REVIEW ARTICLE

SPECIAL SERIES ON FOUNDATIONS OF CIRCADIAN MEDICINE

Revised: 4 February 2023

Data integration and analysis for circadian medicine

Lena Baum¹ | Marco Johns¹ | Maija Poikela¹ | Ralf Möller² | Bharath Ananthasubramaniam³ | Fabian Prasser¹

¹Berlin Institute of Health at Charité-Universitätsmedizin Berlin, Berlin, Germany

²Institute of Information Systems, University of Lübeck, Lübeck, Germany

³Institute for Theoretical Biology, Humboldt-Universität zu Berlin, Berlin, Germany

Correspondence

Bharath Ananthasubramaniam, Institute for Theoretical Biology, Humboldt-Universität zu Berlin, Berlin, Germany. Email: bharath.ananthasubramaniam@ hu-berlin.de

Fabian Prasser, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany. Email: fabian.prasser@charite.de

Abstract

Data integration, data sharing, and standardized analyses are important enablers for data-driven medical research. Circadian medicine is an emerging field with a particularly high need for coordinated and systematic collaboration between researchers from different disciplines. Datasets in circadian medicine are multimodal, ranging from molecular circadian profiles and clinical parameters to physiological measurements and data obtained from (wearable) sensors or reported by patients. Uniquely, data spanning both the time dimension and the spatial dimension (across tissues) are needed to obtain a holistic view of the circadian system. The study of human rhythms in the context of circadian medicine has to confront the heterogeneity of clock properties within and across subjects and our inability to repeatedly obtain relevant biosamples from one subject. This requires informatics solutions for integrating and visualizing relevant data types at various temporal resolutions ranging from milliseconds and seconds to minutes and several hours. Associated challenges range from a lack of standards that can be used to represent all required data in a common interoperable form, to challenges related to data storage, to the need to perform transformations for integrated visualizations, and to privacy issues. The downstream analysis of circadian rhythms requires specialized approaches for the identification, characterization, and discrimination of rhythms. We conclude that circadian medicine research provides an ideal environment for developing innovative methods to address challenges related to the collection, integration, visualization, and analysis of multimodal multidimensional biomedical data.

K E Y W O R D S

chronomedicine, data science, data integration, data visualization, time-series data

Bharath Ananthasubramaniam and Fabian Prasser contributed equally.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. Acta Physiologica published by John Wiley & Sons Ltd on behalf of Scandinavian Physiological Society.

Acta Physiologica

ACTA PHYSIOLOGICA

1 | INTRODUCTION AND OBJECTIVE

The circadian clock system is a network of tissue timekeepers ("clocks") that regulate most aspects of human physiology¹ and thus generate 24-h rhythms in many biological processes. The emerging field of circadian medicine leverages the interaction between the circadian clock system and human physiology in health and disease for prevention, diagnosis, and treatment.² To advance circadian medicine, there is a particularly high need for coordinated and systematic collaboration between researchers from different disciplines to uncover relationships between circadian biology and health, as well as to translate them into clinical practice. Data relevant to circadian medicine are multimodal, ranging from molecular circadian profiles to clinical parameters (also including what is often called physiological measurements) and data obtained from (wearable) sensors or data reported by patients.³ As rhythms are central to circadian medicine, large parts of the data are collected in the form of time series.

To study relationships between circadian properties and physiology, the circadian clock needs to be profiled across both time and space,⁴ all the while confronting the heterogeneity of humans and their clocks. This means that relevant data types must be integrated, visualized, and analyzed across different temporal resolutions ranging from milliseconds and seconds to minutes and up to several hours.⁵ Important challenges encountered when integrating such types of data include the need for specialized storage solutions, a lack of common standards, transformations needed for intuitive visualizations as well as data protection. On the data analysis side, challenges include the identification of periodic patterns, their characterization (e.g., in regards to phase, amplitude, period, and waveform), as well as their differential analysis across conditions. When studying human circadian rhythms there are additional challenges due to the variation of clock parameters in different dimensions and the inability to measure relevant clock tissue repeatedly, which means standard approaches need to be modified and populationbased approaches need to be adopted. The aim of this article is to provide an overview of the state-of-the-art in data integration and analysis for circadian medicine, to highlight challenges and possible directions for future work.

2 | IMPORTANT DATA TYPES IN CIRCADIAN MEDICINE

Research into and applications of circadian medicine require the measurement of potentially heterogeneous clock parameters and their influences on a range of physiological processes. As illustrated in Figure 1, this requires the integrated analysis of a broad spectrum of different data types.

Due to the cell-autonomous gene-regulatory basis for circadian rhythms, molecular data (omics data, reporter assays) are essential to quantify circadian rhythms. Indeed, not only have circadian rhythm disruptions been associated with human pathologies⁶ but also variations in clock-related genes and their protein products have been associated with increased or decreased risk of disorders.⁷ Timing of drug administration⁸ and timing of clinical procedures⁹ can determine the clinical outcome, and mining of routine clinical data¹⁰ can help generate new hypotheses for reverse translation. Wearable sensors as well as questionnaires administered through mobile devices are essential to capture patient-reported data on circadian rhythms, environmental factors, sleep quality, and functional status outside of the healthcare setting.³

2.1 | Molecular data

Modern molecular measurement technologies are key to identifying and explaining variation between individuals.¹¹ In circadian medicine, molecular data are usually collected over a 1- or 2-day period with sampling intervals of between 2 and 12h. Common types of molecular data include genomics, gene expression/transcriptomics, proteomics, metabolomics, and lipidomics data as well as time series from reporter assays. While transcriptomic and proteomic data provide a relatively focused readout of the clock output, metabolomic and lipidomic data provide a much more dynamic readout, which aggregates the molecular and metabolic state of the cell/tissue. Circadian rhythms are present in almost all cell types in humans¹² and regulate most physiological processes.¹ Thus, measurements from different tissues are relevant for studying different pathologies adding yet another dimension to the data. Although relatively new in circadian research, single-cell sequencing technologies allow profiling across cell types within a tissue,¹³ only some of which might be relevant to a pathology adding yet more dimensions to the molecular data. Omics data are also a great source to mine for new clock components¹⁴ and biomarkers for different clock parameters, such as the circadian phase.¹⁵⁻¹⁷ Circadian tissue cultured in vitro have also been shown to retain the circadian properties of the human donor.¹⁸ Thus, time series recordings from cell/tissue cultures transfected with a clock gene/protein reporter are another source of circadian molecular data. In summary, various types of molecular data are important for studying interactions between circadian rhythms and environmental,



behavioral, and lifestyle patterns as well as disease development and progression.5

2.2 **Clinical data**

Clinical data refers to all data from healthcare processes, clinical trials, and patient registries including data on diagnosis, treatment, drugs, laboratory tests, and physiological monitoring data collected at varying frequencies.¹⁹ While vital parameters, such as heart rate, core body temperature²⁰ (CBT), and blood pressure are usually measured in intervals of several minutes to hours, other measurements require a higher frequency and precision, such as electrocardiograms (ECG) or electroencephalograms (EEG), which have sampling rates of several milliseconds (see also next section). Circadian rhythmicity is commonly determined in the clinical setting from CBT measurements^{21,22} with the minimum of CBT serving as a phase marker. However, this requires a sufficiently high sampling rate of CBT.²³ Measurement of hormones relevant to circadian rhythms, cortisol, and melatonin, from saliva, urine, or blood is common for chronobiological profiling of subjects, which is essential for human circadian analysis (see Section 5.4); the current gold-standard phase marker, dim light melatonin onset (DLMO),²⁴ is based on melatonin. Similarly altered peaks or troughs of circadian blood pressure rhythms have long been used to flag worsened cardiovascular outcomes.²⁵ Thus, clinical data collected at different frequencies and over different lengths of time are essential for research in circadian medicine and the translation of its results.

