only those genes having counts per million (CPM) ≥ 0.5 in at least 75% of samples (irrespective of timepoint). To avoid unrealistically low log-transformed CPM values and artificially inflated effect size estimates, for each sample-gene combination that had zero counts, we set the counts to the minimum of the non-zero counts across all samples for that gene.

Quantifying rhythmicity using LimoRhyde2

We ran LimoRhyde2 using three knots (or using a cosinor model where noted), a period of 24 h, and either limma-trend (for GSE59396) or limma-voom (for GSE67305 and GSE72095). Where applicable, we calculated 90% equal-tailed credible intervals based on 200 posterior samples.

Detecting rhythmicity using BooteJTK

To compare performance of a prior method for detecting rhythmicity in genomic data, we used BooteJTK (source code from https://github.com/alanlhutchison/BooteJTK). BooteJTK performs hypothesis testing for rank-order correlation between a feature's time series and a set of reference waveforms, using parametric bootstrapping to account for measurement uncertainty. We ran BooteJTK using the default settings. In particular, we generated 25 bootstrap resamples of each gene's time series, used a period of 24 h, treated samples collected 24 h apart as replicates, and searched for phases and asymmetries at 2-h intervals from 0 to 22 h. For each dataset, we passed BooteJTK the log-transformed microarray expression values or the $log_2(cpm + 1)$ values for the same set of genes that we passed to LimoRhyde2. We adjusted the resulting p-values using the Benjamini-Hochberg (BH) method to control the false discovery rate [30]. To quantify agreement between LimoRhyde2 and BooteJTK in ranking rhythmic genes, we calculated Cohen's kappa using the irr R package.

Results

To demonstrate how LimoRhyde2 quantifies rhythmicity, we applied it to transcriptome data from three mouse tissues (liver, lung, and suprachiasmatic nucleus - SCN) in which samples were acquired every 2 or 4 h for at least 24 h in a 12-h light:12-h dark cycle [31–33] (Table S1). We used LimoRhyde2 to fit periodic spline-based linear models for each gene (raw fits), moderate the raw fits using Mash (producing posterior fits), and calculate each gene's rhythm statistics based on the raw and posterior fits (Fig. 1). As expected, the posterior fits tended to have lower rhythm amplitudes compared to the raw fits (Fig. 2), and higher standard errors in the raw fits led to greater amplitude reduction in the posterior fits (Fig. S1). In addition, the periodic splines allowed LimoRhyde2 to fit rhythms that appeared non-sinusoidal, avoiding the occasional poor fit of the cosinor model (Fig. S2).

To compare LimoRhyde2 to a prior method for detecting rhythmicity, we analyzed the same data using BooteJTK [34], a refinement of the popular JTK_CYCLE [35] that accounts for

measurement uncertainty. Like other prior methods, BooteJTK calculates a p-value (an estimate of the probability that a given feature is *not* rhythmic) for each genomic feature, which can then be used for ranking. As LimoRhyde2 does not calculate p-values, but instead estimates a gene's rhythm, a convenient way to use its output to rank genomic features is by posterior amplitude. Importantly, the raw fits do not account for standard error and are therefore unreliable indicators of true effect size.

Across the three datasets, we found the adjusted p-values from BooteJTK were only modestly correlated with the amplitudes from LimoRhyde2 (Fig. 3A; mean Spearman correlation 0.72). Whereas BooteJTK prioritized monotonic rhythms with high signal-to-noise but perhaps relatively low amplitude, LimoRhyde2 prioritized high-amplitude rhythms of various shapes (Fig. 3B). Accordingly, the two methods showed relatively weak agreement in the top-ranked genes in each tissue (Fig. 3C). The two methods also differed in runtime: 2 minutes for LimoRhyde2 and 73 minutes for BooteJTK (mean per study, LimoRhyde2 run in parallel on 6 cores, BooteJTK does not run in parallel).

As a typical use case, we further analyzed the posterior fits and statistics from LimoRhyde2 for each dataset. Here, we also used LimoRhyde2's ability to sample from the posterior distributions to quantify uncertainty as 90% credible intervals (Fig. 4). As expected, a noisy fit, such as that of Nr1d1 in the SCN, led to a credible interval for amplitude that spanned zero (Fig. 4A-C). Consistent with prior work [36], most clock genes showed high amplitudes in each tissue (Fig. S3), although some of the highest-amplitude genes were tissue-specific (Fig. 4D). Contrary to previous findings [37], the highest-amplitude genes tended to be moderately expressed, not the most highly expressed (Fig. S4A). Furthermore, among the top 25% of genes by amplitude, the joint distribution of amplitude and phase, as well as the marginal distribution of phase, differed widely by tissue (Fig. S4). Overall, these results illustrate the value of LimoRhyde2's approach to quantifying rhythmicity.

Discussion

Genomic data have yielded valuable insights into circadian systems and their relevance to human health, but previous methods for genomic analysis of biological rhythms have multiple limitations, perhaps the most severe being an overreliance on p-values and statistical significance.

LimoRhyde2's posterior estimates—fitted curves, resulting statistics, and credible intervals—account for uncertainty in the raw estimates and for different patterns of rhythmicity. This aspect of LimoRhyde2 is enabled by Mash [14], whose Empirical Bayes procedure shares information among genomic features. Such sharing may reduce goodness of fit for any one genomic feature, but is a well-validated strategy to prevent overfitting and improve estimates [14,19], while avoiding the danger that "statistically significant" results can imply overestimated effect sizes [38]. Thus, in LimoRhyde2, genome-scale data are neither a burden (for multiple testing) nor a curse (of dimensionality), but rather, an advantage.

Traditionally, efforts to understand the core circadian clock in various species have focused less on rhythm amplitude than on period and phase. However, the goal of genomic studies is often not to understand the core clock itself, but rather to determine how the clock and daily rhythms influence physiology. In this respect, amplitude may be as relevant as period and phase. For example, other things being equal, a protein whose mRNA shows a rhythm amplitude of 3 $\log_2$CPM is likely a more promising target than one whose mRNA shows an amplitude of only 0.3.

Given the distinct goals of LimoRhyde2 compared to previous methods, the relevant differences between the methods' output are not quantifiable in terms related to binary classification (precision, recall, false positive, etc.). Therefore, we opted to analyze real circadian transcriptome data rather than simulated data, whose generation requires many simplifying assumptions. Although the true effect sizes in these data are unknown, our results indicate that LimoRhyde2 efficiently prioritizes large effects that have functional significance in the circadian system.

