

Systems biology

Normalized coefficient of variation (nCV): a method to evaluate circadian clock robustness in population scale data

Gang Wu , Lauren J. Francey, Marc D. Ruben and John B. Hogenesch*

Divisions of Human Genetics and Immunobiology, Center for Circadian Medicine, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA

*To whom correspondence should be addressed.

Associate Editor: Can Alkan

Received on July 30, 2021; revised on September 22, 2021; editorial decision on October 13, 2021; accepted on October 27, 2021

Abstract

Summary: Robust oscillation of clock genes is a core feature of the circadian system. Relative amplitude (rAMP) measures the robustness of clock gene oscillations but only works for longitudinal samples. We lack a method for estimating robust oscillations from human samples without labeled time. We show that the normalized coefficient of variation (nCV) of 10 clock genes is linearly correlated with their normalized rAMP, independent of time labels. We found that the mean nCV of clock genes are consistently decreased in tumors compared to nontumors, suggesting a new therapeutic target in cancer treatment by enhancing clock robustness. nCV can provide a simple measure of the clock robustness in population-level datasets.

Availability and implementation: The nCV package (<https://github.com/gangwug/nCV>) and web application (<https://github.com/gangwug/nCVapp>) are available on the GitHub repository.

Contact: john.hogenesch@cchmc.org

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

The feedback loop of the clock gene network drives robust clock gene oscillations (Gonze *et al.*, 2002; Ukai and Ueda, 2010). Clock gene knockouts (e.g. *Arntl*, *Per1/2* and *Cry1/2*) and disease states (e.g. cancers, metabolic syndrome and sleep disorders) can reduce (Anafi *et al.*, 2017; Bae *et al.*, 2001; Bungler *et al.*, 2000; Kume *et al.*, 1999) robustness of clock gene oscillations (hereafter referred to as clock gene robustness). Therefore, it is important to measure clock gene robustness, which can be thought of as relative amplitude (rAMP, the magnitude of the wave). The rAMP calculation requires time labeled samples (Wu *et al.*, 2020b). However, most human samples lack time information. There are currently no methods to directly estimate clock gene robustness in human samples when the time of collection is unknown.

Here, we show that the normalized coefficient of variation (nCV) is linearly correlated with normalized rAMP for clock genes in longitudinal mouse and human datasets. Importantly, nCV can be applied to *unordered* datasets. Therefore, we applied nCV to analyze multiple cancer types and found that clock robustness is consistently reduced in tumors compared to adjacent nontumor samples.

2 Materials and methods

The rAMP is the ratio between amplitude and baseline (Wu *et al.*, 2020b), which are calculated by MetaCycle or CYCLOPS for each gene. The nrAMP is the ratio between the rAMP of a gene and the mean rAMP of all genes. The nCV is the ratio between the coefficient of variation (CV) of a clock gene and the mean CV of all genes. All datasets used in this study are available in the GEO database and FireBrowse (Supplementary Table S1). All statistical analyses were performed in R.

3 Results

Many lines of evidence suggest that the clock is dysregulated in cancer [reviewed in Sancar and Van Gelder (2021)]. Clock function is a network property that can be assessed by measuring phase relationships of multiple clock genes. On the other hand, clock gene robustness refers to the magnitude of circadian oscillation and is measured at the single-gene level by rAMP. rAMP indicates more robust circadian oscillation of *ARNTL* in normal human breast