2.3 Sensor data

Continuous data relevant to circadian medicine can come from many sources, ranging from patient monitoring systems in critical care to wearable sensors capturing data on circadian rhythms and environmental factors outside of healthcare settings and research facilities.⁵ Analogously, there is a broad scale of measurements that can be obtained, including information on light and sound, vital parameters, CBT, or EEG with sampling rates ranging from milliseconds to minutes or hours. Wearable sensors can measure many of the same physiological parameters collected under clinical conditions with a considerably lower subject burden allowing for longer monitoring times.²⁶ For studying sleep and circadian rhythms, there is a wide range

of metrics that can be collected from such devices, such as measurements of light exposure, geographical location, body temperature, heart rate, skin conductivity, and blood oxygen levels, which are all associated with the circadian system.²⁷ Actigraphy uses accelerometry to record activity patterns and body movements, such as exercise, and can be used to predict circadian phase as accurately as the clinical gold-standard DLMO assay.²⁸ For detecting sleep-wake patterns actigraphy is a sensitive and reliable instrument that has high agreement rates with polysomnography.²⁹ Wearable heart rate data might also provide a stable but independent phase marker compared to DLMO.³⁰

2.4 Patient-reported data

Patients can report their perceived health status in combination with data on circadian rhythms during their daily life in questionnaires. These patient-reported outcomes (PROs) can be generic, capturing functional status and quality of life, or be disease-specific and tailored toward specific conditions and symptoms. The PROs can be scored in so-called PRO measures (PROMs) in validated questionnaires.³¹ The common validated circadian and sleep assessments questionnaires are the Horne-Östberg Morningness-Eveningness, Seasonal Pattern Assessment, Pittsburgh Sleep Quality Index, and the Munich Chronotype Questionnaires,³ which are also commonly used for chronobiological profiling. Increasingly, such feedback from subjects or patients is obtained via custom-designed or generic wellness mobile apps.⁵ While patient-generated data provide important information on outcomes of circadian medicine, they are subjective by nature and care must be taken in using and interpreting them.³²

3 DATA INTEGRATION **CHALLENGES**

The comprehensive integration of relevant data types is an essential foundation for performing the multidimensional and multimodal data analyses required for the study of human circadian rhythms and the implementation of related medical applications. Figure 2 illustrates the major steps required for this purpose.

Data integration challenges range from efforts to access relevant data sources through existing interfaces to data standardization and harmonization as well as to provide



FIGURE 2 Schematic view of data generation, standardization, integration, and analytics for circadian medicine.

appropriate visualization and analysis capabilities. Along all those dimensions, there are cross-cutting requirements that need to be considered,³³ such as data quality management.¹⁹ Furthermore, circadian medicine is centered around data from patients and study subjects, and hence, requires adequate information security and privacy protection measures to be put in place.³⁴

3.1 Molecular data integration

Compared to other types of data, individual types of omics data are relatively well standardized with formats such as FASTQ being able to represent a wide range of molecular measurements³⁵ and projects such as Ensembl providing systems and ontologies of stable identifiers for genes, transcripts, proteins, and exons.³⁶ However, there is a lack of standards for multi-omics data.³⁷ In regard to tooling, a wide variety of solutions for multi-omics integration, analysis, or visualization have been developed and made available to the community.³⁸ However, not many tools specific to circadian data integration and visualization exist. A more general example is the cBioPortal, an open-source solution that allows users to explore and analyze large-scale cancer genomics data sets. It has recently been adapted to visualize data regarding the circadian regulation of cancer hallmarks.³⁹ Specific solutions for circadian medicine exist, but they often focus on a limited set of studies from certain labs and cannot easily be extended. Important examples are CircaDB,⁴⁰ which is a database of circadian gene expression profiles, and CircadiOmics,⁴¹ a large data repository and web-based analytics tool for high-throughput omic

circadian time series data. When aiming to support a broad range of circadian medicine projects, the integration with phenotypic data, the particularly high security and data protection requirements for many types of human omics data,³⁷ and the temporal alignment of multimodal data pose particular challenges. Additionally, the storage and processing of omics data are also very resource intensive.

3.2 **Clinical data integration**

Clinical data often suffers from a high degree of structural heterogeneity and a lack of standard terminologies.⁴² The reasons for this are manifold, ranging from the autonomy with which data are collected in different healthcare processes and research projects, to a lack of support by vendors and lack of awareness among researchers. While approaches such as common data models exist to cope with this challenge, data are usually transformed into such representations retrospectively, potentially leading to information loss due to fundamental heterogeneity that cannot be easily resolved.¹⁹ A common approach for integrating data for research purposes is to establish data warehouses, which are specific types of databases designed for analytical processing of heterogeneous data. To reduce the efforts required to transform data into common representations, pay-as-you-go approaches can be utilized.⁴³ These approaches are based on the principle of not aiming for a fully integrated dataset from the beginning, but instead incrementally harmonizing it into a common representation as it is needed, for example, for a circadian medicine research project together with domain experts.

This implies that efforts are invested when needed, leading to a more efficient integration process. Modern interoperability standards, like Health Level 7 (HL7) Fast Healthcare Interoperability Resources (FHIR),⁴⁴ have the potential to improve the homogeneity of clinical data, as they can foster their structured and standardized collection at the source and simplify their exchange with other clinical information systems and research contexts.45 Another issue, particularly important to circadian medicine research, is that timestamps of clinical samples or measurements might not be semantically consistent (e.g., time of sample collection, analysis, or result) making a joint analysis across centers and healthcare institutions challenging. Additionally, systemic biases, like the timing of clinical samples (usually collected in the morning, and at other times only due to urgent clinical concerns), or dosing time of medication may exist due to clinical processes and are confounding factors that need to be considered when integrating and interpreting population patterns.⁴⁶

3.3 | Sensor data integration

A particular challenge in regard to the integration of sensor data is the need to calibrate scales of measurements from different devices, analogous to what needs to be done to make laboratory values comparable across institutions. The challenge is aggravated by the rapid development of sensor technology, algorithms and methods, which need to be considered when integrating data from different devices, studies, and populations.³ Also, the standardization of the collected data can be an issue.⁴⁷ As with the other data types, multi-modal integration is necessary to put sensor data into context with other critical sources of information to facilitate their interpretation. In the context of remote patient monitoring and wearable sensors, raw sensor data (e.g., accelerometry) might be not easily accessible to many consumer-oriented devices due to proprietary components.^{3,27} In addition, it is important to apply data quality control and normalization techniques, as well as to validate consumer-oriented device data with their clinical equivalents before incorporation into clinical research.⁵ Due to the high frequency of the collection process, sensor data can also lead to unusual storage requirements. These challenges, while typical for data from wearable technologies and often relevant in study contexts, also apply to sensor data collection and integration in clinical routine settings.