For example, among the genes ranked considerably higher by LimoRyde2 (amplitude) than by BooteJTK in the liver were Rgs16, Dhrs9, and Mt2. Rgs16 regulates daily rhythms of G-protein signaling in the SCN [47] and substrate oxidation in hepatocytes [54]. Dhrs9 belongs to a set of genes whose expression forms a robust biomarker of internal circadian time from human blood [55]. Mt2 (metallothionein 2) was previously shown in a targeted study to have a dramatic diurnal rhythm [56]. Among genes ranked highly by LimoRhyde2 in the lung were Fkbp5 and Adamts4. The former encodes a negative-feedback regulator of glucocorticoid signaling [57], which plays an important role in synchronizing daily rhythms across tissues [58], while the latter is a clock-controlled gene in mouse cartilage [59] and human corneal endothelial cells [60]. Among genes ranked highly in the SCN were Nfkbib, which encodes an inhibitor of NF-κB signaling and whose expression is regulated in microglia by the clock gene Nr1d1 [61], and Id1, which may play a role in photic entrainment of the circadian system [62].

Despite its advantages, LimoRhyde2 still has limitations and opportunities for future improvements. Importantly, we have currently only validated LimoRhyde2 for quantifying rhythmicity within a single condition, not for quantifying differences in rhythmicity between conditions. Second, LimoRhyde2 assumes that the rhythms have a user-specified period shared by all genomic features. Although it is possible to run LimoRhyde2 multiple times on the same dataset, varying the period each time, we recommend instead using a single period and allowing the periodic spline to capture non-monotonic (and potentially ultradian) rhythms. Third, LimoRhyde2 assumes that each feature's rhythm is fixed. Consequently, LimoRhyde2 does not model amplitude decay [63], although it can model time-dependent trends in mesor.

Just as prior methods cannot determine what level of adjusted p-values qualifies as "significant" (the conventional level of 0.05 being arbitrary), LimoRhyde2 cannot determine what magnitude of a given rhythm statistic is biologically meaningful. Such values likely vary from one gene to another anyway. Thus, when performing follow-up computational analyses, we recommend using the full distribution of a given statistic rather than drawing artificial cutoffs such as

"rhythmic" and "arrhythmic". As a convenient reference for rhythm amplitudes, however, our results indicate that the core clock genes in wild-type mouse liver have a median peak-to-trough amplitude of ~3.3 ($\log_2$CPM in RNA-seq).

By directly estimating biological rhythms and their uncertainty, LimoRhyde2 seeks to shift the focus of an analysis from detecting statistical significance to interpreting biological relevance. Although we have so far only validated LimoRhyde2 on bulk transcriptome data, recent work showed that the same state-of-the-art methods that underlie LimoRhyde2 are well-suited to analysis of single-cell RNA-seq data [43]. Thus, LimoRhyde2 may provide a basis for using various genomic techniques to improve our understanding of biological rhythms.

Acknowledgements

References

1. Patke A, Young MW, Axelrod S. Molecular mechanisms and physiological importance of circadian rhythms. Nat Rev Mol Cell Biol. 2020;21: 67–84. doi:10.1038/s41580-019-0179-2

2. Yoo S-H, Yamazaki S, Lowrey PL, Shimomura K, Ko CH, Buhr ED, et al. PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. Proc Natl Acad Sci U S A. 2004;101: 5339–5346. doi:10.1073/pnas.0308709101

3. Finger A-M, Jäschke S, Del Olmo M, Hurwitz R, Granada AE, Herzel H, et al. Intercellular coupling between peripheral circadian oscillators by TGF-β signaling. Sci Adv. 2021;7. doi:10.1126/sciadv.abg5174

4. Crosby P, Hamnett R, Putker M, Hoyle NP, Reed M, Karam CJ, et al. Insulin/IGF-1 Drives PERIOD Synthesis to Entrain Circadian Rhythms with Feeding Time. Cell. 2019;177: 896–909.e20. doi:10.1016/j.cell.2019.02.017

5. Sinturel F, Gos P, Petrenko V, Hagedorn C, Kreppel F, Storch K-F, et al. Circadian hepatocyte clocks keep synchrony in the absence of a master pacemaker in the suprachiasmatic nucleus or other extrahepatic clocks. Genes Dev. 2021. doi:10.1101/gad.346460.120

6. Qu M, Qu H, Jia Z, Kay SA. HNF4A defines tissue-specific circadian rhythms by beaconing BMAL1::CLOCK chromatin binding and shaping the rhythmic chromatin landscape. Nat Commun. 2021;12: 6350. doi:10.1038/s41467-021-26567-3

7. Guan D, Xiong Y, Trinh TM, Xiao Y, Hu W, Jiang C, et al. The hepatocyte clock and feeding control chronophysiology of multiple liver cell types. Science. 2020;369: 1388–1394. doi:10.1126/science.aba8984

8. Ruben MD, Wu G, Smith DF, Schmidt RE, Francey LJ, Lee YY, et al. A database of

tissue-specific rhythmically expressed human genes has potential applications in circadian medicine. Sci Transl Med. 2018;10. doi:10.1126/scitranslmed.aat8806

9. Singer JM, Hughey JJ. LimoRhyde: A Flexible Approach for Differential Analysis of Rhythmic Transcriptome Data. J Biol Rhythms. 2018; 748730418813785. doi:10.1177/0748730418813785

10. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev Camb Philos Soc. 2007;82: 591–605. doi:10.1111/j.1469-185X.2007.00027.x

11. Halsey LG. The reign of the p-value is over: what alternative analyses could we employ to fill the power vacuum? Biol Lett. 2019;15: 20190174. doi:10.1098/rsbl.2019.0174

12. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15: 550. doi:10.1186/s13059-014-0550-8

13. Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics. 2019;35: 2084–2092. doi:10.1093/bioinformatics/bty895

14. 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

15. Wang Y, Ke C, Brown MB. Shape-invariant modeling of circadian rhythms with random effects and smoothing spline ANOVA decompositions. Biometrics. 2003;59: 804–812. doi:10.1111/j.0006-341x.2003.00094.x

16. Thaben PF, Westermark PO. Differential rhythmicity: detecting altered rhythmicity in biological data. Bioinformatics. 2016;32: 2800–2808. doi:10.1093/bioinformatics/btw309

17. 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. 2019. doi:10.1093/bioinformatics/btz730

18. Hughey JJ, Hastie T, Butte AJ. ZeitZeiger: supervised learning for high-dimensional data from an oscillatory system. Nucleic Acids Res. 2016;44: e80. doi:10.1093/nar/gkw030

19. Stephens M. False discovery rates: a new deal. Biostatistics. 2017;18: 275–294. doi:10.1093/biostatistics/kxw041

20. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014;15: R29. doi:10.1186/gb-2014-15-2-r29

21. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43: e47. doi:10.1093/nar/gkv007

22. Schoenbachler JL, Hughey JJ. The seeker R package: simplified fetching and processing of transcriptome data. bioRxiv. 2022. p. 2022.08.30.505820. doi:10.1101/2022.08.30.505820

23. Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, et al. Evolving gene/transcript

definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res. 2005;33: e175. doi:10.1093/nar/gni179

24. Krueger F, James F, Ewels P, Afyounian E, Schuster-Boeckler B. FelixKrueger/TrimGalore: v0.6.7 - DOI via Zenodo. 2021. doi:10.5281/zenodo.5127899

