

Quantifying stochastic noise in cultured circadian reporter cells

Peter C. St. John¹ and Francis J. Doyle III^{1,*}

¹Department of Chemical Engineering, University of California Santa Barbara, Santa Barbara, California 93106-5080

*Email: doyle@engineering.ucsb.edu

February 17, 2015

Running head:

Quantifying circadian stochastic noise

Keywords:

Systems Biology | Circadian Rhythms

Gene Regulatory Network | Stochastic | Synchronization

Introduction

Circadian rhythms in mammals are daily changes in gene expression and physiology that persist even in the absence of external environmental cues (6). Such rhythms are generated by a large network of interacting regulatory elements, in which time-delayed negative feedback gives rise to sustained oscillations (17). The functional roles of different species in circadian regulation have traditionally been studied using behavioral-level data and genetic knockout experiments (18). Bioluminescence-based cellular circadian reporters offer a more direct view of the gene regulatory network (2), and are amenable to high-throughput screens, allowing genome-wide exploration into factors which affect circadian rhythmicity (21). Additionally, cultured circadian reporter cells allow the change in transcriptional amplitude following a perturbation to be quantified. This additional parameter has proven useful in differentiating between perturbations with the same effect on period (15), and has led to the search for small-molecule therapeutics which boost clock amplitude (4).

Transcription at the single-cell level is strongly affected by the low molecular counts of the mRNA and protein species involved. As a result, bioluminescence traces of individual cells are stochastic, with significant period-to-period variability (19). The collective behavior of thousands of cells results in more reliable oscillations at the tissue-level, especially in the suprachiasmatic nucleus (SCN) where cell-to-cell coupling keeps individual oscillators at a consistent phase (7). While communication between cells is common in many tissues, it is currently thought that coupling between circadian oscillations outside the SCN, such as in peripheral tissues or cultured reporter cells, is very weak, if present (5, 11). In populations which lack cell-to-cell coupling, stochastic noise at the single-cell level is manifested in damped oscillations at the population-level (19). Despite the averaging which occurs at the population-level, stochastic noise at the single-cell level plays an important role in determining the function of the circadian oscillator. A recent study showed that stochasticity is critical to the population-level response to a neuropeptide and forms the basis for how the SCN entrains to light-mediated cues (1). Additional studies have even argued that the basis of single-cell rhythmicity depends on stochastic noise, as models of deterministically damped oscillators - when simulated stochastically - better capture the noise characteristics seen in single-cell fibroblast data (20).

Despite the importance of single-cell stochasticity in circadian rhythms, measuring stochastic noise currently requires careful plating and recording of fibroblast cells and subsequent image processing (10). As a result, while circadian perturbations have been postulated to affect single-cell stochasticity (13), no study has experimentally quantified changes to single-cell noise as a result of a small molecule or genetic perturbation. In this study, we demonstrate single-cell stochasticity can be reliably inferred from the damping rate of population-level bioluminescence recordings. Additionally, we show that a small-molecule modulator is able to change circadian noise in a dose-dependent fashion. Finally, we calculate the genome-wide effects of siRNA knock-down on single-cell noise, and demonstrate that population-level damping rate is independent of other circadian parameters, such as period or amplitude. Our results should prove especially important in the future search for small molecule circadian therapeutics, as it allows the effect of candidate drugs on single-cell noise to be quantified in a high-throughput manner.

Materials and Methods

Fitting a damped sinusoid to experimental data

A damped sinusoid, specified by:

$$\hat{y}(t) = Ae^{-dt} \sin\left(\frac{2\pi t}{T} + \theta\right)$$

is fit to experimental data $x_i(t_i)$, $i \in \{0, \dots, N-1\}$. The x_i data points are first detrended using Hodrick-Prescott filter with a smoothing parameter $\gamma = 0.05 (24 \text{ hrs}/s)^4$, in which s is the sampling rate (in hours) (12). The detrended data is then filtered using a low-pass filter to remove high-frequency noise (forward-backward Butterworth filter with $n = 5$, $w_c = 0.1$). We denote the detrended and filtered experimental data by $y_i(t_i)$. For numerical efficiency, the period, T , and damping rate, d , parameters are fit first using a matrix pencil method (9), reviewed in (22). Amplitude, A , and phase, θ , parameters are subsequently fit using a linear least-squares regression. Overall R^2 values for the regression were calculated from the residual error between the detrended data and fitted sinusoid:

