

Supplemental Figures for:

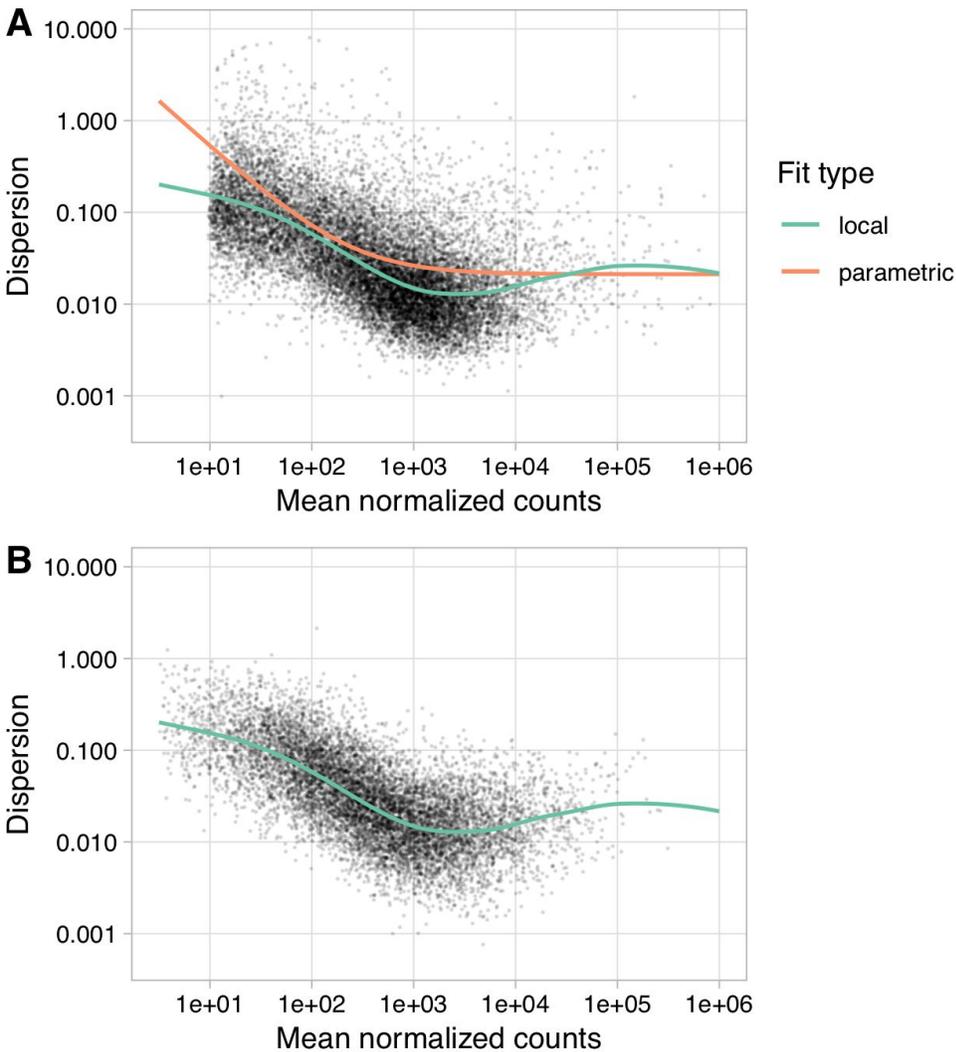
## Simphony: simulating large-scale, rhythmic data

Jordan M. Singer<sup>1</sup>, Darwin Y. Fu<sup>1</sup>, Jacob J. Hughey<sup>1,2,\*</sup>

<sup>1</sup>Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee; <sup>2</sup>Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee

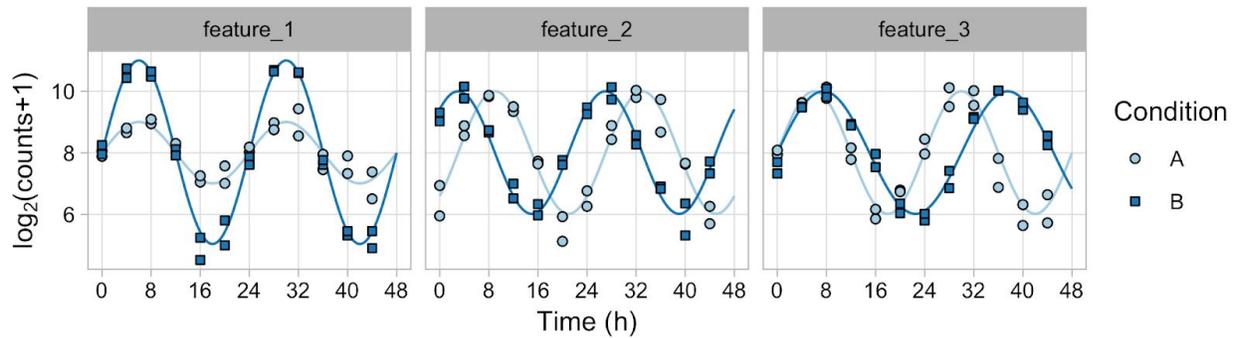
\*To whom all correspondence should be addressed: [jakejhughey@gmail.com](mailto:jakejhughey@gmail.com)