25. Andrews S. FastQC: A quality control analysis tool for high throughput sequencing data. Github; Available: https://github.com/s-andrews/FastQC

26. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32: 3047–3048. doi:10.1093/bioinformatics/btw354

27. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14: 417–419. doi:10.1038/nmeth.4197

28. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. 2015;4: 1521. doi:10.12688/f1000research.7563.2

29. Stolarczyk M, Reuter VP, Smith JP, Magee NE, Sheffield NC. Refgenie: a reference genome resource manager. Gigascience. 2020;9. doi:10.1093/gigascience/giz149

30. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Series B Stat Methodol. 1995;57: 289–300. Available: http://www.jstor.org/stable/2346101

31. Haspel JA, Chettimada S, Shaik RS, Chu J-H, Raby BA, Cernadas M, et al. Circadian rhythm reprogramming during lung inflammation. Nat Commun. 2014;5: 4753. doi:10.1038/ncomms5753

32. Pembroke WG, Babbs A, Davies KE, Ponting CP, Oliver PL. Temporal transcriptomics suggest that twin-peaking genes reset the clock. Elife. 2015;4. doi:10.7554/eLife.10518

33. Janich P, Arpat AB, Castelo-Szekely V, Lopes M, Gatfield D. Ribosome profiling reveals the rhythmic liver translatome and circadian clock regulation by upstream open reading frames. Genome Res. 2015;25: 1848–1859. doi:10.1101/gr.195404.115

34. Hutchison AL, Allada R, Dinner AR. Bootstrapping and Empirical Bayes Methods Improve Rhythm Detection in Sparsely Sampled Data. J Biol Rhythms. 2018;33: 339–349. doi:10.1177/0748730418789536

35. 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

36. Zhang R, Lahens NF, Ballance HI, Hughes ME, Hogenesch JB. A circadian gene expression atlas in mammals: implications for biology and medicine. Proc Natl Acad Sci U S A. 2014;111: 16219–16224. doi:10.1073/pnas.1408886111

37. Laloum D, Robinson-Rechavi M. Methods detecting rhythmic gene expression are biologically relevant only for strong signal. PLoS Comput Biol. 2020;16: e1007666.

doi:10.1371/journal.pcbi.1007666

38. Gelman A, Carlin J. Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. Perspect Psychol Sci. 2014;9: 641–651. doi:10.1177/1745691614551642

39. Chen H, Gao L, Yang D, Xiao Y, Zhang M, Li C, et al. Coordination between the circadian clock and androgen signaling is required to sustain rhythmic expression of Elovl3 in mouse liver. J Biol Chem. 2019;294: 7046–7056. doi:10.1074/jbc.RA118.005950

40. Yang Y, Duguay D, Fahrenkrug J, Cermakian N, Wing SS. USP2 regulates the intracellular localization of PER1 and circadian gene expression. J Biol Rhythms. 2014;29: 243–256. doi:10.1177/0748730414544741

41. Chowdhury D, Wang C, Lu A, Zhu H. Identifying Transcription Factor Combinations to Modulate Circadian Rhythms by Leveraging Virtual Knockouts on Transcription Networks. iScience. 2020;23: 101490. doi:10.1016/j.isci.2020.101490

42. Gagliano O, Luni C, Li Y, Angiolillo S, Qin W, Panariello F, et al. Synchronization between peripheral circadian clock and feeding-fasting cycles in microfluidic device sustains oscillatory pattern of transcriptome. Nat Commun. 2021;12: 6185. doi:10.1038/s41467-021-26294-9

43. Carrillo GL, Su J, Monavarfeshani A, Fox MA. F-spondin Is Essential for Maintaining Circadian Rhythms. Front Neural Circuits. 2018;12: 13. doi:10.3389/fncir.2018.00013

44. Antle MC, Silver R. Orchestrating time: arrangements of the brain circadian clock. Trends Neurosci. 2005;28: 145–151. doi:10.1016/j.tins.2005.01.003

45. Webb AB, Angelo N, Huettner JE, Herzog ED. Intrinsic, nondeterministic circadian rhythm generation in identified mammalian neurons. Proc Natl Acad Sci U S A. 2009;106: 16493–16498. doi:10.1073/pnas.0902768106

46. He Y, Jones CR, Fujiki N, Xu Y, Guo B, Holder JL Jr, et al. The transcriptional repressor DEC2 regulates sleep length in mammals. Science. 2009;325: 866–870. doi:10.1126/science.1174443

47. Doi M, Ishida A, Miyake A, Sato M, Komatsu R, Yamazaki F, et al. Circadian regulation of intracellular G-protein signalling mediates intercellular synchrony and rhythmicity in the suprachiasmatic nucleus. Nat Commun. 2011;2: 327. doi:10.1038/ncomms1316

48. Tao W, Wu J, Zhang Q, Lai S-S, Jiang S, Jiang C, et al. EGR1 regulates hepatic clock gene amplitude by activating Per1 transcription. Sci Rep. 2015;5: 15212. doi:10.1038/srep15212

49. Parcha V, Patel N, Gutierrez OM, Li P, Gamble KL, Musunuru K, et al. Chronobiology of Natriuretic Peptides and Blood Pressure in Lean and Obese Individuals. J Am Coll Cardiol. 2021;77: 2291–2303. doi:10.1016/j.jacc.2021.03.291

50. Takeda N, Maemura K. Circadian clock and the onset of cardiovascular events. Hypertens Res. 2016;39: 383–390. doi:10.1038/hr.2016.9

51. Yamamuro D, Takahashi M, Nagashima S, Wakabayashi T, Yamazaki H, Takei A, et al. Peripheral circadian rhythms in the liver and white adipose tissue of mice are attenuated by

constant light and restored by time-restricted feeding. PLoS One. 2020;15: e0234439. doi:10.1371/journal.pone.0234439

52. McGlincy NJ, Valomon A, Chesham JE, Maywood ES, Hastings MH, Ule J. Regulation of alternative splicing by the circadian clock and food related cues. Genome Biol. 2012;13: R54. doi:10.1186/gb-2012-13-6-r54

53. Schick S, Becker K, Thakurela S, Fournier D, Hampel MH, Legewie S, et al. Identifying Novel Transcriptional Regulators with Circadian Expression. Mol Cell Biol. 2016;36: 545–558. doi:10.1128/MCB.00701-15

54. Bai X, Liao Y, Sun F, Xiao X, Fu S. Diurnal regulation of oxidative phosphorylation restricts hepatocyte proliferation and inflammation. Cell Rep. 2021;36: 109659. doi:10.1016/j.celrep.2021.109659

55. Wittenbrink N, Ananthasubramaniam B, Münch M, Koller B, Maier B, Weschke C, et al. High-accuracy determination of internal circadian time from a single blood sample. J Clin Invest. 2018. doi:10.1172/JCI120874

