Published online 15 June 2018 *Nucleic Acids Research, 2018, Vol. 46, Web Server issue* ***W157–W162***
doi: 10.1093/nar/gky441

CircadiOmics: circadian omic web portal

Nicholas Ceglia[1,2,†], Yu Liu[1,2,†], Siwei Chen[1,2], Forest Agostinelli[1,2], Kristin Eckel-Mahan[3], Paolo Sassone-Corsi[2,4,5] and Pierre Baldi[1,2,4,5,*]

[1]Department of Computer Science, University of California, Irvine, CA 92617, USA, [2]Institute for Genomics and Bioinformatics, University of California, Irvine, CA 92617, USA, [3]Institute of Molecular Medicine, McGovern Medical School, The University of Texas Health Science Center, Houston, TX 77030, USA, [4]Center for Epigenetics and Metabolism, School of Medicine, University of California, Irvine, CA 92617, USA and [5]Department of Biochemistry, University of California, Irvine, CA 92617, USA

Received February 09, 2018; Revised April 17, 2018; Editorial Decision May 03, 2018; Accepted June 13, 2018

ABSTRACT

Circadian rhythms play a fundamental role at all levels of biological organization. Understanding the mechanisms and implications of circadian oscillations continues to be the focus of intense research. However, there has been no comprehensive and integrated way for accessing and mining all circadian omic datasets. The latest release of CircadiOmics (http://circadiomics.ics.uci.edu) fills this gap for providing the most comprehensive web server for studying circadian data. The newly updated version contains high-throughput 227 omic datasets corresponding to over 74 million measurements sampled over 24 h cycles. Users can visualize and compare oscillatory trajectories across species, tissues and conditions. Periodicity statistics (e.g. period, amplitude, phase, *P*-value, q-value etc.) obtained from BIO_CYCLE and other methods are provided for all samples in the repository and can easily be downloaded in the form of publication-ready figures and tables. New features and substantial improvements in performance and data volume make CircadiOmics a powerful web portal for integrated analysis of circadian omic data.

INTRODUCTION

Circadian rhythms are a ubiquitous phenomenon in biology that is deeply rooted in evolution (1,2). Circadian oscillations of molecular species maintain homeostatic balance by regulating a variety of physiological and metabolic processes. These processes include sleep/wake cycle, hormone secretion, diet related metabolism and neural function (3–6). Disruption in circadian rhythms can lead to a wide range of health problems such as diabetes, obesity and premature aging (7–11).

It is well known that circadian oscillations at the transcriptomic level are pervasive and well coordinated (4,12,2). Oscillation in transcription is strongly regulated by a number of key transcription factors, such as CLOCK, BMAL1, PERs and CRYs that comprise the *core clock* (13). These transcript level oscillations form regulatory feedback loops that oscillate throughout the transcriptome (14–15,2). Moreover, a large number of metabolites and proteins in cells exhibit circadian oscillations and may play a key role within the organization of genetic circadian regulation (16–19). Strikingly, the circadian landscape in a cell can be drastically different depending on genetic and epigenetic conditions (17,12,2,20). The process by which these circadian landscapes evolve is understood as circadian reprogramming. Reprogramming can be induced by external perturbations such as inflammation or dietary challenge (21–24). The large repository of omic data provided in CircadiOmics, together with several comparative analysis tools, provide a foundational platform that can be used to analyze these complex mechanisms and their implications.

MATERIALS AND METHODS

Dataset collection

The omic datasets available on CircadiOmics are compiled from project collaborations, automated discovery and manual curation. Over 6400 individual time points spanning 227 separate circadian experiments are available for search and visualization. In aggregate, these datasets form the largest single repository of circadian data available, including all datasets from other repositories including CircaDB (25). Table 1 shows a break down of the number of datasets available on several other sources. Eight species are currently available on CircadiOmics. The majority are collected from *Mus musculus* and *Papio anibus*.