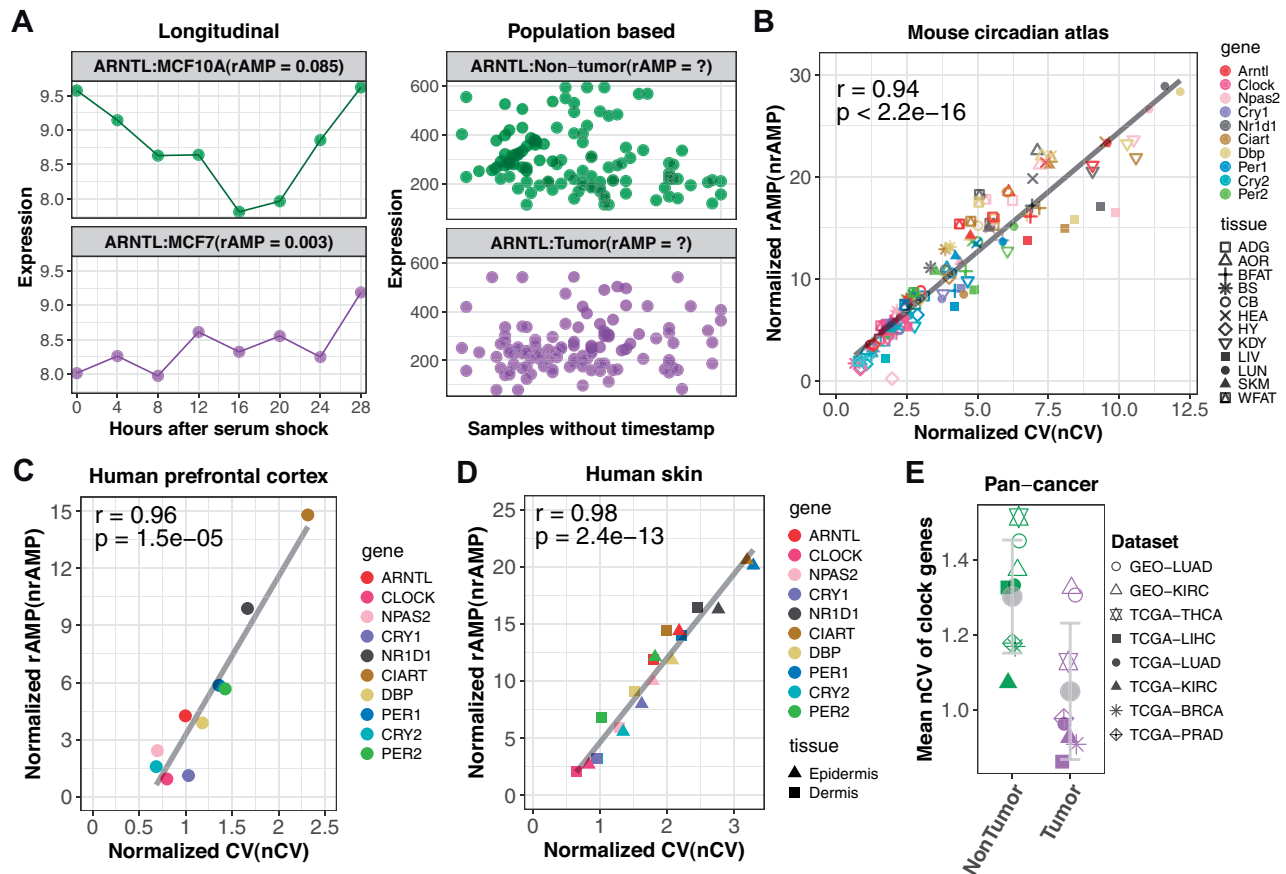


Fig. 1. nCV measures clock robustness. (A) rAMP is uninformative when comparing human breast tumors versus adjacent nontumor samples without labeled time. nCV of 10 clock genes is linearly correlated with nrAMP in mouse tissues (B), in human prefrontal cortex samples with labeled time (C) and in CYCLOPS ordered human epidermis and dermis samples (D). (E) The mean nCV of 10 clock genes from nontumor (green) and tumor samples (purple) per dataset (Supplementary Table S1). Gray point represents the average value in nontumor or tumor group, with error bar indicating \pm SD (Color version of this figure is available at *Bioinformatics* online.)

epithelial cells (MCF10A; rAMP = 0.085) compared to breast cancer cells (MCF7; rAMP = 0.003) (Fig. 1A, left panel). The rAMP calculation requires the sample collection time (Fig. 1A, right panel). However, the vast majority of patient datasets available do not have recorded collection times. Other metrics, such as the clock correlation distance (Shilts et al., 2018), can indicate clock function in the absence of collection time, but cannot measure clock gene robustness (Supplementary Fig. S1). Therefore, we need a way to estimate clock gene robustness in population data where sampling time is not recorded.

To address this, we selected a group of 10 clock genes that: (i) are important in circadian regulation, (ii) cycle in multiple tissues, (iii) are phased across the full circadian cycle and (iv) represent a range of rAMP values (Supplementary Fig. S2). Then, we searched for a time-independent measure of variance that is linearly correlated with rAMP. Median absolute deviation (MAD), standard deviation (SD) and the CV are common measures of transcript level variance. We found that the CV correlates with rAMP ($r = 0.96$) much better than with either MAD ($r = 0.35$) or SD ($r = 0.37$) for circadian genes in mouse liver (Supplementary Fig. S3). We used the normalized CV (nCV) to adjust for systemic differences in CV between datasets, and validated this new metric with the normalized rAMP (nrAMP) for a variety of datasets.

The nCVs of the 10 clock genes show strong linear correlation ($r = 0.94$ and $P < 2.2e-16$) to nrAMPs in 12 mouse tissues (Fig. 1B). Importantly, the nCV is reduced in several genetic knockout models of the clock (*Cry1/Cry2* double knockout mouse livers; *Arntl* knockout cells and mouse tissues; *Clock* knockout mouse kidney; *Nr1d1/Nr1d2* double knockout mouse livers) (Supplementary Fig. S4). The mean nCV of 10 clock genes (hereafter referred to as clock

robustness) in the knockout group is smaller (Wilcoxon test, $P = 0.01$) than the wild-type group (Supplementary Fig. S5). We next validated nCV in human samples with labeled time. nCV is strongly correlated with nrAMP in human prefrontal cortex samples (Fig. 1C; $r = 0.96$ and $P = 1.5e-05$) (Chen et al., 2016) and human skin samples ordered by CYCLOPS (Fig. 1D; $r = 0.98$ and $P = 2.4e-13$) (Wu et al., 2020a). In sum, nCV is a strong surrogate measure for nrAMP even when sample collection time is unknown.