56. Zhang D, Jin T, Xu Y-Q, Lu Y-F, Wu Q, Zhang Y-KJ, et al. Diurnal-and sex-related difference of metallothionein expression in mice. J Circadian Rhythms. 2012;10: 5. doi:10.1186/1740-3391-10-5

57. Wochnik GM, Rüegg J, Abel GA, Schmidt U, Holsboer F, Rein T. FK506-binding Proteins 51 and 52 Differentially Regulate Dynein Interaction and Nuclear Translocation of the Glucocorticoid Receptor in Mammalian Cells*. J Biol Chem. 2005;280: 4609–4616. doi:10.1074/jbc.M407498200

58. Oster H, Challet E, Ott V, Arvat E, de Kloet ER, Dijk D-J, et al. The Functional and Clinical Significance of the 24-Hour Rhythm of Circulating Glucocorticoids. Endocr Rev. 2017;38: 3–45. doi:10.1210/er.2015-1080

59. Gossan N, Zeef L, Hensman J, Hughes A, Bateman JF, Rowley L, et al. The circadian clock in murine chondrocytes regulates genes controlling key aspects of cartilage homeostasis. Arthritis Rheum. 2013;65: 2334–2345. doi:10.1002/art.38035

60. Nakai H, Tsuchiya Y, Koike N, Asano T, Ueno M, Umemura Y, et al. Comprehensive Analysis Identified the Circadian Clock and Global Circadian Gene Expression in Human Corneal Endothelial Cells. Invest Ophthalmol Vis Sci. 2022;63: 16. doi:10.1167/iovs.63.5.16

61. Griffin P, Dimitry JM, Sheehan PW, Lananna BV, Guo C, Robinette ML, et al. Circadian clock protein Rev-erbα regulates neuroinflammation. Proc Natl Acad Sci U S A. 2019;116: 5102–5107. doi:10.1073/pnas.1812405116

62. Duffield GE, Watson NP, Mantani A, Peirson SN, Robles-Murguia M, Loros JJ, et al. A role for Id2 in regulating photic entrainment of the mammalian circadian system. Curr Biol. 2009;19: 297–304. doi:10.1016/j.cub.2008.12.052

63. De Los Santos H, Hurley JM, Collins EJ, Bennett KP. Circadian Rhythms in Neurospora Exhibit Biologically Relevant Driven and Damped Harmonic Oscillations. ACM BCB. 2017;2017: 455–463. doi:10.1145/3107411.3107420

Figures

Figure 1
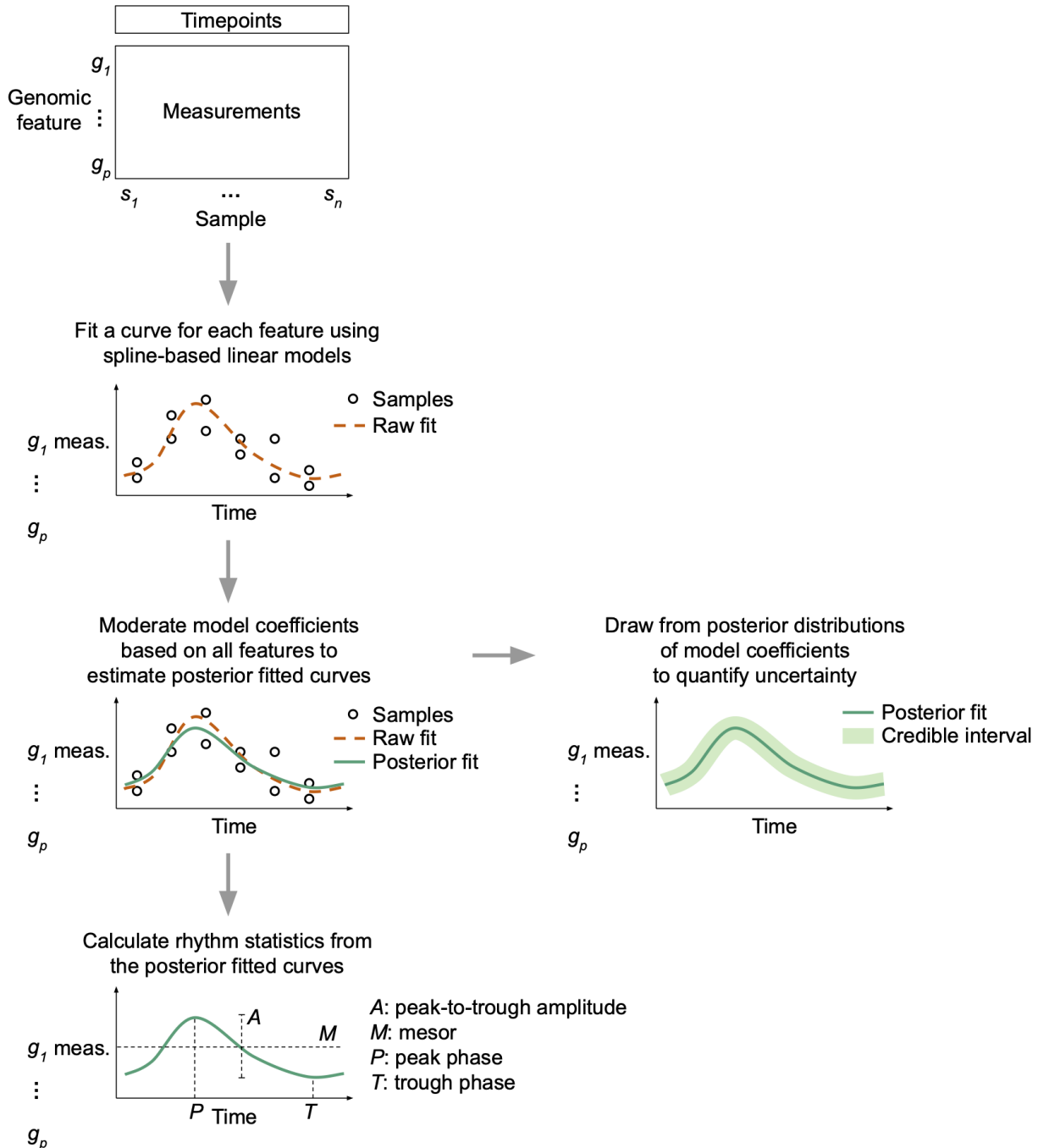
Schematic overview of LimoRhyde2's approach to quantifying rhythmicity in genomic data. Given a genomic feature (row) by sample (column) matrix of measurements, LimoRhyde2 fits a curve (*dashed orange line*) based on periodic splines describing how each feature's measurements change as a function of time. To adjust for noise and uncertainty in the fits,

LimoRhyde2 uses Mash to moderate the model coefficients, yielding a posterior fit (*solid green line*) for each feature. Using the posterior fits and their distributions, LimoRhyde2 calculates rhythm statistics and credible intervals.

