Compass in the data ocean: Toward chronotherapy

Rikuhiro G. Yamada[a] and Hiroki R. Ueda[a,b,1]

In the globalized modern society, the world is continuously moving 24 h a day 7 d a week. Prominent cities on the earth can be seen brightly lit at night from outer space. People on earth have daily (circadian) rhythms based more on their social circumstances than on natural cycles of day and night. This discrepancy of rhythms poses new challenges to our bodies that have never been met in evolutionary history. The circadian clock system regulates various aspects of physiology. Even a healthy person may suffer once internal rhythms break from environmental ones. The misalignment happens, for example, when a person travels on an airplane across many time zones (jet lag), when a person oversleeps on the weekend and wakes up early to go to work on Monday (social jet lag), or when a person works during the night or early morning with an intermittent or rotating schedule (shiftwork). Several reports have shown that internal body time varies by 5–6 h in healthy humans (1) and by as much as 10–12 h in shift workers (2). Accumulating evidence suggests that those misalignments may be a link to health risks, including obesity (3) and psychiatric disorders (4). Also, the severity of some diseases may be affected by the time of day, and some drugs are known to have different potency or toxicity depending on their administration time. The advancement achieved in the past two decades in the field of circadian biology has revealed that the core of the mammalian circadian clock is composed of about 20 transcription factors (5). Recently, a research group reported that a majority of mammalian genes are under the clock regulation, and that markedly different genes show circadian oscillation in each tissue (6). Importantly, they reported that a substantial number of top-selling drugs in the United States have circadian targets (6). Based on those findings, a convenient and precise molecular measurement of tissue molecular time is needed. The report published in PNAS by Anafi et al. (7) from the same research group strives to achieve this precise molecular measurement of tissue molecular time.

The conventional method to monitor body time is to take blood samples periodically over 24 h, measuring the level of hormones such as melatonin or cortisol, which have robust circadian oscillation in the blood (1, 8). However, this method ignores the underlying molecular machinery and differences between organs. A straightforward approach to address this problem is to perform time series sampling from each organ. However, subjecting people to tissue biopsies with a constant time interval for a couple of days is a huge burden on the people and is impractical. Choosing an easily accessible tissue such as hair follicles can reduce the burden, but this approach precludes investigation of internal organs. Another way is to develop a molecular timetable method that can estimate internal body time with just one or a few samples (9).

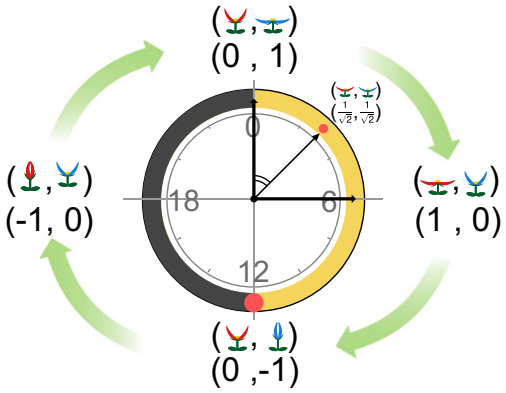

Fig. 1. Simplified numerical representation of Linné's flower clock. A hypothetical flower A is represented with red petals, and flower B is represented with blue petals around a 24-h clock. The openness of flowers is indicated as −1 (closed), 0 (half-open), or 1 (open). The 2D vectors that contain elements of flower openness are displayed below the corresponding flowers. The two thick black arrows form a basis of the plane, and any arbitrary vector of flower openness locates on the unit circle on the plane. A thin black vector is depicted as an example, and its phase (the time of day) is recapitulated as the angle between the axis and the vector (double arc). The yellow and black bands around the clock represent the day and night times of a day. In an actual situation, the dimension of the vector will be the number of biological molecules used in the analysis, or the number of nodes in the previous layer of the neural network.

[a]Laboratory for Synthetic Biology, RIKEN Quantitative Biology Center, 1-3 Yamadaoka, Suita, Osaka 565-0871, Japan; and [b]Department of Systems Pharmacology, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Author contributions: R.G.Y. and H.R.U. wrote the paper.

The authors declare no conflict of interest.

See companion article on page 5312.

[1]To whom correspondence should be addressed. Email: uedah-tky@umin.ac.jp.