$$R^2 = 1 - \frac{\sum_{i=0}^{N-1} (y_i(t_i) - \hat{y}(t_i))^2}{\sum_{i=0}^{N-1} (y_i(t_i) - \bar{y}(t_i))^2}$$

Processing single-cell bioluminescence data

Single-cell bioluminescence data for 79 cells was obtained from Leise *et al.*, 2012 (10). As was done in the original study, a discrete wavelet transform (using PyWavelets, <http://www.pybytes.com/pywavelets>) was performed to detrend and remove noise.

Sorting cells by noise level

As in the original study, various parameters describing the average noise level of each cell were collected. Traces were denoised and detrended by keeping only the (8hr, 258hr) wavelet components. From these smoothed trajectories, a Hilbert transform was used estimate points at which the phase crossed 0 to find period and amplitude coefficients of variation. An additional noise parameter, the standard deviation in the (1hr, 8hr) wavelet components divided by the overall rhythm amplitude, was used to quantify the high-frequency noise of the system. From these three noise metric, a combined noise variable was constructed via a principle component analysis (using scikit-learn, <http://scikit-learn.org/>). Cells were ranked according to this combined noise metric, and a high-noise group and low-noise group were constructed by taking the 39 highest-noise and lowest-noise cells, respectively. The raw bioluminescence profiles were not initially synchronized, so in order to simulate the gradual desynchronization of a group of oscillators the traces were offset to have the same starting phase. This was accomplished by starting each trace at the first phase zero-crossing, which were found using a Hilbert transform.

Bootstrap estimations of the damping rate difference

Averaged traces for low and high-noise group displayed a damped sinusoidal rhythm. The first 4 days of rhythms ($s = 0.5$, $N = 192$) were fit using a damped sinusoid. To ensure the difference in damping rate between groups was statistically significant, a bootstrap analysis was performed. In each of 10,000 bootstrap trials, the cells were randomly assigned evenly to either the low-noise

or the high-noise group. The absolute difference in damping rate between the two populations was recorded to yield a two-tailed test. The observed test statistic, $|d_h - d_l| = 6.65 \times 10^{-3}$, was found to be significant at the $\alpha = 0.05$ confidence threshold ($p = 0.0264$).

Quantifying dose-dependent effects of small molecule modulators

Bioluminescence traces ($s = 1.67$, $N = 71$) from the application of increasing concentrations of small molecule were fit using the method described in a previous section. Because the small molecules were toxic to the cells at very high concentrations, experiments were removed from further analysis when the R^2 of the sinusoidal regression fell below 0.80.

In silico prediction of small molecule experiments

A previously published mathematical model of circadian rhythms was used to predict the effects on population damping rate from the dose-dependent small molecule experiments (8). The parameters used to capture the effects of each small molecule were the same as described previously (15). The model was converted to a stochastic biochemical system and subsequently simulated using StochKit2 (14). Population-level rhythms were found by taking the average of 1,000 noninteracting oscillators, starting at identical initial conditions. The only parameter left unspecified by the deterministic model was the cell volume Ω , which controlled the amount of noise in the system. For a given Ω , the population-level damping rate was found by finding the phase diffusivity parameter which gave the closest fit to the simulated population-level rhythms, as described in (16). An R^2 value was calculated for each fit, taking into account all eight state variables.