Over 62 tissues grouped into 18 categories are represented in the database. Within these categories, liver and brain experiments comprise the majority. Diverse experimental con-

*To whom correspondence should be addressed. Tel: +1 949 824 5809; Fax: +1 949 824 9813; Email: pfbaldi@uci.edu
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Table 1. Data volumes of publicly available circadian omic databases

Source	Experiments	Tissues	Species	Total data pts. (est.)
CircadiOmics	227	23	8	≈74 600 000
CircaDB	30	15	2	<1 800 000
DIURNAL	11	3	3	≈3 009 600
BIOCLOCK	2	2	2	≈3 600 000
CirGRDB	50	<20	2	≈9 000 000

Comparison of CircadiOmics with other circadian repositories. Experiments refers to the total number of experimental level datasets from each source. An experimental level dataset should contain at least two time points, more than one replicate at each time point, and time series data for a substantial number of molecular species–at least 1000 for transcriptome and acetylome, and at least 100 for metabolome and proteome–and each replicate. Total data points provide an estimate of the total number of individual measurements taken across different time points, replicates and molecular species. Numbers are collected from internal statistics for CircadiOmics and from publications, or official websites, for the other sources. Details are provided in Supplementary Material.

ditions grouped into nine broad categories are available for comparison. Unique conditions include chronic and acute ethanol consumption, high-fat diet, traumatic brain injury, fibroblast undergoing myogenic reprogramming and several cancer-specific datasets (26,27). At last, CircadiOmics is the only tool that includes transcriptome, metabolome, acetylome and proteome experiments. Figure 1 summarizes the number of available datasets by detailed categories. The full table of datasets is available, with a short description and experimental details such as number of replicates, on the CircadiOmics web portal.

Increased interest in circadian rhythms is driving a continuous increase in publicly available omic datasets. Automated discovery of datasets has become necessary to maintain the most current and comprehensive repository. A Python framework built with *scholarly* and *geotools* Python packages is used to continuously search the literature for new circadian omic studies and datasets. Automated discovery based on keyword searches in published abstracts is filtered using several features including publishing journal, author and provided supplementary materials. A logistic regression step is used to classify datasets that are good candidates for inclusion in CircadiOmics. Results produced by this automated pipeline are then manually inspected for quality, based primarily on the time point resolution of the dataset. The minimum sampling density for any dataset in the repository is every eight hours over a 24-h cycle. Additionally, the CircadiOmics team and collaborating biologists periodically search recent publications for new datasets that qualify for inclusion in CircadiOmics.

Statistics

All datasets are processed with both BIO_CYCLE and JTK_CYCLE to provide oscillation statistics (e.g. period, amplitude, phase) for each set of samples (28,29). Primary identification of oscillatory species is made using p-values and accompanying q-values at a selected threshold. Technical details for calculating P-values and q-values are provided in the cited articles for the respective methods. BIO_CYCLE results have consistently shown to be an improvement in determining periodicity over older methods (28). The BIO_CYCLE portal within CircadiOmics at http://circadiomics.ics.uci.edu/biocycle allows users to upload an unpublished dataset for processing with BIO_CYCLE. For each experiment and each molecular species, individual P-value, q-value, period, amplitude and phase can be obtained. Additionally, summary figures are generated for the distribution of each statistic in the user provided dataset. Trends for individual trajectories in user-provided data are available for search and visualization through the supplied set of molecular IDs. An example dataset is provided to give the user a sample of portal features and provide a template for desired data format. The main CircadiOmics documentation page provides additional guidance. The BIO_CYCLE R package is also available for download through the main portal.

Implementation

CircadiOmics is available as a pubic domain website at http://circadiomics.ics.uci.edu. The CircadiOmics web application is constructed as a three-tier Model View Controller architecture. The web server is implemented with the Flask Python library. The interface is generated dynamically with Twitter Bootstrap and Google Charts. Fast query response times are accomplished by caching JSON serialized datasets on disk as the server is started. Figure 2 describes the web application architecture and corresponding technology. The interface loads with an example search of ARNTL (CLOCK-BMAL) in a sample liver control dataset. Dynamic filtering of the available datasets is provided based on tissue and experimental perturbations. Examples of filtering options are provided in the documentation on the main web server in the context of various sample workflows. Downloadable results for each search include high resolution images in PNG or SVG format, and an excel table of BIO_CYCLE reported statistics. Dataset documentation includes a short technical description as well as a link to the corresponding article in PubMed. At last, additional help information on the features of CircadiOmics is provided through a link on the main page of the web server.