Figure 2
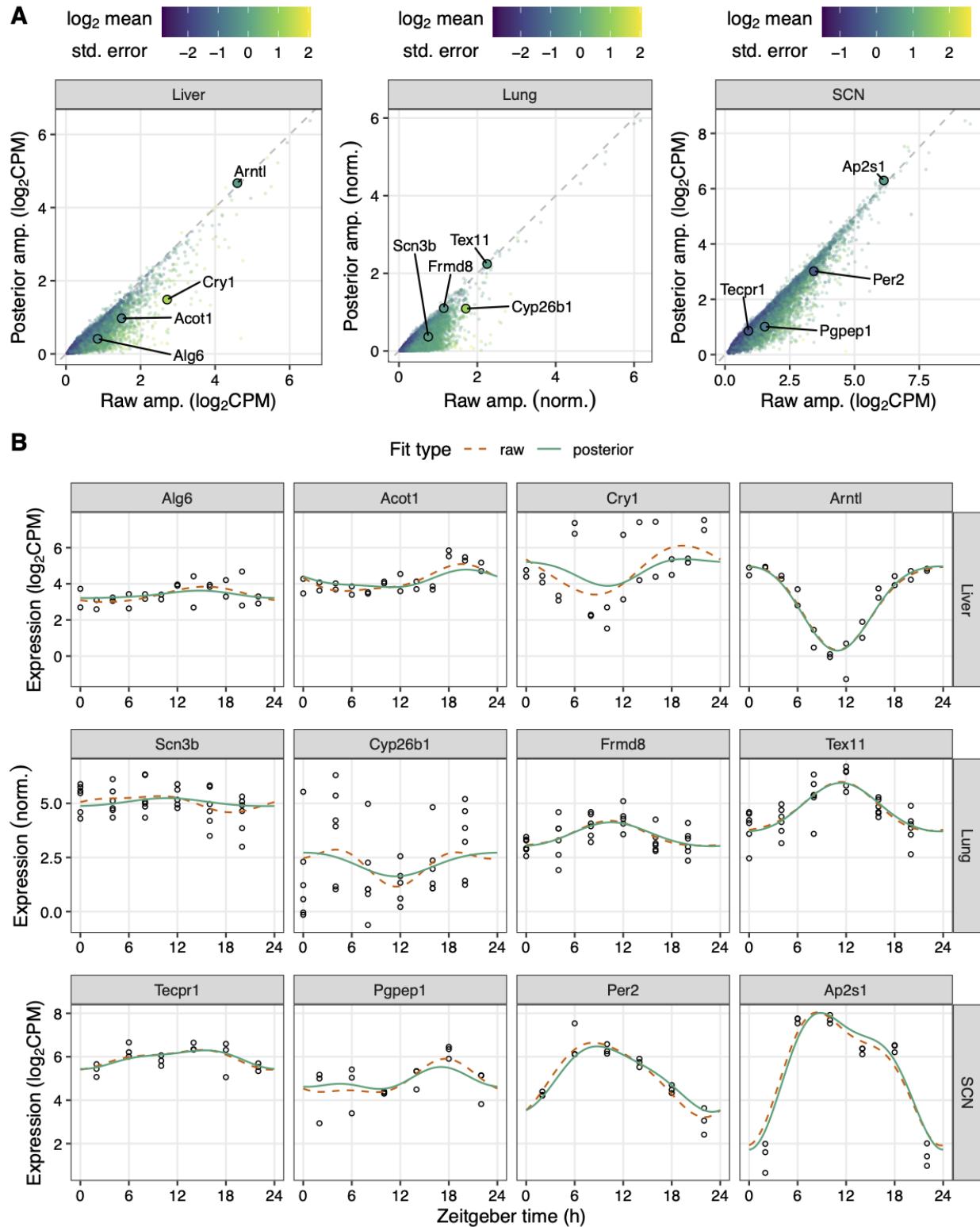
LimoRhyde2 uses raw fits and their standard errors to obtain posterior fits. (A) Scatterplots of posterior peak-to-trough amplitude vs. raw peak-to-trough amplitude for

transcriptome data from mouse liver (GSE67305), lung (GSE59396), and SCN (GSE72095). *Points* represent genes, *color* represents $\log_2$ mean standard error of the gene's raw fit. *Dashed lines* indicate y = x. **(B)** Time-courses of expression of genes (sorted by posterior amplitude) labeled in (A) in the respective tissue. *Curves* represent fits calculated by LimoRhyde2. *Points* represent samples.

Figure 3
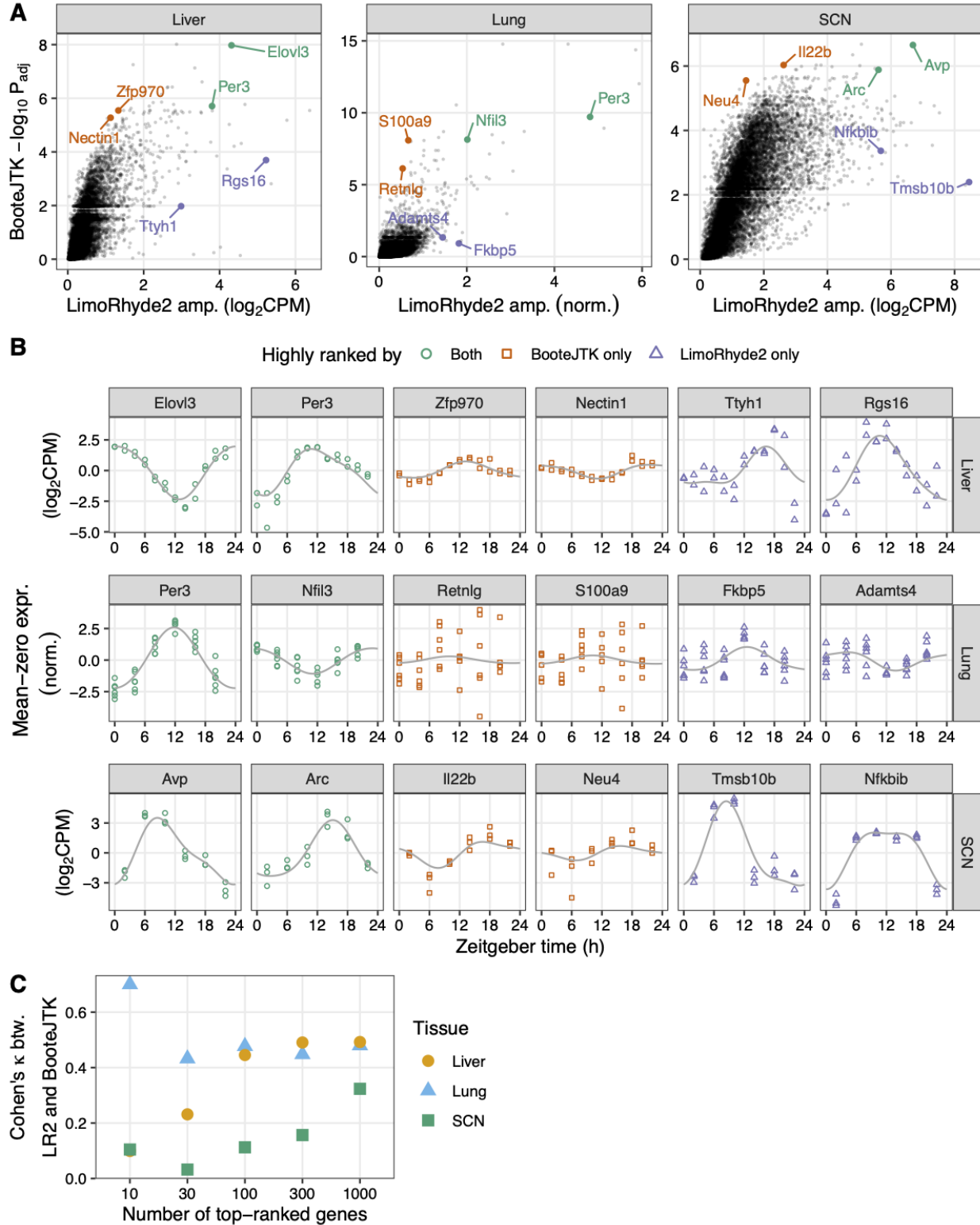
LimoRhyde2 prioritizes high-amplitude rhythms, compared to BooteJTK. (A) Scatterplots of -$\log_{10}$ adjusted p-value calculated by BooteJTK vs. posterior peak-to-trough amplitude