The molecular timetable method was inspired by the Linné's flower clock, which is a conceptual clock named after Carl von Linné, the father of modern taxonomy. In the 18th century, Linné realized that several flowers open or close at a particular time of day, and proposed a list of flowers so that the combination of the flowers indicates the time of day. By substituting the flowers with good time-indicating molecules, we can construct a molecular timetable with which to infer internal body time. Molecules can be, but are not limited to, mRNA (9–12), metabolites (13–15), and protein (16). The fundamental step of the method is to find good time-indicating molecules. Because good time-indicating molecules are oscillatory with the circadian period, analytical techniques developed for searching for oscillatory molecules can be used, including the Pearson correlation between variation of molecular levels and the sinusoidal curve (9, 13, 14, 16), a nonparametric correlation test such as JTK_CYCLE (17), and singular value decomposition or its relatives (18, 19). However, the notable caveat of those analysis techniques is that the experiment should be well designed and controlled with the intention of detecting oscillatory molecules. Therefore, those analysis tools usually require a set of new experiments when the subject changes. The recently emerging machine learning approaches may mitigate this problem by training a model that can be applied to datasets taken from different organs (20). However, samples with known circadian time are required for the training. As Anafi et al. (7) point out, the Gene Expression Omnibus, a public gene expression database maintained by the National Center for Biotechnology Information, has a collection of more than 1 million human gene expression data, but most data lack the sampling time information. Hence, they embarked on a challenge to develop a method that could make use of the large-scale data by developing an unsupervised learning method, CYCLOPS (CYCLic Ordering by Periodic Structure).

First, they focused on genes that were detected as oscillatory in the mouse in the previous study (6). Then, they relied on a mathematical concept of oscillation. The basic concept they used in the work can be envisaged with the metaphor of Linné's flower clock. For simplicity, let us assume we have two imaginary flowers A and B, and that (*i*) "flower A" opens at 0600 hours and closes at 1800 hours, (*ii*) "flower B" opens at 0000 hours (midnight) and closes at 1200 hours (noon), and (*iii*) flowers A and B open or close gradually; that is, flower A is half-open at midnight and noon, and flower B is half-open at 0600 and 1800 hours (Fig. 1). Notice that it would be difficult to know the time of day by looking at just flower A or just flower B under this assumption. For example, when you see that flower A is half-open, you may need to sit for hours to see if it is opening (midnight) or closing (noon); that is, you need time series observation. Instead, by looking at both flowers, you would be able to know immediately that it is noon if flower B is also open. If you have more flowers with different opening or closing times, the clock becomes more precise. This simple setting can show another interesting property of oscillation. When the degree of openness of a flower is numerically represented (e.g., 1 for open, −1 for closed, 0 for half-open), we can form two vectors for time-indicating flowers. Each vector contains elements representing the openness at a different time of day (Fig. 1). It should be noted that two orthogonal vectors form a plane on which vectors of arbitrary time of day distribute on the unit circle. The angle between the vector and an axis indicates the time of day of the vector. This observation also holds when the vector is *n*-dimensional; that is, there are $n > 2$ flowers in a vector. This property of oscillation means that the openness of *n* time-indicating flowers at an arbitrary time of day can be encoded in a circle on 2D space. So, the problem of finding time-indicating molecules (flowers) can be rephrased as "What selection of molecules can make a plane on which projected vectors form a circle, or more generally, an ellipse?" Anafi et al. (6) addressed this problem by using unsupervised machine learning, where a neural network model was optimized so that the gene expression profiles were efficiently encoded on an ellipse on a plane. Because their method is an unsupervised approach, it was expected to work with datasets of unknown sampling time. Indeed, they successfully demonstrated that their method could detect aberrant circadian function in human hepatocellular carcinoma samples that were originally sampled without the intention of circadian analysis. Furthermore, they confirmed the circadian expression of a glucose transporter gene in human liver samples. In addition, based on this finding, they succeeded in providing a proof of concept, where the toxicity of a drug that interferes with the transporter could be reduced by controlling administration timing. The results of Anafi et al. may spark the dawn of the "Age of Discovery" in the data ocean of public databases, manifesting CYCLOPS as a useful compass to enhance the translation of knowledge in circadian biology to medicine.