RESULTS

Features

The main functionality of CircadiOmics is the search, comparison and visualization of oscillation trends. The user can search any molecular species in the omic datasets within the repository and overlay multiple searches together to initiate a comparative study. A typical work flow may consist of comparing a set of specific transcripts, metabolites or proteins among several datasets. Intelligent auto-completion

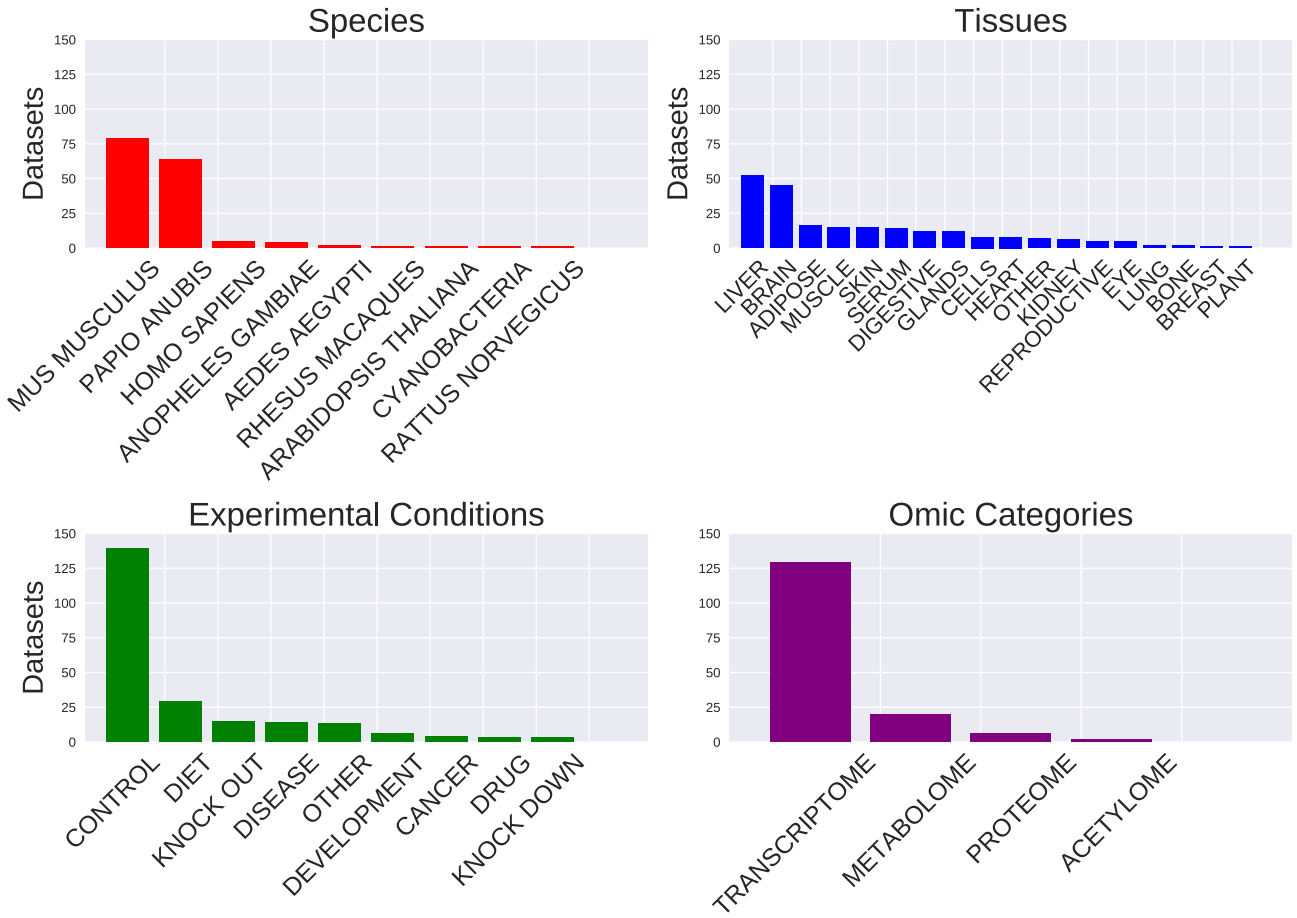

Figure 1. Dataset collection by species, tissues, experimental conditions and omic categories.

facilitates user queries within the currently selected dataset. Searches can be performed individually or in batch on a selected dataset. When datasets do not have the same time course, results are displayed from the minimum to the maximum time point over all selected datasets. Query result for a set of example searches is shown in Figure 3. Documentation available on the web server illustrates common query tasks and results. Datasets with large difference in intensity values at each time point can be dynamically scaled for easy visual comparison. Minimum and maximum values are normalized to zero and one, respectively.