Fitting the volume parameter

We calculated an average experimental damping rate of $d = 0.0151$ from the $0\mu M$ bioluminescence trajectories for both KL001 and longdaysin. *In silico* damping rates were calculated for logarithmically space values of $\Omega \in (100, 500)$. Ten independent groups of 1,000 oscillators were simulated for each Ω , from which the means and standard errors were found. Simulations in which $R^2 < 0.90$ were removed from further analysis. A weighted least-squares regression (using statsmodels, <http://statsmodels.sourceforge.net/>) was performed for $\log d$ vs. $\log \Omega$, using the logSEM of each measurement as a regression weight. The best fit was found to be $\Omega = 226.3 \pm 9.0$.

Parameter knockdown experiments

We mathematically replicated the effects of the small molecules KL001 and longdaysin through the reductions of the *vdcn* and *vac1p* parameters, respectively. Simulations were performed with 20 values of each parameter, linearly spaced between 100% and 15% of their nominal value. Similar to the volume calibration simulations, 10 independent groups of 1,000 oscillators were simulated from an initially entrained state. Means and standard errors in period and damping rate were calculated from each trajectory. Simulations in which $R^2 < 0.90$ were removed from further analysis.

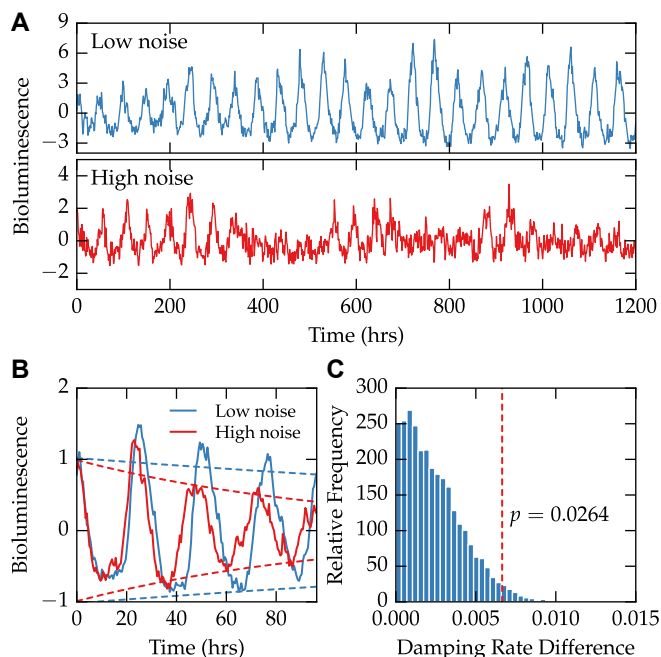


Figure 1: Figure Title. (A) Part A. (B) Part B. (C) Part C.

Fitting the genome-wide siRNA screen

We analyzed the supplemental annotations and data for the 111,743 wells ($s = 2$, $N = 72$) in the Zhang *et al.*, 2009 screen (21). Fits for which the $R^2 < 0.80$ were discarded. We rescaled amplitude values on a logarithmic basis, as these were more normally distributed. Parameters were normalized on a plate-by-plate basis using a robust z-score, median \pm median absolute deviation (3). We chose the “control” wells to be those which contained no siRNA, as these proved to be more numerous than those containing reference siRNA perturbations and were clustered similarly to the highest-density regions of the perturbed fits.

Detecting outlier perturbations

To find which target genes confidently shifted the fitted parameters away from the control population, we performed a test similar to a two-sample t-test with equal variance. Because the control distribution ($n = 11,253$) is much larger than that for any particular gene ID ($n \approx 4$), the outlier test is dependent almost entirely on the control distribution and the perturbed mean. We found the average perturbed response by grouping the perturbed dataset by target gene ID (using pandas, <http://pandas.pydata.org/>). To extend the t-test to two dimensions, we used a one-class support vector machine (SVM) with a radial basis function kernel ($\nu = 0.1$, $\gamma = 0.2$). The one-class SVM, after training with the control dataset, draws a decision boundary around 90% of the control points, flagging the remaining 10% ($n = 1124$) as outliers. The decision boundary was then used to classify the averaged perturbed points, from which 43% ($n = 3139$) were flagged as outliers.