3.4 | Patient-reported data integration

Capturing and integrating patient-reported data, which provide insights into quality-of-life and environmental factors,

Acta Physiologica

come with a range of associated challenges. On the data level, there is heterogeneity in how PROs are measured through various types of questionnaire instruments and are evaluated by the research community as well as heterogeneity in the data representations of questionnaire software and their licensing models.³ Recent standardization efforts, such as the inclusion of questionnaire structures into the HL7 FHIR model⁴⁵ and the representation of items from the Patient Reported Outcomes Measurement Information System (PROMIS)⁴⁸ as Logical Observation Identifiers Names and Codes (LOINC)⁴⁹ provide an important step forward. Still, the joint use and common analysis of patient-reported data usually require researchers to align regarding the instruments used upfront.⁵⁰ Collecting patient-generated data through apps and web-based surveys also comes with specific infrastructural requirements and information security challenges.⁵ The distinguishing characteristic of these types of data is that they are collected through an information channel into the domestic environment of the patients and study subjects. Such connections can be complex to implement for healthcare institutions, which often use highly protected and isolated IT environments due to the sensitivity of the data managed. This means that different security perimeters must be bridged during integration, which requires particular attention.⁵¹ Furthermore, the data might be influenced by various types of self-reporting bias⁵² or privacy concerns, which might lead to the patients providing incomplete or inaccurate information.

4 | MULTI-MODAL TIME SERIES DATA INTEGRATION, VISUALIZATION, AND PROTECTION

As outlined in the previous sections, studying circadian mechanisms requires integrating different types of data captured at different temporal resolutions. Additionally, since the study of circadian rhythms requires measurements to be collected over time, as well as over "space", that is, sample from different tissues, which are relevant to different aspects of physiology, large amounts of data are collected for the same modality. This leads to special requirements for biomedical integration architectures, which are often centered around relational data representations that are well suited for capturing longitudinal clinical data or around specific types of data stores, for example, Clinical Data Repositories for HL7 FHIR or systems for images or omics data.

Storing, querying, and analyzing multimodal time series data with different temporal resolutions is better supported by specialized data representations or specific

cta Physiologica

database management systems, such as time series databases. Bringing both worlds together can be challenging⁵³ with solutions ranging from modules for relational database management systems, such as TimescaleDB for PostgreSQL⁵⁴ to specialized database management systems, such as InfluxDB.⁵⁵ Also, integrating multimodal data is complicated by the fact that no common standards are able to capture all types of data in an integrated manner.⁵⁶ Since physiology is time-dependent, data obtained from different sources, collections, and modalities need to be temporally aligned and registered to allow for a meaningful analysis, which is especially true for molecular and sensor data collected in clinical settings, as well as from wearables technologies. Further, as they likely include data points sampled at different frequencies, specialized transformation steps are needed to convert variable-length time series into common, fixed-frequency representations without loss of information.⁵⁷ On the visualization side, a wide range of scalable open-source solutions have become available in recent years, many of which were originally developed for displaying monitoring and telemetry data about computer systems. However, effective visualizations can be challenging, because the display of time already accounts for one dimension of the 2-dimensional image, even making, for example, the visualization of a single RNA-seq experiment with multiple features difficult.

An important example of an open-source solution is Grafana,⁵⁸ which has recently also been adopted for visualizing a wide range of biomedical information. Examples of applications of Grafana in the healthcare and biomedical research context include the work by Drake et al,⁵⁹ which focuses on health-related data from smartphone sensors. the work by Cruz et al.⁶⁰ on real-time processing and visualization of intensive care data, and the work by Calhan et al.⁶¹, which focuses on patient monitoring. An example visualization and comparison of circadian rhythms with Grafana is provided in Figure 3, showing log₂ relative gene expression levels for CRY1, CRY2, DPD, FKBP5, NR1D2, PER1, and PER2 across four healthy male subjects.¹⁷ Using this and related frameworks for circadian medicine data may need the inclusion of dimensionality reduction techniques considering time dependencies.⁶² For example, composite phase deviations (CPD) density plots have been developed to support the visualization of circadian misalignment.63

Data protection is an important orthogonal topic when human biomedical data are collected, shared, and integrated for research purposes. Due to the need to integrate different types of data of the same individual as well as data about cohorts of selected individuals, integration architectures for circadian medicine require services for securely managing the identity of patients and study participants³⁴ as well as consistent pseudonymization⁶⁴ and data-linkage functionalities. While these requirements



FIGURE 3 Visualization of relative gene expression levels series across multiple subjects using Grafana. The data shows the gene expression in monocytes of four young healthy males under constant routine conditions.¹⁷

can be fulfilled using central services typically available at research institutions, there are some specific challenges arising from the nature of the data required. For example, secure communication channels are needed to collect patient-reported data and data from wearable sensors in remote settings (see above).⁵ Also, as medical data are highly sensitive, care must be taken to preserve the individuals' privacy.⁶⁵ For example, sharing time series data with other researchers in a privacy-preserving manner can be challenging if its dimensionality is high,⁶⁶ however, even coarse data does not guarantee privacy.⁶⁷ We therefore expect that research on circadian medicine will particularly benefit from innovative Deep Learning-based data synthetization mechanisms, that is, methods to produce artificial data from original data by training a model to reproduce its essential properties,⁶⁸ for time series data which is currently a very active field of research.

5 | DATA ANALYSIS FOR CIRCADIAN MEDICINE

Circadian rhythms are defined as (i) endogenously generated periodic patterns that repeat approximately every 24h under constant conditions, (ii) entrain to external stimuli, and (iii) are temperature compensated.⁶⁹ While these rigorously defined rhythms have been validated in model organisms, studies in humans typically test (i) and assume the rest hold true as well; no study to our knowledge has tested (i), (ii), and (iii) simultaneously in humans. Moreover, it is debatable whether rhythms under constant conditions ("circadian") or under normal periodic environments of light, sleep, feeding, and temperature ("daily/diurnal") are more relevant to circadian medicine or indeed biomedical research. We therefore take a broader view of circadian medicine in this review as encompassing physiology and medicine associated with the circadian clock rather than the strict definition given above. The dynamic nature of this phenomenon poses not only challenges to diagnostics but also to experimental design, data collection, and data analysis of circadian studies. The analysis of data generated typically involves the identification, characterization, and discrimination of rhythms, as illustrated in Figure 4A.

5.1 Experimental design

The study of dynamic phenomena such as circadian rhythms requires collecting time series of measurements. Experiments on circadian rhythms have to make important decisions regarding the total number of time samples, the spacing between samples, and the number

Acta Physiologica

for the design of circadian experiments. This is vital as statistically underpowered studies cannot conclusively test the desired hypotheses or detect desired effects. The analysis approaches and their statistical properties (see Section 5.2) too must be borne in mind to calculate effect and sample sizes.

There are consensus guidelines for the design of circadian experiments in model organisms.⁷¹ Sample sizes can be calculated for such situations using computational tools, such as TimeTrial⁷² for transcriptomic data. Nevertheless, key challenges remain to the design of human circadian medicine experiments due to the unavailability of effect sizes needed for power calculations and the lack of statistical tools to incorporate the heterogeneity of human clocks into the calculations.