A table of statistics is compiled and displayed beneath the main search window after each query. Statistics can be updated dynamically to reflect results obtained with BIO_CYCLE. The table can be downloaded in several formats compatible with Excel. Individual searches can be removed from both the search view and the statistics table. Figure 3 shows an example result obtained from searches for ARNTL, PER1 and CRY1 in an example dataset.

With a rapidly expanding dataset collection, filtering candidate dataset within the interface has become necessary. The filtering menu allows the user to limit the scope of datasets displayed under drop-down menus for each dataset type. Filtering can be done by species, tissues and experimental conditions. Similar experimental conditions are categorically grouped together in the filtering menu. These include knock-downs, knock-outs, diet changes and drug treatments. The full set of available conditions for filtering is summarized in Figure 1. The search interface uses an abbreviated dataset identification. Upon selection of a dataset, the user can quickly verify the source of the data through a corresponding literature citation. Additional details for each dataset can be found in tabular form under the dataset tab. These details include a brief description of the experimental protocol.

The Metabolic Atlas web portal (http://circadiomics.ics.uci.edu/metabolicatlas) is also available under the CircadiOmics umbrella. In addition to metabolite time series, interactive metabolic networks can be generated and visualized. These networks are derived in part from the KEGG database (30) and can be filtered using BIO_CYCLE statistics.

Improvements

The new version of CircadiOmics considerably increases the amount of data available to the user. In particular, the number of experiment-level datasets increased from 50 to 227, the number of species increased from 1 to 8, the number of transcriptomic datasets increased from 40 to 169, the number of proteomic datasets increased from 1 to 8, the number

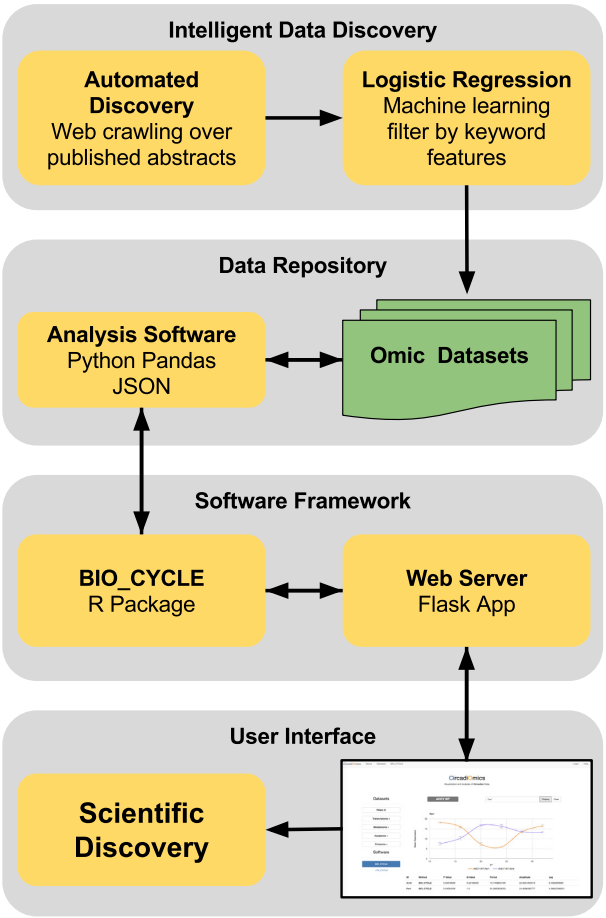

Figure 2. Three-tier Model-View-Controller architecture of the CircadiOmics web portal. Intelligent data discovery supplies candidate datasets for inclusion in the repository using a machine learning filter applied to key word features derived from web crawling published abstracts. BIO_CYCLE results are obtained and stored for all datasets. The user interface sends requests and displays results from the web server allowing for interactive hypothesis generation and scientific discovery.