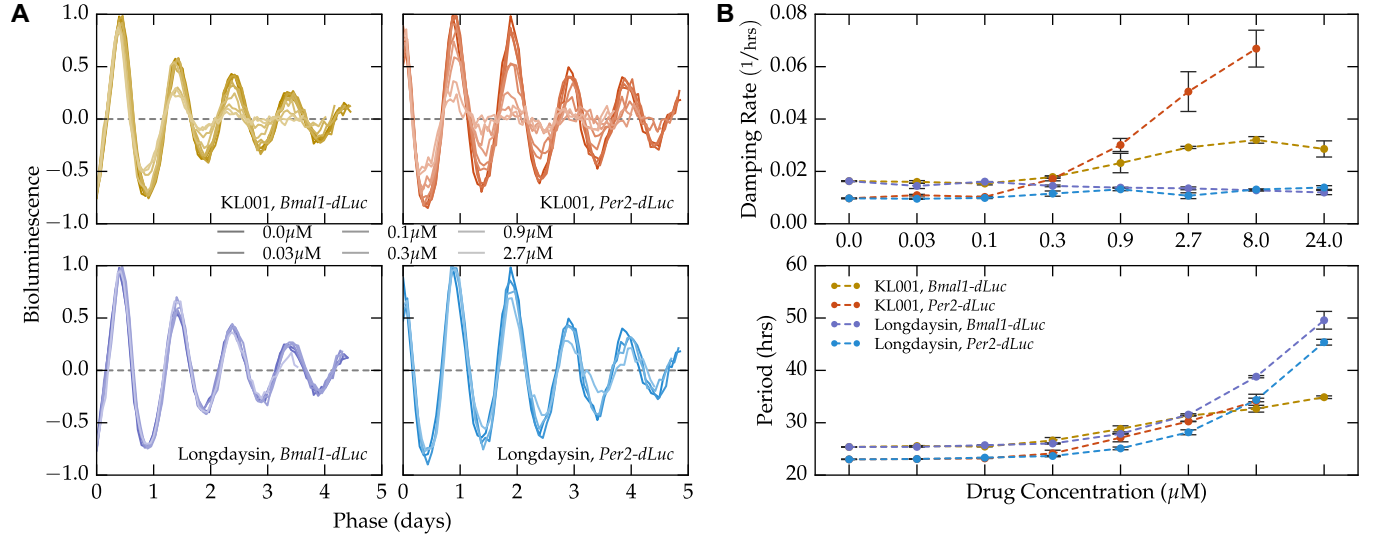


Figure 2: Figure Title. (A) Part A. (B) Part B.