5.2 | Identification of rhythms

The first step in any circadian analysis is to determine whether measurements show patterns of near 24-h periodicity. The standard approach to test this hypothesis requires a time series of measurements, that is, a sequence of measurements made at preferably regular intervals. The measurements must be made at sufficiently small sampling intervals and often multiple cycles of the periodic pattern are required to draw reliable conclusions. The approaches to identifying rhythms differ based on the type of data. Behavioral assays, such as actigraphy, produce long recordings at relatively high resolution, while omics time series data are typically short and often span only one cycle of the rhythm.

For long high-resolution time series, classical approaches include the χ^2 -periodogram⁷³ and a standard periodogram via a Fourier transform. A modified Fourier approach termed Lomb–Scargle periodogram⁷³ even works with missing time points and can be considered the best option. Wavelet analyses that resolve rhythms in both frequency and across the length of the time series can be applied to measurements that vary across the recording length, which is very common and often of interest in biological settings. Wavelet analysis needs multiple cycles of data at high resolution, but the approach is effective. Implementations of this sophisticated approach are available in easy-to-use software packages, such as CIRCADA⁷⁴ and pyBOAT.⁷⁵

For short time series, the statistical power can be improved with biological replicates. For this purpose, Cosinor analysis⁷⁶ is a fast, robust, and highly flexible



FIGURE 4 Overview of important steps when analyzing circadian data. (A) The four key steps in the execution of circadian studies in which analysis plays an important role. (B) When rhythms need to be compared across two different conditions, they must not be analyzed individually and then the results compared. This results in a high rate of false positives in the results. Instead, the rhythms should be compared directly across the two conditions (figure taken from Pelikan et al., 2022⁷⁰).

method that is based on linear regression. The only drawback of this approach for rhythm detection is the need for the user to specify the frequency (or cycle period) of interest. Cosinor analysis further assumes a sinusoidal waveform with normally distributed noise in the measurements.⁷⁷ More complex waveforms can be identified by Cosinor analysis using more frequencies at the cost of estimating more parameters (which needs more measurements) and ambiguously defined circadian parameters.⁷⁷ However, rhythms with harmonic periods (6, 8, and 12 h) are also 24 h periodic patterns and are often overlooked. These possibly physiologically important⁷⁸ rhythms can be detected using a different cosinor formulation.⁷⁹ Two other methods circumvent some or most of these assumptions. The very popular JTKcycle algorithm⁸⁰ assumes only that the waveform is sine-like. A related approach, RAIN⁸¹ and an improvement on JTKcycle's statistical properties, eJTKcycle,⁸² relax this assumption even further to allow other asymmetric waveforms. Since the previous three approaches are only concerned with the ordering of values, they do not explicitly incorporate measurement uncertainty, a shortcoming that the BooteJTK algorithm rectifies in the context of genomewide expression studies.⁸³ The latter four approaches can search for rhythms over a range of periods and can be applied to any data type with almost no knowledge of the underlying distribution of the measurements. The computational burden of calculating p-values increases steadily from JTKcycle, eJTKcycle, and BooteJTKcycle to RAIN. For non-programmers, Nitecap⁸⁴ provides webbased interactive access to these tools including visualization albeit for simpler, standard study designs.

We also highly recommend filtering identified rhythms using a biologically informed amplitude threshold to consider effect size (amplitude in our case, see Section 5.2) and hence biological relevance in addition to p-values. Small effect sizes also result in failed validation.⁸⁵

5.3 Characterization of rhythms

After their detection, rhythms need to be characterized by determining their parameters. As is shown in Figure 4A, the canonical circadian parameters are the rhythm period, amplitude, acrophase or peak phase, and mesor or magnitude. Typical rhythm detection methods provide an estimate of the period, with different accuracies depending on the data. The remaining rhythm parameters are mathematically well-defined only when a sinusoidal waveform is assumed. They can be estimated using the cosinor method introduced in the previous section, with the period either obtained from rhythm detection or fixed at 24 h.

5.4 Discrimination of rhythms

Circadian medicine research requires comparisons of circadian rhythms between different conditions. Intuitively, it appears that performing rhythm identification in each condition and then comparing the results is correct. However, it was recently shown that this approach is flawed and that it results in a high rate of false positives.⁷⁰ This flaw is exacerbated in high-throughput studies, where the comparison is often visualized using Venn diagrams, as is illustrated in Figure 4B. Thus, readers might use Venn diagrams as a flag to more closely inspect the method used for comparison. To control false positives at the desired level, the parameters of the circadian rhythms (see previous section) must be directly compared across conditions. In the case of two conditions, this can be achieved within either a hypothesis testing or a model selection framework. In the former, the null hypothesis is tested that the coefficients of the cosinor fits to the two conditions are the same. In the latter, a collection of linear models representing the four categories of rhythmic patterns in the two conditions (a gain of, a loss of, a change of, or unaltered rhythms) are fitted and the "best" category (model) based on a Bayesian model selection criterion is selected. The former and latter approaches are available in different implementations,^{86–88} whose relative merits are discussed elsewhere.⁷⁰ Both approaches are implemented in an easy-to-use R package compareRhythms,⁷⁰ which also features amplitudebased effect-size filtering of the identified rhythms for improved biological relevance.

5.5 | Rhythm analysis in humans

Circadian rhythm analysis in humans brings its own challenges because humans and their internal clocks are heterogeneous (as compared to isogenic animal models). Even in healthy individuals, circadian clock parameters vary across age, sex, genetic background,⁸⁹ seasons, and environmental light exposure.⁹⁰ There are several reasons to account for this heterogeneity: (i) to quantify this feature of the clock, (ii) to eliminate it in order to get at the underlying circadian phenomena, and (iii) to personalize circadian medicine. Under disease, the combination of disease-clock interactions, psychological stress, effects of therapeutics, comorbidities, Zeitgeber-disrupted environments (e.g., intensive care units⁹¹), and altered rhythmic behaviors can further compound the heterogeneity. The standard approaches described in the previous sections therefore need to be modified in several ways in the context of human studies. Since studies in humans are often longitudinal, measurements from each subject are correlated, unlike (typical) samples from model organisms. As one solution, cosinor analysis can be made to account for correlations using linear mixed models.¹⁶ Rhythms can be averaged across study subjects to reduce the impact of noise and assess a population rhythm if the study cohort is sufficiently homogenous. In circadian studies, this homogeneity extends to the ability to align the internal time of subjects by correcting the sampling times to characterize the underlying common circadian rhythm.¹⁶ This often requires a complete chronobiological profile of the subjects. When the subjects under study are highly heterogeneous and chronobiological profiling is infeasible, circadian rhythms can be analyzed at the

individual level using amplitude as the metric.⁹² This latter approach makes fewer assumptions and allows stratification of the cohort according to chronobiological factors, but is less statistically powerful.

5.6 | Population approaches for human circadian analysis

The other key challenge in studying human rhythms is our inability to repeatedly sample most tissues from single individuals. If one settles for population rhythms instead, only one tissue sample per human would be needed. This depends critically on the circadian variation exceeding the inter-individual variation, which has been shown to be the case in human skin.¹⁶ In fact, such single tissue biopsies are available in many public databases, such as The Cancer Genome Atlas, or are part of tissue banks. Of course, accurate time labels for these samples are still needed to construct a population-level time series. Unfortunately, such time labels are generally missing as they were not collected with circadian studies in mind.