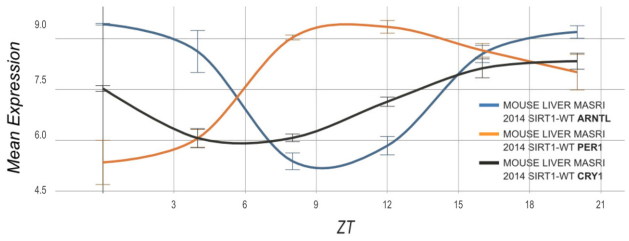

Figure 3. Visualization of queries for ARNTL, PER1 and CRY1 in a control mouse dataset. Any number of queries, across any number of datasets, can be displayed simultaneously.

of acetylome datasets increased from 1 to 8 and the number of metabolomic datasets increased from 5 to 32.

Beyond the multi-fold increase in the underlying data repository, the new version of CircadiOmics comes with several other significant improvements, including a new, more robust, architecture and software infrastructure. In addition, all circadian statistics are computed using the latest version of BIO_CYCLE with the capability to systematically apply any updates on the fly, as new versions of BIO_CYCLE are created and released. Thus, together with intelligent data discovery, CircadiOmics provides state-of-the-art statistical tools for integrating and analyzing circadian data. The server-side code has improved security through encrypted HTTPS connection and enabled user-specific content visibility for unpublished data.

In combination, the new features enable CircadiOmics users to conduct end-to-end circadian analyses, starting from the generation of new hypotheses all the way to the generation of results suitable for publication.

DISCUSSION

Central to the study of circadian rhythms are large-scale reprogramming events. Understanding these events at the molecular level critically depends on being able to access and compare significant amounts of high-throughput circadian omic data. CircadiOmics, with its advanced search features and unprecedented amount of high quality circadian data, is a primary enabling tool for such studies. In a circadian reprogramming event, changes in oscillation of one molecular species can often be related to changes in other molecular species (31,2). One of the main qualities of CircadiOmics is the flexibility of the comparative analyses it enables. For instance, a user can compare transcripts across species, or relate metabolites to proteins and transcripts and identify underlying oscillatory trends. An important example can be seen in the loss of oscillation in the metabolite NAD+ as a response to changes in the transcriptomic oscillatory landscape (17). As a result, CircadiOmics has proven to be highly effective for hypothesis generation in new studies. To date, the web server has contributed to multiple studies that have been published in high impact journals. The server has been accessed more than 250 000 times in total traffic in 2017 alone.

Figure 4 details some examples of the impact of CircadiOmics. For instance, Eckel-Mahan *et al.* utilized CircadiOmics to analyze three related omic datasets in mouse liver (17). They found that core clock genes regulate the acetylation of the enzyme AceCS1. AceCS1 is responsible for changes in the oscillation of the metabolite acetyl-CoA, a key metabolite involved in fatty acid synthesis (Figure 4 A). Similarly, Masri *et al.* compared liver transcriptomic data with metabolomic data in mice afflicted with cancer using CircadiOmics (Figure 4 B). They discovered that a distal tumor-bearing lung can reprogram the liver circadian transcriptome through inflammatory pathways and insulin related metabolic pathways (27). More recently, CircadiOmics has been used to examine the role of circadian regulation in myogenic reprogramming of fibroblast (https://www.biorxiv.org/content/early/2017/06/18/151555). It was observed that the *core clock* is completely disrupted during this process. However, exogenous MYOD1 gains rhythmicity during transition to muscle cell. As a result, MYOG and a majority of critical transcription factors related to muscle development known to be regulated by MYOD1 synchronize oscillation. This behavior was identified in CircadiOmics through visualization and confirmed by BIO_CYCLE reported phase lag (Figure 4 C). At last, aggregating all mouse transcriptomic