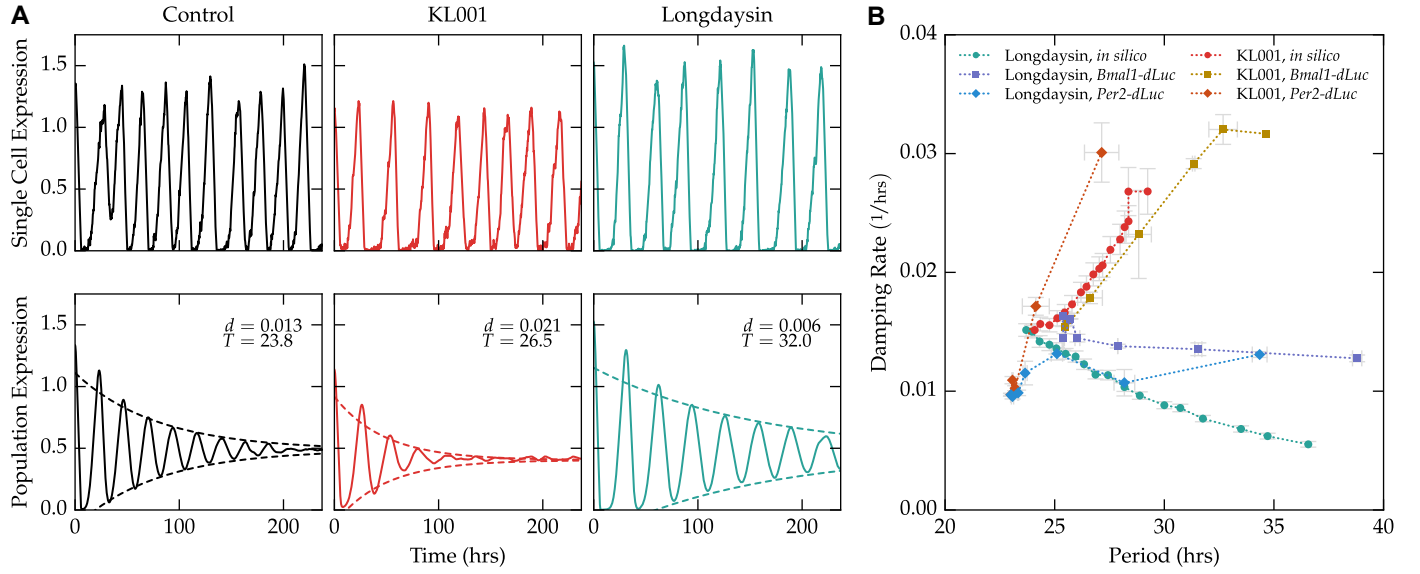


Figure 3: Figure Title. (A) Part A. (B) Part B.

	T		$\ln A$		d	
	C	P	C	P	C	P
μ	-0.234	0.187	0.443	-0.343	0.043	0.090
σ	0.774	1.820	0.778	1.753	0.878	1.688
Skew	0.153	0.367	-1.823	-0.580	-0.107	0.371
Kurt	3.772	0.591	8.329	0.476	2.423	0.373

Table 1: Fitted Parameters

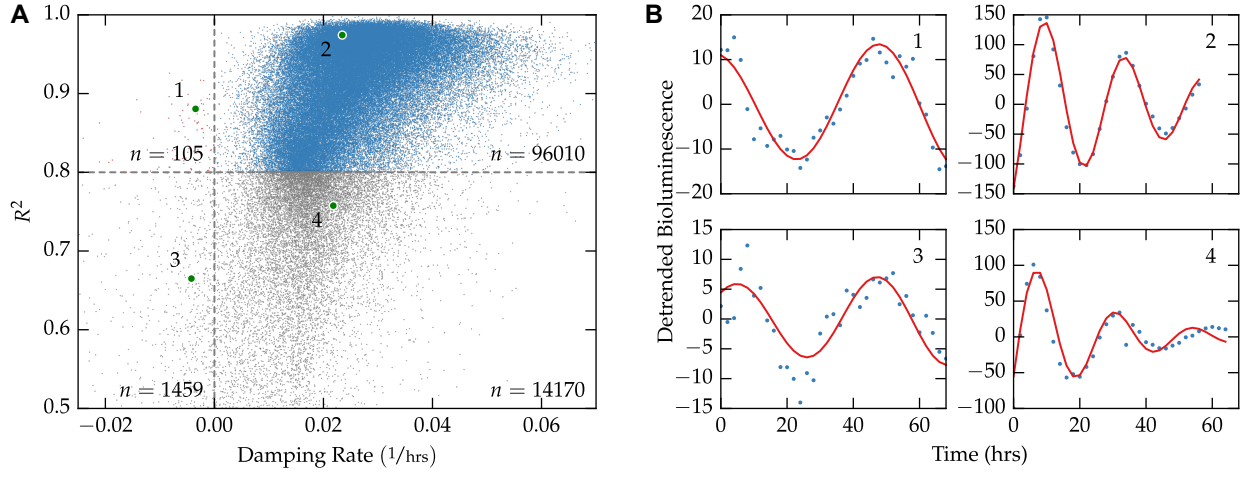


Figure 4: Figure Title. (A) Part A. (B) Part B.