The vast majority of human datasets do not have time of day information available. This includes most cancer datasets. Thus, we applied nCV to study clock gene robustness in tumor versus adjacent nontumor samples without labeled time. The majority of clock genes showed reduced nCV in tumor samples from patients with hepatocellular carcinoma, lung adenocarcinoma, clear cell renal cell carcinoma, breast invasive carcinoma and thyroid carcinoma (Supplementary Fig. S6). The nCV of *ARNTL* and *PER2* is consistently decreased in tumor samples from all eight tested datasets. In sum, clock robustness is reduced in tumors compared to nontumors in all eight datasets (Wilcoxon test, $P = 0.007$; Fig. 1E).

Studies report inconsistent changes in clock gene expression between different human cancer types (Savvidis and Koutsilieris, 2012; Ye et al., 2018). However, we show that clock robustness using nCV is consistently reduced across human cancers. Clock robustness may in fact be an important therapeutic target in cancer. For example, enhancing the circadian clock function in cancer cells inhibits tumor growth (Kiessling et al., 2017). Beyond cancer, nCV can be used to estimate clock robustness in any context where population scale data are available. In sum, nCV provides a straightforward measure of circadian clock function in the absence of labeled time, the vast majority of all human datasets.

Acknowledgements

We thank Dr Andrew Liu, Dr Eric Zhang and Dr Rafael Irizarry for helpful discussions about circadian amplitude. We thank Dr Ron Anafi for discussing these issues.

Funding

This work was supported by the National Cancer Institute [1R01CA227485-01A1 to Ron Anafi and J.B.H.]; National Institute of Neurological Disorders and Stroke [5R01NS054794-13 to Andrew Liu and J.B.H.]; and the National Heart, Lung and Blood Institute [5R01HL138551-02 to Eric Bittman and J.B.H.].

Data availability

The datasets were derived from sources in the public domain: [GEO, <https://www.ncbi.nlm.nih.gov/geo/>] and [FireBrowse, <http://firebrowse.org/>].

Conflict of Interest: none declared.

References

Anafi,R.C. *et al.* (2017) CYCLOPS reveals human transcriptional rhythms in health and disease. *Proc. Natl. Acad. Sci. USA*, **114**, 5312–5317.

- Bae,K. *et al.* (2001) Differential functions of mPer1, mPer2, and mPer3 in the SCN circadian clock. *Neuron*, **30**, 525–536.
- Bunger,M.K. *et al.* (2000) Mop3 is an essential component of the master circadian pacemaker in mammals. *Cell*, **103**, 1009–1017.
- Chen,C.-Y. *et al.* (2016) Effects of aging on circadian patterns of gene expression in the human prefrontal cortex. *Proc. Natl. Acad. Sci. USA*, **113**, 206–211.
- Gonze,D. *et al.* (2002) Robustness of circadian rhythms with respect to molecular noise. *Proc. Natl. Acad. Sci. USA*, **99**, 673–678.
- Kiessling,S. *et al.* (2017) Enhancing circadian clock function in cancer cells inhibits tumor growth. *BMC Biol.*, **15**, 13.
- Kume,K. *et al.* (1999) mCRY1 and mCRY2 are essential components of the negative limb of the circadian clock feedback loop. *Cell*, **98**, 193–205.
- Sancar,A. and Van Gelder,R.N. (2021) Clocks, cancer, and chronochemotherapy. *Science*, **371**, eabb0738.
- Savvidis,C. and Koutsilieris,M. (2012) Circadian rhythm disruption in cancer biology. *Mol. Med.*, **18**, 1249–1260.
- Shilts,J. *et al.* (2018) Evidence for widespread dysregulation of circadian clock progression in human cancer. *PeerJ*, **6**, e4327.
- Ukai,H. and Ueda,H.R. (2010) Systems biology of mammalian circadian clocks. *Annu. Rev. Physiol.*, **72**, 579–603.
- Wu,G. *et al.* (2020a) A population-based gene expression signature of molecular clock phase from a single epidermal sample. *Genome Med.*, **12**, 73.
- Wu,G. *et al.* (2020b) Genome-wide studies of time of day in the brain: design and analysis. *Brain Sci. Adv.*, **6**, 92–105.
- Ye,Y. *et al.* (2018) The genomic landscape and pharmacogenomic interactions of clock genes in cancer chronotherapy. *Cell Syst.*, **6**, 314–328.e2.