Stochastic modeling of variability in circadian rhythms utilizing measured variance

Alan L. Hutchison, Elzbieta Kula-Eversole, Fangke Xu, Ravi Allada, and Aaron R. Dinner

Circadian rhythms exist in all kingdoms of life. Computational models of circadian clocks incorporate molecular and behavioral period measurements of phenotypes of clock mutations. Stochastic models of clocks exist to understand the role of noise, yet none attempt to explain experimental variances in period. We develop an improved method for detecting rhythmicity in genome-wide time series, which we apply to a rich RNA-Seq dataset in *Drosophila.* We identify novel rhythmic genes and test knockouts to identify circadian phenotypes. We use these results to explore the effects of stochasticity on predicting period variance and compare our simulation to prior results.

Background

Circadian rhythms are endogenous rhythms with approximately 24-hour periods. In *Drosophila,* a series of transcriptional feedback loops creates the rhythm-generating core clock. Mutations in the core clock can lead to changes in the period of oscillations and strength of rhythmicity, which can be measured by an activity profile (actogram) of organismal behavior over time.

Many methods exist for detecting rhythms in time series data. When data are noisy, sparse, and contain many false positives, as genome-wide time series data does, successfully detecting rhythms becomes more difficult. One leading method is JTK_CYCLE [1], which uses non-parametric correlations with reference waveforms to detect rhythmicity. This method is limited by an overly conservative multiple hypothesis test correction and by only using symmetric reference waveforms.

Many models for the core circadian clock exist that attempt to replicate measured mean protein, transcript, and behavioral dynamics in wild type and mutant phenotypes. Most of these models are deterministic [2-4]. However, measurements produce distributions of these different dynamics, which can only be simulated via stochastic models. Stochastic models of differing complexity do exist [5,6], but no known detailed stochastic models attempt to match the variance observed in the behavioral period.

Results

We develop an improved method for detecting rhythmicity in circadian genome-wide data by using asymmetric reference waveforms to identify asymmetric rhythmic time series and using Monte Carlo simulations to empirically correct the p-values for multiple hypothesis testing [7]. We show that this gives greater sensitivity and specificity for rhythm detection in comparison to six other methods, including the original JTK_CYCLE method.

We apply our method to *Drosophila melanogaster* RNA-Seq data to identify novel cycling genes. We test knockouts of newly identified circadian genes for changes in circadian behavior. We compare these phenotypes to our stochastic models of the core clock network, with particular attention to the variance in period as well as the mean value. Reflecting a previously unappreciated asymmetry in period variance in actogram experiments, we find that the period distribution in our simulations tends to skew below the mean value in the deterministic models. We find that care must be taken to accurately assign the period of oscillation in a stochastic simulation while taking into account circadian arrhythmia, which affects a sizeable fraction of core clock mutants.

To inform our understanding of the role of noise in our models, we add noise selectively to each species to observe how it propagates through the whole clock. We compare these results to clock mutants as well as mutants of the newly identified circadian genes to verify our predictions of the mechanism by which these genes affect the core clock.

Conclusion

We conclude that empirically correcting for multiple hypothesis testing and searching for asymmetric waveforms provides improved rhythm detection over other methods. We find that using stochastic simulations to explicitly model the distribution of the period in organismal circadian activity provides a useful means of understanding the effects of circadian mutants on the core clock. Future work includes applying our methods to better understand the temperature-independence (temperature compensation) of circadian rhythms and phenotypic mutants.

References

[1] Hughes ME, Hogenesch JB, Kornacker K. (2010) "JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets." J Biol Rhythms. Oct;25(5):372-80.

[2] Tyson JJ, Hong CI, Thron CD, Novak B. (1999) "A simple model of circadian rhythms based on dimerization and proteolysis of PER and TIM." Biophys. J. 77:2411-2417.

[3] Ruoff P, Christensen MK, Sharma VK. (2005) "PER/TIM-mediated amplification, gene dosage effects and temperature compensation in an interlocking-feedback loop model of the Drosophila circadian clock," J Theor Biol. 237:1:41-57.

[4] Bagheri N, Lawson MJ, Stelling J, Doyle III FJ. (2008) "Modeling the Drosophila melanogaster circadian oscillator via phase optimization," J Biol Rythms, 23:525-37

[5] Gonze D, Goldbeter A. (2006) "Circadian rhythms and molecular noise," Chaos, 16:2:026110.

[6] Bagheri N, Taylor SR, Meeker K, Petzold LR, Doyle III FJ. (2008) "Synchrony and entrainment properties of robust circadian oscillators," J R Soc Interface, 5:1:517-28.

[7] Hutchison AL, Maienschein-Cline M, Chiang AH, et al. (2015) "Improved Statistical Methods Enable Greater Sensitivity in Rhythm Detection for Genome-Wide Data." PLoS Comput Biol (in press).