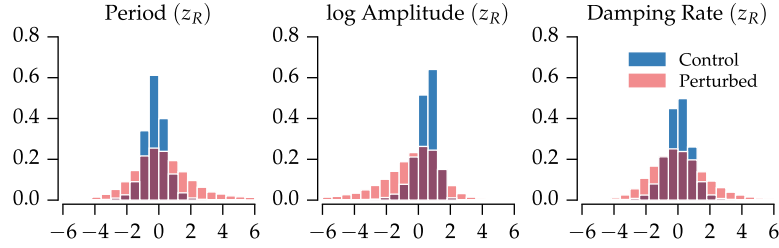


Figure 5: Figure Title. (A) Part A. (B) Part B.

	d	$\ln A$	T	θ
d	1.000	0.285	-0.142	-0.269
$\ln A$	0.285	1.000	-0.022	-0.112
T	-0.142	-0.022	1.000	-0.113
θ	-0.269	-0.112	-0.113	1.000

Table 2: Correlation matrix

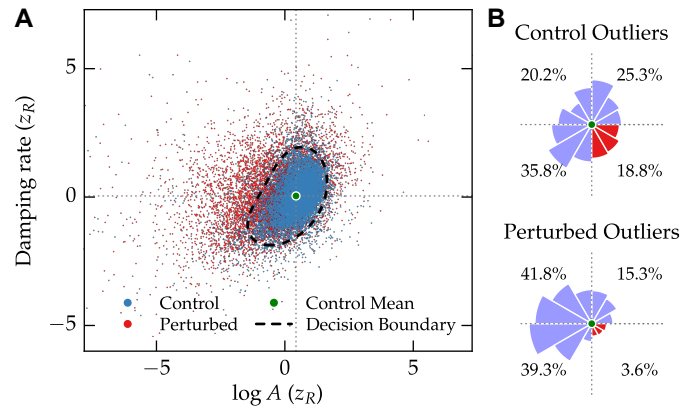


Figure 6: Figure Title. (A) Part A. (B) Part B.

Results and Discussion

Conclusion

Acknowledgments

This work was supported by the National Institutes of Health/National Institute of General Medical Sciences under award number 1R01GM096873-01 and by the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 from the U.S. Army Research Office.

References

1. An S, Harang R, Meeker K, Granados-Fuentes D, Tsai Ca, Mazuski C, Kim J, Doyle FJ, Petzold LR, Herzog ED (2013) A neuropeptide speeds circadian entrainment by reducing intercellular synchrony. *Proc Natl Acad Sci U S A* **110**: E4355–61
2. Balsalobre A, Damiola F, Schibler U (1998) A serum shock induces circadian gene expression in mammalian tissue culture cells. *Cell* **93**: 929–37
3. Birmingham A, Selfors LM, Forster T, Wrobel D, Kennedy CJ, Shanks E, Santoyo-Lopez J, Dunican DJ, Long A, Kelleher D, Smith Q, Beijersbergen RL, Ghazal P, Shamu CE (2009) Statistical methods for analysis of high-throughput RNA interference screens. *Nat Methods* **6**: 569–75
4. Chen Z, Yoo SH, Takahashi JS (2013) Small molecule modifiers of circadian clocks. *Cell Mol Life Sci* **70**: 2985–98
5. Guenther CJ, Luitje ME, Pyle La, Molyneux PC, Yu JK, Li AS, Leise TL, Harrington ME (2014) Circadian rhythms of Per2::Luc in individual primary mouse hepatocytes and cultures. *PLoS One* **9**: e87573
6. Herzog ED (2007) Neurons and networks in daily rhythms. *Nat Rev Neurosci* **8**: 790–802