The results of Anafi et al. may spark the dawn of the "Age of Discovery" in the data ocean of public databases, manifesting CYCLOPS as a useful compass to enhance the translation of knowledge in circadian biology to medicine.

Acknowledgments

We thank the laboratory members at RIKEN and Quantitative Biology Center, especially A. Millius for his critical reading and editing of the manuscript. This research was supported by grants from the strategic programs for R&D (President's Discretionary Fund) of RIKEN (to H.R.U.), an intramural grant-in-aid from the RIKEN Quantitative Biology Center (to H.R.U.), and a grant-in-aid for Challenging Exploratory Research (JSPS KAKENHI 26640021, to R.G.Y.).

1 Wright KP, Jr, Gronfier C, Duffy JF, Czeisler CA (2005) Intrinsic period and light intensity determine the phase relationship between melatonin and sleep in humans. *J Biol Rhythms* 20:168–177.
2 Horowitz TS, Cade BE, Wolfe JM, Czeisler CA (2001) Efficacy of bright light and sleep/darkness scheduling in alleviating circadian maladaptation to night work. *Am J Physiol Endocrinol Metab* 281:E384–E391.
3 Roenneberg T, Allebrandt KV, Merrow M, Vetter C (2012) Social jetlag and obesity. *Curr Biol* 22:939–943.
4 Wulff K, Gatti S, Wettstein JG, Foster RG (2010) Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease. *Nat Rev Neurosci* 11:589–599.
5 Ukai H, Ueda HR (2010) Systems biology of mammalian circadian clocks. *Annu Rev Physiol* 72:579–603.
6 Zhang R, Lahens NF, Ballance HI, Hughes ME, Hogenesch JB (2014) A circadian gene expression atlas in mammals: Implications for biology and medicine. *Proc Natl Acad Sci USA* 111:16219–16224.
7 Anafi RC, Francey LJ, Hogenesch JB, Kim J (2017) CYCLOPS reveals human transcriptional rhythms in health and disease. *Proc Natl Acad Sci USA* 114:5312–5317.
8 Weitzman ED, et al. (1971) Twenty-four hour pattern of the episodic secretion of cortisol in normal subjects. *J Clin Endocrinol Metab* 33:14–22.

9 Ueda HR, et al. (2004) Molecular-timetable methods for detection of body time and rhythm disorders from single-time-point genome-wide expression profiles. *Proc Natl Acad Sci USA* 101:11227–11232.
10 Akashi M, et al. (2010) Noninvasive method for assessing the human circadian clock using hair follicle cells. *Proc Natl Acad Sci USA* 107:15643–15648.
11 Lech K, et al. (2016) Evaluation of mRNA markers for estimating blood deposition time: Towards alibi testing from human forensic stains with rhythmic biomarkers. *Forensic Sci Int Genet* 21:119–125.
12 Laing EE, et al. (2017) Blood transcriptome based biomarkers for human circadian phase. *eLife* 6:6.
13 Kasukawa T, et al. (2012) Human blood metabolite timetable indicates internal body time. *Proc Natl Acad Sci USA* 109:15036–15041.
14 Minami Y, et al. (2009) Measurement of internal body time by blood metabolomics. *Proc Natl Acad Sci USA* 106:9890–9895.
15 Martinez-Lozano Sinues P, et al. (2014) Circadian variation of the human metabolome captured by real-time breath analysis. *PLoS One* 9:e114422.
16 Narumi R, et al. (2016) Mass spectrometry-based absolute quantification reveals rhythmic variation of mouse circadian clock proteins. *Proc Natl Acad Sci USA* 113:E3461–E3467.
17 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* 25:372–380.
18 Alter O, Brown PO, Botstein D (2000) Singular value decomposition for genome-wide expression data processing and modeling. *Proc Natl Acad Sci USA* 97:10101–10106.
19 Hughey JJ, Hastie T, Butte AJ (2016) ZeitZeiger: Supervised learning for high-dimensional data from an oscillatory system. *Nucleic Acids Res* 44:e80.
20 Agostinelli F, Ceglia N, Shahbaba B, Sassone-Corsi P, Baldi P (2016) What time is it? Deep learning approaches for circadian rhythms. *Bioinformatics* 32:i8–i17.