In 2017, cyclic ordering by periodic structure (CYCLOPS),⁹³ a linear auto-encoder with a circular node, was used to reorder unlabeled human samples to construct population-level time series datasets from postmortem human tissue expression data.¹² The key idea is to "code" a sample consisting of multiple features (e.g., the transcriptome measured at a single time) as two numbers that are constrained to lie on a circle and represent the time (or phase) of the sample. Although the authors test their approach on mouse data and a single human dataset where the ordering is known, it is unclear if and how well their approach works in general. Moreover, prior biological data from mice need to be used to filter the search among conserved rhythmic genes, which makes this approach potentially biased. Recently, the fact that this human tissue dataset included multiple samples from the same human was used to improve the time labeling by means of an expectation-maximization algorithm.94 Similarly, Oscope95 was designed to find oscillatory genes in unsynchronized single-cell RNA-seq data by finding pairs of oscillating genes that can provide good circular ordering followed by clustering of the pairs and nearest-neighbor insertion to find time labels for individual samples. The main drawback of this approach is the combinatorial growth in the search space for good pairs of genes for ordering. Thus, not only are improved computational algorithms necessary to reconstruct rhythms from population-sampled data despite the above-described heterogeneity, but validation studies must also be undertaken to evaluate the true accuracy of these reconstructions.

5.7 | Novel biomarkers for human circadian clock analysis

In recent years, several laboratories have developed methods to read the circadian clock in humans, in particular its phase (i.e., chronotype). Molecular biomarkers identified using machine learning can quantify the circadian clock phase using only one or a few biosamples (for reviews, see Münch and Kramer¹⁵ and Dijk and Duffy⁹⁶). This is a major advance because traditionally the analysis of the human circadian system required the acquisition of long time series of markers, such as melatonin or CBT, which required specific laboratory protocols to exclude environmental and behavioral influences on rhythmicity (e.g., the constant routine (CR) protocol⁹⁷). These protocols are still the gold standard but have the disadvantage that they are time-consuming, expensive, and therefore not suitable for larger studies, nor for everyday clinical use. To date, it remains unclear how robust the developed innovative biomarkers are and they still need to be validated against gold-standard methods in different patient populations and under varying non-laboratory conditions.

Moreover, it is imprecise to refer to "reading the circadian clock" in humans because there is not a circadian clock but rather a network of tissue clocks. These tissue clocks are separated by fixed time (phase) differences that possibly vary with health status, environmental conditions (e.g., during jet lag), and between individuals. Current phase biomarkers measure the clock in peripheral tissue (blood, skin) and infer the central clock time (time according to DLMO) or wall time assuming a fixed phase relationship between them. The tissue in which time is in fact measured and which time ought to be measured for a certain application of circadian medicine remain open. Finally, only the circadian phase has received attention thus far, and circadian amplitude (and associated biomarkers) might be at least if not more important; for example, there is a loss of circadian central clock amplitude with aging⁹⁸ in mice.

6 SUMMARY AND CONCLUSION

In this article, we have presented an overview of requirements, available solutions, and open challenges for data integration and analysis for studying, implementing, and advancing circadian medicine. Considering the types of data particularly relevant to research in this field and the fact that those need to be captured and analyzed in the form of time series with various resolutions and rhythmic properties, there is a range of challenges associated with building adequate research platforms and handling associated data quality issues. On the data integration level, these challenges range from a lack of standards that can be used to represent all necessary data types in a common, interoperable form, to challenges associated with data storage, the need to perform timedependent transformations for integrated visualizations, and to data protection requirements. On the data analysis level, challenges exist in regard to the determination of periodicity in repeated-measures designs and differential rhythmicity analysis, when rhythms are not sinusoidal. Moreover, specific challenges in the study of human circadian rhythms relate to the heterogeneity of clock parameters and the inability to repeatedly sample most human tissues.

Many of these challenges point toward important future research directions for data platforms supporting circadian medicine. These are also highly relevant for translating circadian medicine into clinical practice. While this review focuses mostly on clinical research, there are related technical and organizational challenges that need to be considered in clinical routine processes and systems pointing toward synergies that could be leveraged. For example, performing circadian medicine research with clinical routine data requires accurate and consistent (across institutions and systems) timestamps for sample collection, analysis, and result reporting. The same is true for integrating circadian assessments into healthcare processes and for implementing a range of other clinical mechanisms, such as alert systems that are based on lab measurements. As with other novel types of information with clinical relevance, integrating circadian data into electronic health record systems and aggregating information relevant to clinical workflows can be challenging. As a consequence, flexible data standards are needed that can be used to integrate new data types while keeping up with medical developments and automated decision support systems must be developed and validated to provide only information that is relevant at the point of care. In the opposite direction, electronic health record systems already store a wide range of data that is relevant to circadian medicine research and pragmatic interfaces and pay-as-you-go approaches are needed to make it available in this context. Furthermore, data collected for circadian medicine is highly sensitive medical data collected not only in the clinical setting but also in other environments using various sensors. Consequently, adequate information security needs to be maintained for example in terms of establishing secure communication channels, and data protection challenges need to be resolved to ensure individuals' privacy while preserving linkability requirements.

We point out that, to the best of our knowledge, chronobiological analyses are currently not used in routine healthcare. Even in the research context, due to the reasons outlined above, there are currently significant limitations to the application of rhythm analysis to humans and only a few papers describe analysis approaches that are closely associated with translational applications. We conclude that circadian medicine, as an emerging field with collaborations between researchers and clinicians from different disciplines, provides an ideal environment for developing innovative methods for solving challenges associated with the collection, integration, visualization, and analysis of multi-modal biomedical data.

ACKNOWLEDGMENTS

B. A. acknowledges support from the Deutsche Forschungsgemeinschaft (DFG) grant AN 1553/2-1. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Lena Baum ^(b) https://orcid.org/0000-0003-4983-4025 Marco Johns ^(b) https://orcid.org/0000-0003-4706-6595 Maija Poikela ^(b) https://orcid.org/0000-0003-2616-4401 Ralf Möller ^(b) https://orcid.org/0000-0002-1174-3323 Bharath Ananthasubramaniam ^(b) https://orcid. org/0000-0003-4467-1546

Fabian Prasser D https://orcid.org/0000-0003-3172-3095

REFERENCES

- Allada R, Bass J. Circadian mechanisms in medicine. Longo DL, ed. N Engl J Med. 2021;384(6):550-561. doi:10.1056/NEJM ra1802337
- Kramer A, Lange T, Spies C, Finger AM, Berg D, Oster H. Foundations of circadian medicine. *PLoS Biol.* 2022;20(3):e3001567. doi:10.1371/journal.pbio.3001567
- Mazzotti DR, Haendel MA, McMurry JA, et al. Sleep and circadian informatics data harmonization: a workshop report from the sleep research society and sleep research network. *Sleep*. 2022;45(6). doi:10.1093/sleep/zsac002
- Cederroth CR, Albrecht U, Bass J, et al. Medicine in the fourth dimension. *Cell Metab.* 2019;30(2):238-250. doi:10.1016/j. cmet.2019.06.019
- Shameer K, Badgeley MA, Miotto R, Glicksberg BS, Morgan JW, Dudley JT. Translational bioinformatics in the era of realtime biomedical, health care and wellness data streams. *Brief Bioinform.* 2017;18(1):105-124. doi:10.1093/bib/bbv118
- Crnko S, Du Pré BC, Sluijter JPG, Van Laake LW. Circadian rhythms and the molecular clock in cardiovascular biology and disease. *Nat Rev Cardiol.* 2019;16(7):437-447. doi:10.1038/ s41569-019-0167-4
- Sancar A, Van Gelder RN. Clocks, cancer, and chronochemotherapy. *Science*. 2021;371(6524):eabb0738. doi:10.1126/science.abb0738
- 8. Long JE, Drayson MT, Taylor AE, Toellner KM, Lord JM, Phillips AC. Morning vaccination enhances antibody response

over afternoon vaccination: a cluster-randomised trial. *Vaccine*. 2016;34(24):2679-2685. doi:10.1016/j.vaccine.2016.04.032