7. Herzog ED, Aton SJ, Numano R, Sakaki Y, Tei H (2004) Temporal precision in the mammalian circadian system: a reliable clock from less reliable neurons. *J Biol Rhythms* **19**: 35–46
8. Hirota T, Lee JW, St. John PC, Sawa M, Iwaisako K, Noguchi T, Pongsawakul PY, Sonntag T, Welsh DK, Brenner DA, Doyle FJ, Schultz PG, Kay SA (2012) Identification of Small Molecule Activators of Cryptochrome. *Science* **337**: 1094–1097
9. Hua Y, Sarkar T (1990) Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Trans Acoust* **38**: 814–824
10. Leise TL, Wang CW, Gitis PJ, Welsh DK (2012) Persistent cell-autonomous circadian oscillations in fibroblasts revealed by six-week single-cell imaging of PER2::LUC bioluminescence. *PLoS One* **7**: 1–10
11. Noguchi T, Wang LL, Welsh DK (2013) Fibroblast PER2 circadian rhythmicity depends on cell density. *J Biol Rhythms* **28**: 183–92
12. Ravn M, Uhlig H (2002) On adjusting the Hodrick-Prescott Filter for the frequency of observations **84**
13. Rougemont J, Naef F (2007) Dynamical signatures of cellular fluctuations and oscillator stability in peripheral circadian clocks. *Mol Syst Biol* **3**: 93
14. Sanft KR, Wu S, Roh M, Fu J, Lim RK, Petzold LR (2011) StochKit2: software for discrete stochastic simulation of biochemical systems with events. *Bioinformatics* **27**: 2457–8
15. St. John PC, Hirota T, Kay SA, Doyle FJ (2014) Spatiotemporal separation of PER and CRY posttranslational regulation in the mammalian circadian clock. *Proc Natl Acad Sci U S A* **111**: 2040–5
16. St. John PC, Taylor SR, Abel JH, Doyle FJ (2014) Amplitude Metrics for Cellular Circadian Bioluminescence Reporters. *Biophys J* **107**: 2712–2722
17. Ueda HR, Hayashi S, Chen W, Sano M, Machida M, Shigeyoshi Y, Iino M, Hashimoto S (2005) System-level identification of transcriptional circuits underlying mammalian circadian clocks. *Nat Genet* **37**: 187–92
18. Vitaterna MH, King DP, Chang aM, Kornhauser JM, Lowrey PL, McDonald JD, Dove WF, Pinto LH, Turek FW, Takahashi JS (1994) Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior. *Science* **264**: 719–725
19. Welsh DK, Yoo SH, Liu AC, Takahashi JS, Kay SA (2004) Bioluminescence imaging of individual fibroblasts reveals persistent, independently phased circadian rhythms of clock gene expression. *Curr Biol* **14**: 2289–95
20. Westermarck PLO, Welsh DK, Okamura H, Herzel H (2009) Quantification of circadian rhythms in single cells. *PLoS Comput Biol* **5**: e1000580
21. Zhang EE, Liu AC, Hirota T, Miraglia LJ, Welch G, Pongsawakul PY, Liu X, Atwood A, Huss JW, Janes J, Su AI, Hogenesch JB, Kay SA (2009) A genome-wide RNAi screen for modifiers of the circadian clock in human cells. *Cell* **139**: 199–210
22. Zieliński TP, Duda K (2011) Frequency and damping estimation methods - An overview. *Metrol Meas Syst* **18**: 505–528

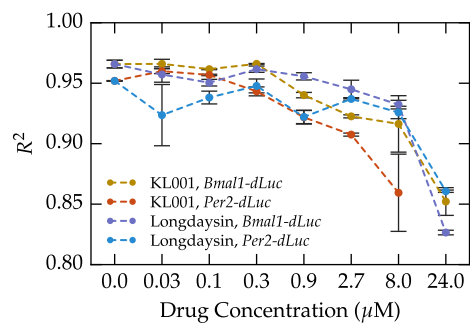


Figure S1: Figure Title. (A) Part A. (B) Part B.