calculated by LimoRhyde2 for each tissue (indicated at *top*). *Points* represent genes. Genes towards the top are prioritized by BooteJTK, genes to the right are prioritized by LimoRhyde2. **(B)** Time-courses of expression of genes (indicated at *top*) labeled in (A) in the respective tissue (indicated at *right*). *Points* represent samples. *Colors* represent relative ranking of the genes according to the two methods. *Curves* represent posterior fits calculated by LimoRhyde2. Each gene's expression is centered at zero to highlight differences in amplitude. **(C)** Interrater agreement, as quantified by Cohen's κ, of top-ranked genes based on LimoRhyde2 posterior amplitude and BooteJTK -$\log_{10}$ adjusted p-value. *Shape* and *color* represent tissue.

Figure 4
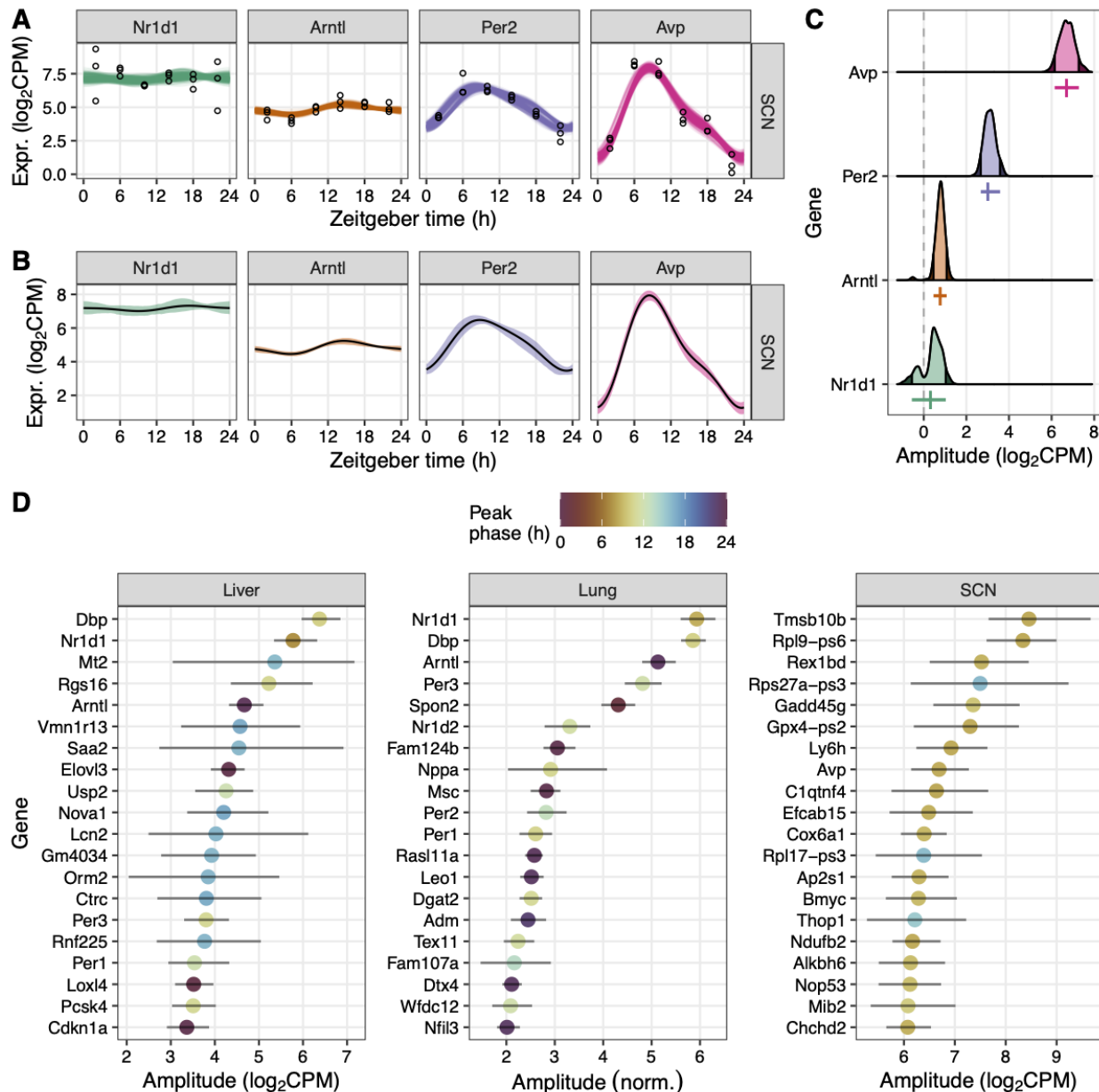
LimoRhyde2 quantifies uncertainty in rhythmicity using credible intervals. (A) Time-courses of expression for selected genes (indicated by *color* and at *top*) in the SCN. *Points* represent samples. *Curves* represent fits from 200 draws from the posterior distributions. **(B)** Time-courses of expression, where *lines* represent posterior means and *ribbons* represent 90% credible intervals calculated from the posterior draws. **(C)** Density plots of peak-to-trough