- Felten M, Ferencik S, Teixeira Alves LG, et al. Ventilatorinduced lung injury is modulated by the circadian clock. *Am J Respir Crit Care Med.* Published online December 2022:rccm.202202-0320OC. doi:10.1164/rccm.202202-0320OC
- Tendler A, Bar A, Mendelsohn-Cohen N, et al. Hormone seasonality in medical records suggests circannual endocrine circuits. *Proc Natl Acad Sci*. 2021;118(7):e2003926118. doi:10.1073/ pnas.2003926118
- Chu SH, Huang M, Kelly RS, et al. Integration of metabolomic and other omics data in population-based study designs: an epidemiological perspective. *Metabolites*. 2019;9(6):E117. doi:10.3390/metabo9060117
- 12. Ruben MD, Wu G, Smith DF, et al. A database of tissue-specific rhythmically expressed human genes has potential applications in circadian medicine. *Sci Transl Med.* 2018;10(458):eaat8806. doi:10.1126/scitranslmed.aat8806
- Fonseca Costa SS, Robinson-Rechavi M, Ripperger JA. Singlecell transcriptomics allows novel insights into aging and circadian processes. *Brief Funct Genomics*. 2020;19(5–6):343-349. doi:10.1093/bfgp/elaa014
- Bhargava A, Herzel H, Ananthasubramaniam B. Mining for novel candidate clock genes in the circadian regulatory network. *BMC Syst Biol.* 2015;9(1):78. doi:10.1186/ s12918-015-0227-2
- Münch M, Kramer A. Timing matters: new tools for personalized chronomedicine and circadian health. *Acta Physiol.* 2019;227(2):e13300. doi:10.1111/apha.13300
- del Olmo M, Spörl F, Korge S, et al. Inter-layer and intersubject variability of circadian gene expression in human skin. *NAR Genom Bioinform*. 2022;4(4):lqac097. doi:10.1093/nargab/ lqac097
- Wittenbrink N, Ananthasubramaniam B, Münch M, et al. High-accuracy determination of internal circadian time from a single blood sample. *J Clin Invest.* 2018;128(9): 3826-3839. doi: 10.1172/JCI120874
- Brown SA, Kunz D, Dumas A, et al. Molecular insights into human daily behavior. *Proc Natl Acad Sci.* 2008;105(5):1602-1607. doi:10.1073/pnas.0707772105
- Weng C. Clinical data quality: a data life cycle perspective. Biostat Epidemiol. 2020;4(1):6-14. doi:10.1080/24709360.2019.1 572344
- Werner A, Gunga HC. Monitoring of Core body temperature in humans. In: Choukèr A, ed. *Stress Challenges and Immunity in Space*. Springer International Publishing; 2020:477-498. doi:10.1007/978-3-030-16996-1_26
- Komarzynski S, Bolborea M, Huang Q, Finkenstädt B, Lévi F. Predictability of individual circadian phase during daily routine for medical applications of circadian clocks. *JCI Insight*. 2019;4(18):e130423. doi:10.1172/jci.insight.130423
- 22. Culver A, Coiffard B, Antonini F, et al. Circadian disruption of core body temperature in trauma patients: a single-center retrospective observational study. *J Intensive Care*. 2020;8(1):4. doi:10.1186/s40560-019-0425-x
- Broussard JL, Reynolds AC, Depner CM, Ferguson SA, Dawson D, Wright KP. Circadian rhythms versus daily patterns in human physiology and behavior. In: Kumar V, ed. *Biological Timekeeping: Clocks, Rhythms and Behaviour.* Springer; 2017:279-295. doi:10.1007/978-81-322-3688-7_13

TA PHYSIOLOGICA

- 24. Pandi-Perumal SR, Smits M, Spence W, et al. Dim light melatonin onset (DLMO): a tool for the analysis of circadian phase in human sleep and chronobiological disorders. *Prog Neuropsychopharmacol Biol Psychiatry*. 2007;31(1):1-11.
- Peixoto AJ, White WB. Circadian blood pressure: clinical implications based on the pathophysiology of its variability. *Kidney Int.* 2007;71(9):855-860. doi:10.1038/sj.ki.5002130
- Mombers C, Legako K, Gilchrist A. Identifying medical wearables and sensor technologies that deliver data on clinical endpoints. *Br J Clin Pharmacol.* 2016;81(2):196-198. doi:10.1111/ bcp.12818
- Depner CM, Cheng PC, Devine JK, et al. Wearable technologies for developing sleep and circadian biomarkers: a summary of workshop discussions. *Sleep.* 2020;43(2):zsz254. doi:10.1093/ sleep/zsz254
- Cheng P, Walch O, Huang Y, et al. Predicting circadian misalignment with wearable technology: validation of wrist-worn actigraphy and photometry in night shift workers. *Sleep*. 2021;44(2):zsaa180. doi:10.1093/sleep/zsaa180
- van Hees VT, Sabia S, Jones SE, et al. Estimating sleep parameters using an accelerometer without sleep diary. *Sci Rep.* 2018;8(1):12975. doi:10.1038/s41598-018-31266-z
- Bowman C, Huang Y, Walch OJ, et al. A method for characterizing daily physiology from widely used wearables. *Cell Rep Methods*. 2021;1(4):100058. doi:10.1016/j.crmeth.2021.100058
- van der Willik EM, Terwee CB, Bos WJW, et al. Patient-reported outcome measures (PROMs): making sense of individual PROM scores and changes in PROM scores over time. *Nephrol Carlton Vic.* 2021;26(5):391-399. doi:10.1111/nep.13843
- Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD. What is sufficient evidence for the reliability and validity of patientreported outcome measures? *Value Health*. 2007;10:S94-S105. doi:10.1111/j.1524-4733.2007.00272.x
- Prasser F, Kohlbacher O, Mansmann U, Bauer B, Kuhn K. Data integration for future medicine (DIFUTURE): an architectural and methodological overview. *Methods Inf Med.* 2018;57(S 01):e57-e65. doi:10.3414/ME17-02-0022
- 34. Holub P, Kohlmayer F, Prasser F, et al. Enhancing reuse of data and biological material in medical research: from FAIR to FAIR-health. *Biopreserv Biobank*. 2018;16(2):97-105. doi:10.1089/bio.2017.0110
- Cock PJA, Fields CJ, Goto N, Heuer ML, Rice PM. The sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res.* 2009;38(6):1767-1771. doi:10.1093/nar/gkp1137
- 36. Poinsignon T, Poulain P, Gallopin M, Lelandais G. Working with omics data, an interdisciplinary challenge at the crossroads of biology and computer science. In: Colliot, O, eds. *Machine Learning for Brain Disorders*; Humana Press; 2023. https://hal.science/hal-03738102
- Zhao J, Feng Q, Wei WQ. Integration of omics and phenotypic data for precision medicine. In: Bai JPF, Hur J, eds. *Systems Medicine*. Vol 2486. Springer US; 2022:19-35. doi:10.1007/978-1-0716-2265-0_2
- Santiago-Rodriguez TM, Hollister EB. Multi 'omic data integration: a review of concepts, considerations, and approaches. *Semin Perinatol.* 2021;45(6):151456. doi:10.1016/j. semperi.2021.151456
- 39. Yalçin M, El-Athman R, Ouk K, Priller J, Relógio A. Analysis of the circadian regulation of cancer hallmarks by a cross-platform