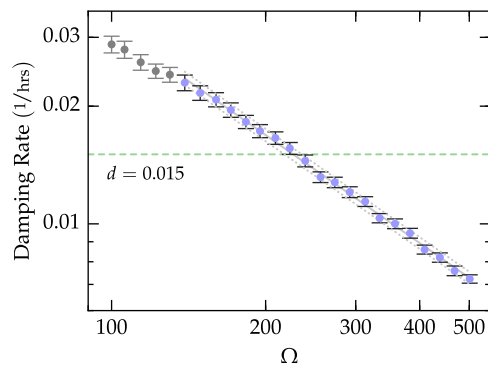


Figure S2: Figure Title. (A) Part A. (B) Part B.

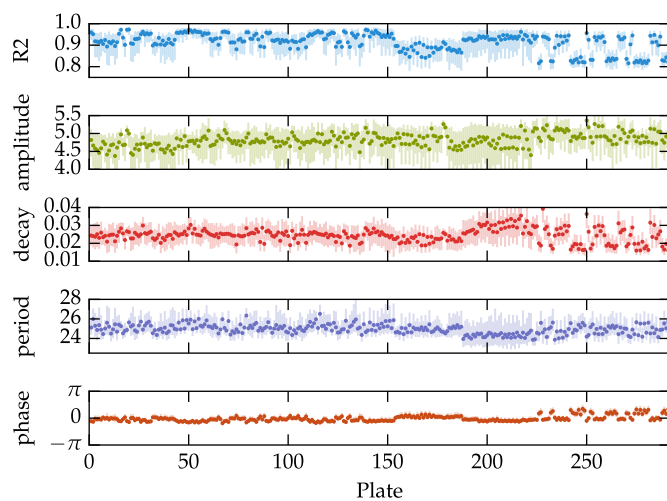


Figure S3: figure title. (a) part a. (b) part b.

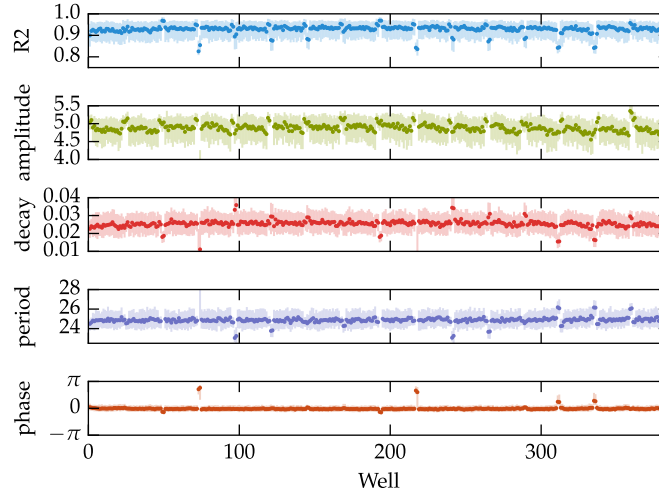


Figure S4: figure title. (a) part a. (b) part b.

Dep. Variable:	Damping Rate	R-squared:	0.169
Model:	OLS	Adj. R-squared:	0.169
Method:	Least Squares	F-statistic:	4782.
Date:	Wed, 11 Feb 2015	Prob (F-statistic):	0.00
Time:	16:26:22	Log-Likelihood:	$-1.7248e + 05$
No. Observations:	94053	AIC:	$3.450e + 05$
Df Residuals:	94048	BIC:	$3.450e + 05$

	coef	std err	t	P> t	[95.0% Conf. Int.]
const	-0.0370	0.014	-2.572	0.010	-0.065, -0.009
amplitude	0.2375	0.003	86.282	0.000	0.232, 0.243
period	-0.1521	0.003	56.798	0.000	-0.157, -0.147
phase	-0.2354	0.003	85.598	0.000	-0.241, -0.230
type	0.3197	0.015	20.664	0.000	0.289, 0.350

Omnibus:	9769.391	Durbin-Watson:	1.876
Prob(Omnibus):	0.000	Jarque-Bera (JB):	18459.719
Skew:	0.697	Prob(JB):	0.00
Kurtosis:	4.664	Cond. No.	8.34

Table S1: OLS Regression Results