Figure S1



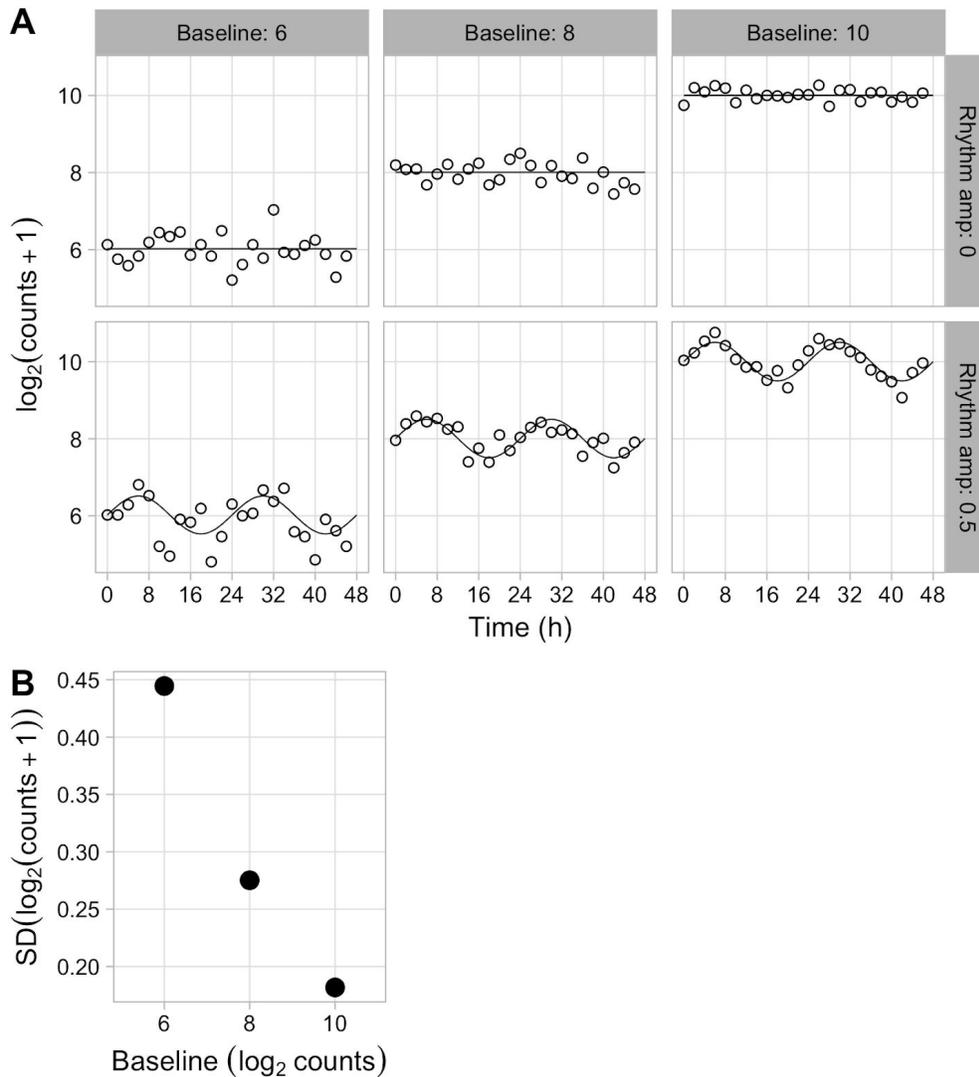
Estimating distributions of mean and dispersion and the relationship between the two. **(A)** Scatterplot of gene-wise means and dispersions in RNA-seq data from mouse liver (PRJNA297287), along with parametric and local regression-based curves, calculated using DESeq2. The curve based on local regression was chosen as the default in Symphony. **(B)** Scatterplot of sampled means and dispersions ( $n = 10,000$ ), with values drawn from the estimated distributions of mean normalized counts and residual log dispersion (around the local regression-based curve).

Figure S2



Examples of differentially rhythmic gene expression data generated by Symphony. Gene expression values were sampled from the negative binomial family. Rhythms followed a sinusoid. The baseline  $\log_2$  counts was 8. Time points in each of two conditions were spaced 2 h apart, with 2 replicates per time point. Colored circles and squares show the sampled gene expression values in each condition. Colored lines show the expected expression over time in each condition (feature\_1 shows a difference in amplitude, feature\_2 shows a difference in phase, feature\_3 shows a difference in period).

Figure S3



Examples of non-rhythmic and rhythmic gene expression data generated by Symphony, with various values of baseline expression and rhythm amplitude. The interval between time points was 2 h, with 1 replicate per time point. **(A)** Each plot shows the time-course for one simulated gene. Circles show the sampled expression values (from the negative binomial family) and black lines show the expected expression over time. **(B)** Standard deviation of  $\log_2(\text{counts} + 1)$  for non-rhythmic genes for each value of baseline expression.