study of colorectal cancer time-series data reveals an association with genes involved in Huntington's disease. *Cancer*. 2020;12(4):963. doi:10.3390/cancers12040963

- Pizarro A, Hayer K, Lahens NF, Hogenesch JB. CircaDB: a database of mammalian circadian gene expression profiles. *Nucleic Acids Res.* 2012;41(D1):D1009-D1013. doi:10.1093/nar/ gks1161
- Samad M, Agostinelli F, Sato T, Shimaji K, Baldi P. CircadiOmics: circadian omic web portal. *Nucleic Acids Res.* 2022;50(W1):W183-W190. doi:10.1093/nar/gkac419
- Schulz S, Stegwee R, Chronaki C. Standards in healthcare data. In: Kubben P, Dumontier M, Dekker A, eds. *Fundamentals of Clinical Data Science*. Springer; 2019. Accessed September 14, 2022. http://www.ncbi.nlm.nih.gov/books/NBK543525/
- Spengler H, Lang C, Mahapatra T, Gatz I, Kuhn KA, Prasser F. Enabling agile clinical and translational data warehousing: platform development and evaluation. *JMIR Med Inform*. 2020;8(7):e15918. doi:10.2196/15918
- Bender D, Sartipi K. HL7 FHIR: an agile and RESTful approach to healthcare information exchange. Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems. IEEE; 2013:326-331. doi:10.1109/CBMS.2013.6627810
- 45. Vorisek CN, Lehne M, Klopfenstein SAI, et al. Fast healthcare interoperability resources (FHIR) for interoperability in health research: systematic review. *JMIR Med Inform*. 2022;10(7):e35724. doi:10.2196/35724
- Ruben MD, Francey LJ, Guo Y, et al. A large-scale study reveals 24-h operational rhythms in hospital treatment. *Proc Natl Acad Sci.* 2019;116(42):20953-20958. doi:10.1073/pnas.1909557116
- 47. Urbauer P, Frohner M, David V, Sauermann S. Wearable activity trackers supporting elderly living independently: a standards based approach for data integration to health information systems. In Hadjileontiadis L, Barroso J, Sandnes FE, eds. Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion. ACM; 2018:302-309. doi:10.1145/3218585.3218679
- Alonso J, Bartlett SJ, Rose M, et al. The case for an international patient-reported outcomes measurement information system (PROMIS[®]) initiative. *Health Qual Life Outcomes*. 2013;11(1):210. doi:10.1186/1477-7525-11-210
- 49. McDonald CJ, Huff SM, Suico JG, et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update. *Clin Chem*. 2003;49(4):624-633. doi:10.1373/49.4.624
- Bennett AV, Jonsson M, Chen RC, Al-Khatib SM, Weinfurt KP, Curtis LH. Applying patient-reported outcome methodology to capture patient-reported health data: report from an NIH Collaboratory roundtable. *Healthc Amst Neth*. 2020;8(3):100442. doi:10.1016/j.hjdsi.2020.100442
- Mehraeen E, Ghazisaeedi M, Farzi J, Mirshekari S. Security challenges in healthcare cloud computing: a systematic review. *Glob J Health Sci.* 2016;9(3):157. doi:10.5539/gjhs.v9n3p157
- 52. Schwarz N. Self-reports: how the questions shape the answers. *Am Psychol.* 1999;54(2):93-105. doi:10.1037/0003-066X.54.2.93
- Schiff S, Gehrke M, Möller R. Efficient enriching of synthesized relational patient data with time series data. *Procedia Comput Sci.* 2018;141:531-538. doi:10.1016/j. procs.2018.10.130
- 54. TimescaleDB. Time-series data simplified. *Timescale*. 2022. Accessed September 14, 2022. https://www.timescale.com/

- InfluxDB. InfluxDB: open source time series database. InfluxData. 2022. Accessed September 14, 2022. https://www. influxdata.com/
- Mazzotti DR. Landscape of biomedical informatics standards and terminologies for clinical sleep medicine research: a systematic review. *Sleep Med Rev.* 2021;60:101529. doi:10.1016/j. smrv.2021.101529
- Lee G, Kang B, Nho K, Sohn KA, Kim D. MildInt: deep learning-based multimodal longitudinal data integration framework. *Front Genet.* 2019;10:617. doi:10.3389/ fgene.2019.00617
- Chakraborty M, Kundan AP. Grafana. Monitoring Cloud-Native Applications. Apress; 2021:187-240. doi:10.1007/ 978-1-4842-6888-9_6
- 59. Drake J, Schulz K, Bukowski R, Gaither K. Collecting and analyzing smartphone sensor data for health. In: Paris J, Milhans J, Hillery B, Geva SB, Schmitz P, Sinkovits R, eds. *Practice and Experience in Advanced Research Computing. PEARC* '21. Association for Computing Machinery; 2021. doi:10.1145/3437359.3465599
- Cruz R, Guimarães T, Peixoto H, Santos MF. Architecture for intensive care data processing and visualization in realtime. *Procedia Comput Sci.* 2021;184:923-928. doi:10.1016/j. procs.2021.03.115
- Çalhan A, Cicioğlu M, Ceylan A. EHealth monitoring testbed with fuzzy based early warning score system. *Comput Methods Programs Biomed.* 2021;202:106008. doi:10.1016/j. cmpb.2021.106008
- 62. Nguyen M, Purushotham S, To H, Shahabi C. *M-TSNE: a framework for visualizing high-dimensional multivariate time series.* 2017. 10.48550/ARXIV.1708.07942
- Fischer D, Vetter C, Roenneberg T. A novel method to visualise and quantify circadian misalignment. *Sci Rep.* 2016;6(1):38601. doi:10.1038/srep38601
- Aamot H, Kohl CD, Richter D, Knaup-Gregori P. Pseudonymization of patient identifiers for translational research. *BMC Med Inform Decis Mak.* 2013;13:75. doi:10.1186/ 1472-6947-13-75
- Gostin LO, Halabi SF, Wilson K. Health data and privacy in the digital era. JAMA. 2018;320(3):233. doi:10.1001/ jama.2018.8374
- 66. Shou L, Shang X, Chen K, Chen G, Zhang C. Supporting pattern-preserving anonymization for time-series data. *IEEE Trans Knowl Data Eng.* 2013;25(4):877-892. doi:10.1109/ TKDE.2011.249
- Rocher L, Hendrickx JM, de Montjoye YA. Estimating the success of re-identifications in incomplete datasets using generative models. *Nat Commun.* 2019;10(1):3069. doi:10.1038/ s41467-019-10933-3
- Chen RJ, Lu MY, Chen TY, Williamson DFK, Mahmood F. Synthetic data in machine learning for medicine and healthcare. *Nat Biomed Eng.* 2021;5(6):493-497. doi:10.1038/ s41551-021-00751-8
- Dunlap JC, Loros JJ, DeCoursey PJ. Chronobiology: Biological Timekeeping. Sinauer Associates; 2004.
- Pelikan A, Herzel H, Kramer A, Ananthasubramaniam B. Venn diagram analysis overestimates the extent of circadian rhythm reprogramming. *FEBS J.* 2022;289(21):6605-6621. doi:10.1111/ febs.16095