amplitude based on the posterior draws. *Dashed line* indicates 0 amplitude. *Shaded regions* represent lower and upper bounds of the 90% credible intervals. *Vertical colored lines* represent posterior means of amplitude. *Horizontal colored lines* represent 90% credible intervals. **(D)** Amplitudes and corresponding 90% credible intervals for the top 20 genes ranked by amplitude in each tissue (indicated at *top*). *Color* represents peak phase.
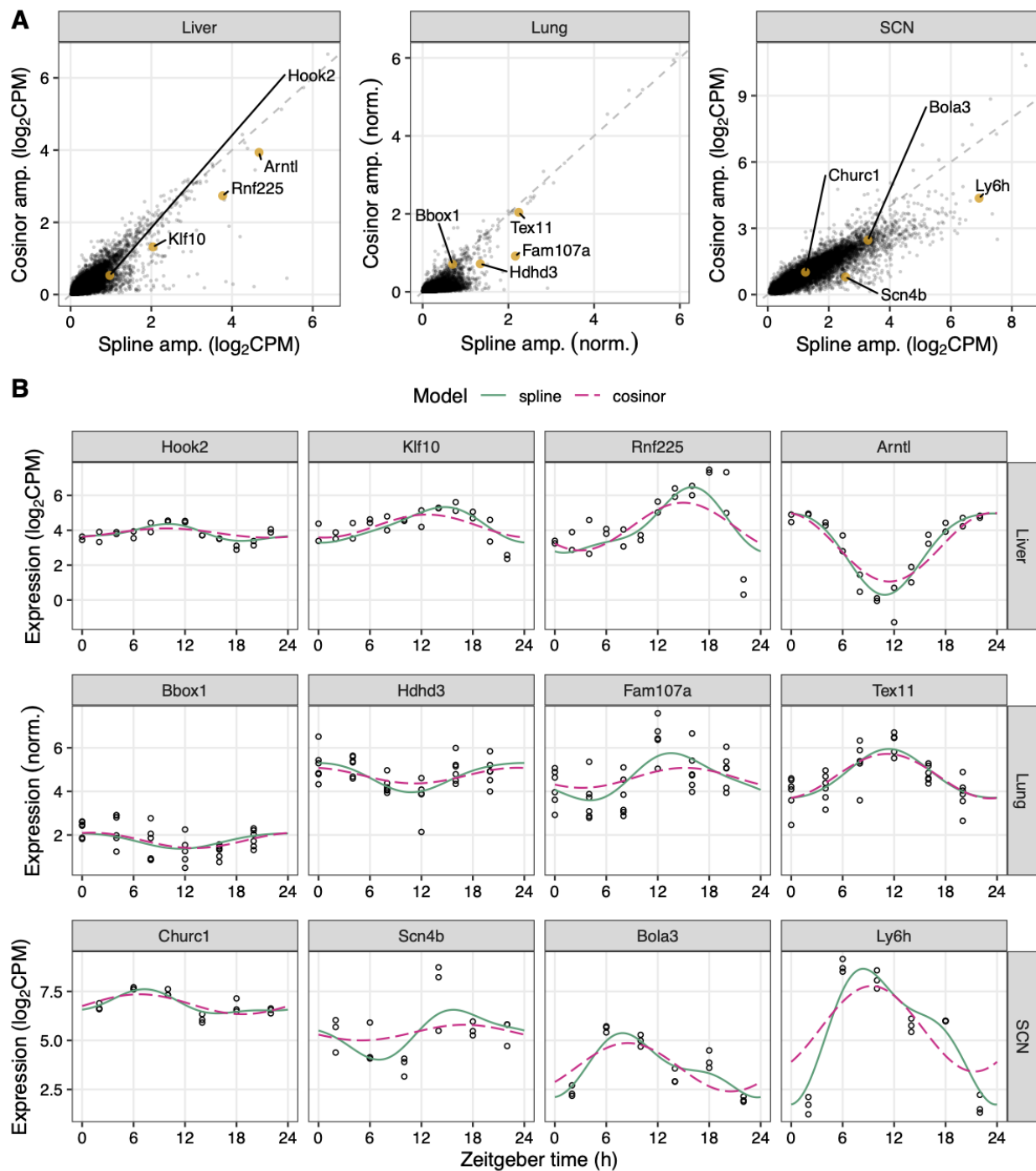
Figure S1
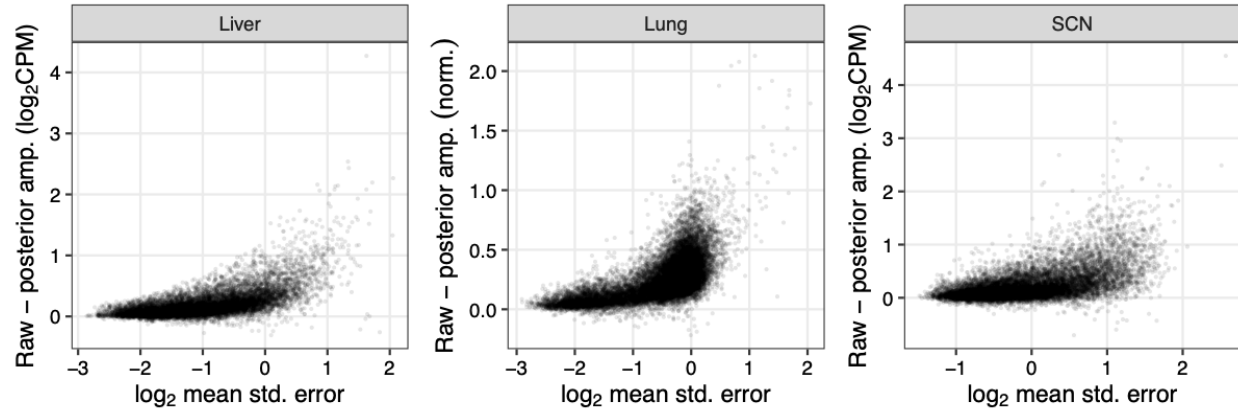
LimoRhyde2 moderates amplitude based on standard errors of genes. Scatterplots of difference between raw and posterior peak-to-trough amplitude vs $\log_2$ mean standard error of the raw fit for genes in each tissue. *Points* represent genes.

Figure S2

LimoRhyde2's spline-based model is more flexible and tends to give higher amplitude than the cosinor model. (A) Scatterplots of cosinor posterior peak-to-trough amplitude vs. spline posterior peak-to-trough amplitude for each tissue (indicated at *top*). *Points* represent genes. *Dashed lines* indicate y = x. **(B)** Time-courses of expression of genes labeled in (A) in the respective tissue (indicated at *right*). *Points* represent samples. *Curves* represent posterior fits for the two models.