- Hughes ME, Abruzzi KC, Allada R, et al. Guidelines for genome-scale analysis of biological rhythms. *J Biol Rhythms*. 2017;32(5):380-393. doi:10.1177/0748730417728663
- 72. Ness-Cohn E, Iwanaszko M, Kath WL, Allada R, Braun R. TimeTrial: an interactive application for optimizing the design and analysis of transcriptomic time-series data in circadian biology research. *J Biol Rhythms*. 2020;35(5):439-451. doi:10.1177/0748730420934672
- Refinetti R, Cornélissen G, Halberg F. Procedures for numerical analysis of circadian rhythms. *Biol Rhythm Res.* 2007;38(4):275-325. doi:10.1080/09291010600903692
- 74. Cenek L, Klindziuk L, Lopez C, et al. CIRCADA: shiny apps for exploration of experimental and synthetic circadian time series with an educational emphasis. *J Biol Rhythms*. 2020;35(2):214-222. doi:10.1177/0748730419900866
- Schmal C, Mönke G, Granada AE. Analysis of complex circadian time series data using wavelets. In: Solanas G, Welz P, eds. *Circadian Regulation*. Springer; 2022:35-54. doi:10.1101/2020.04.29.067744
- Cornelissen G. Applications of cosinor rhythmometry in pharmacology. *J Pharmacokinet Pharmacodyn*. 2021;48(3):339-359. doi:10.1007/s10928-021-09748-x
- Bingham C, Arbogast B, Guillaume GC, Lee JK, Halberg F. Inferential statistical methods for estimating and comparing cosinor parameters. *Chronobiologia*. 1982;9(4):397-439.
- Asher G, Zhu B. Beyond circadian rhythms: emerging roles of ultradian rhythms in control of liver functions. *Hepatology*. 2022;hep.32580. doi:10.1002/hep.32580
- 79. Ananthasubramaniam B, Diernfellner A, Brunner M, Herzel H. Ultradian rhythms in the transcriptome of *Neurospora crassa*. *iScience*. 2018;9:475-486. doi:10.1016/j.isci.2018.11.012
- Hughes ME, Hogenesch JB, Kornacker K. JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *J Biol Rhythms*. 2010;25(5):372-380. doi:10.1177/0748730410379711
- Thaben PF, Westermark PO. Detecting rhythms in time series with RAIN. J Biol Rhythms. 2014;29(6):391-400. doi:10.1177/0748730414553029
- Hutchison AL, Maienschein-Cline M, Chiang AH, et al. Improved statistical methods enable greater sensitivity in rhythm detection for genome-wide data. *PLoS Comput Biol.* 2015;11(3):e1004094. doi:10.1371/journal.pcbi.1004094
- Hutchison AL, Allada R, Dinner AR. Bootstrapping and empirical Bayes methods improve rhythm detection in sparsely sampled data. J Biol Rhythms. 2018;33(4):339-349. doi:10.1177/0748730418789536
- Brooks TG, Mrčela A, Lahens NF, et al. Nitecap: an exploratory circadian analysis web application. *J Biol Rhythms*. 2022;37(1):43-52. doi:10.1177/07487304211054408
- 85. Nuzzo R. Scientific method: statistical errors. *Nature*. 2014;506(7487):150-152. doi:10.1038/506150a
- Singer JM, Hughey JJ. LimoRhyde: a flexible approach for differential analysis of rhythmic transcriptome data. J Biol Rhythms. 2019;34(1):5-18. doi:10.1177/0748730418813785
- Thaben PF, Westermark PO. Differential rhythmicity: detecting altered rhythmicity in biological data. *Bioinformatics*. 2016;32(18):2800-2808. doi:10.1093/bioinformatics/btw309
- 88. Weger BD, Gobet C, David FPA, et al. Systematic analysis of differential rhythmic liver gene expression mediated by

cta Physiologica

the circadian clock and feeding rhythms. *Proc Natl Acad Sci.* 2021;118(3):e2015803118. doi:10.1073/pnas.2015803118

- Ashbrook LH, Krystal AD, Fu YH, Ptáček LJ. Genetics of the human circadian clock and sleep homeostat. *Neuropsychopharmacology*. 2020;45(1):45-54. doi:10.1038/ s41386-019-0476-7
- Roenneberg T, Merrow M. Entrainment of the human circadian clock. *Cold Spring Harb Symp Quant Biol*. 2007;72(1):293-299. doi:10.1101/sqb.2007.72.043
- McKenna H, van der Horst GTJ, Reiss I, Martin D. Clinical chronobiology: a timely consideration in critical care medicine. *Crit Care*. 2018;22(1):124. doi:10.1186/s13054-018-2041-x
- Lachmann G, Ananthasubramaniam B, Wünsch VA, et al. Circadian rhythms in septic shock patients. *Ann Intensive Care*. 2021;11(1):64. doi:10.1186/s13613-021-00833-5
- Anafi RC, Francey LJ, Hogenesch JB, Kim J. CYCLOPS reveals human transcriptional rhythms in health and disease. *Proc Natl Acad Sci.* 2017;114(20):5312-5317. doi:10.1073/ pnas.1619320114
- 94. Talamanca L, Gobet C, Naef F. Sex-dimorphic and agedependent organization of 24-hour gene expression rhythms in human. *Science*. 2023;379:478-483. doi:10.1126/science. add0846

- Leng N, Chu LF, Barry C, et al. Oscope identifies oscillatory genes in unsynchronized single-cell RNA-seq experiments. *Nat Methods*. 2015;12(10):947-950. doi:10.1038/nmeth.3549
- 96. Dijk DJ, Duffy JF. Novel approaches for assessing circadian rhythmicity in humans: a review. J Biol Rhythms. 2020;35(5):421-438. doi:10.1177/0748730420940483
- Duffy JF, Dijk DJ. Getting through to circadian oscillators: why use constant routines? J Biol Rhythms. 2002;17(1):4-13. doi:10.1177/074873002129002294
- Ramkisoensing A, Meijer JH. Synchronization of biological clock neurons by light and peripheral feedback systems promotes circadian rhythms and health. *Front Neurol.* 2015;6. doi:10.3389/fneur.2015.00128

How to cite this article: Baum L, Johns M, Poikela M, Möller R, Ananthasubramaniam B, Prasser F. Data integration and analysis for circadian medicine. *Acta Physiol.* 2023;237:e13951. doi:<u>10.1111/</u>apha.13951