Figure S3
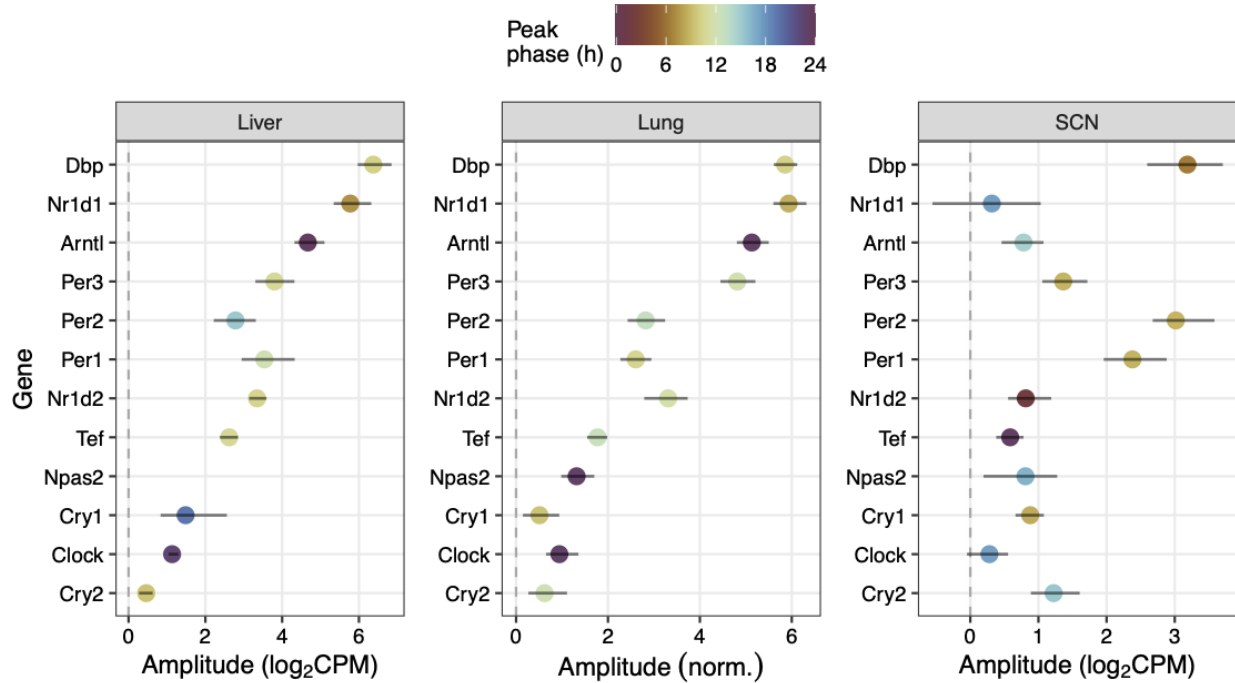
LimoRhyde2 identifies generally strong rhythms of core clock genes. Posterior peak-to-trough amplitudes and corresponding 90% credible intervals for core clock genes in each tissue. *Points* represent genes, *color* represents peak phase for each gene. *Dashed lines* indicate 0 amplitude.

Figure S4
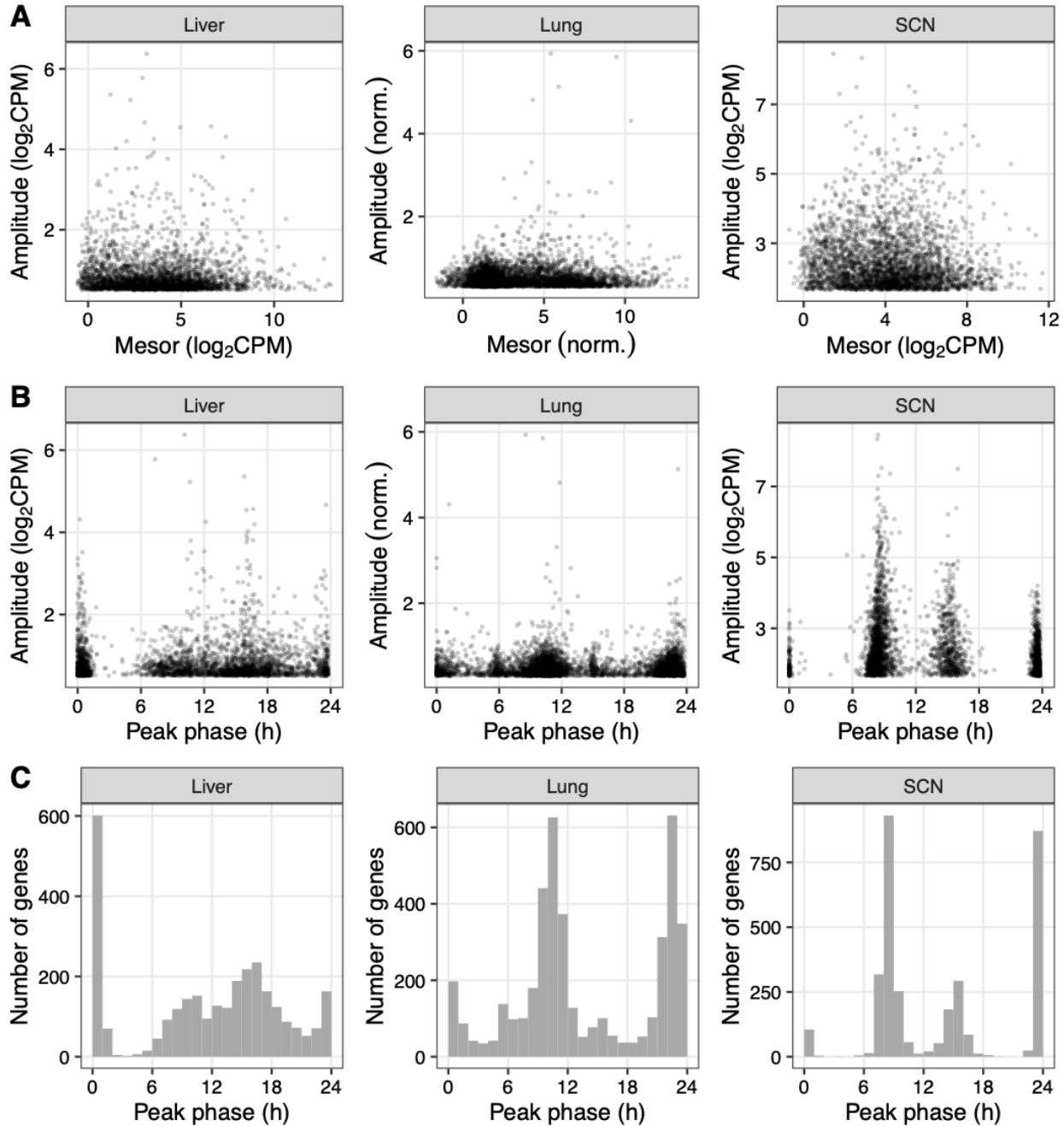
Distributions of rhythmicity based on LimoRhyde2 posterior statistics. Scatterplots of *(A)* peak-to-trough amplitude vs. mesor and **(B)** peak-to-trough amplitude vs. peak phase for genes in each tissue (indicated at *top*). *Points* represent genes. **(C)** Histograms of peak phase. All plots include only the top 25% of genes based on amplitude.