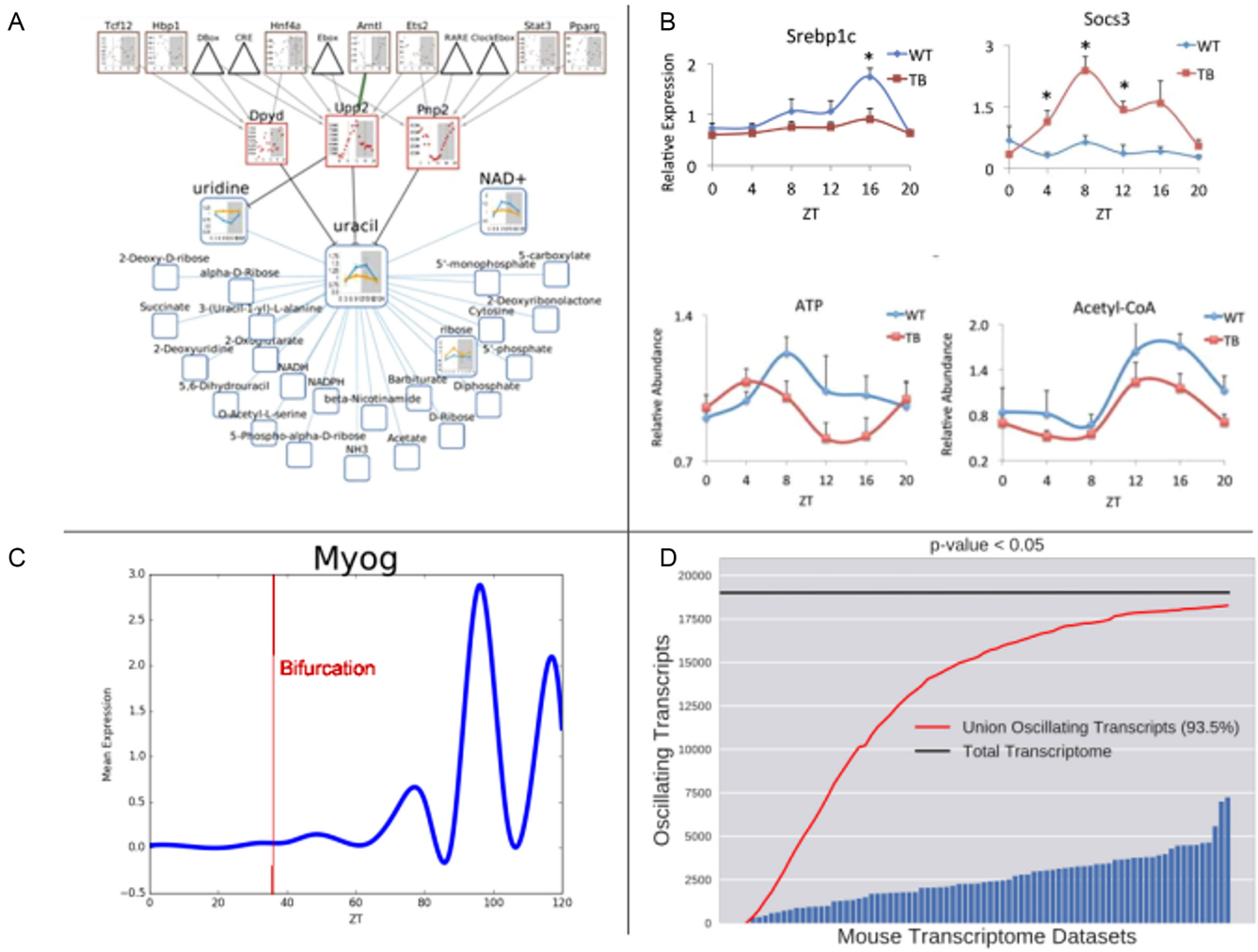

Figure 4. Selected examples of the impact of CircadiOmics. (**A**) CircadiOmics was used to link a multitude of circadian metabolites with functionally related circadian transcripts. Figure taken from Figure 5A of (17). (**B**) CircadiOmics was used to discover reprogrammed circadian transcripts and metabolites related to inflammatory and energy pathways. Figure taken from Figures 2E, 4B and 5D of (27). (**C**) Exogenous MYOD1, during MEF myogenic reprogramming, entrains oscillation in MYOG and related targets in absence of oscillation of the core clock (https://www.biorxiv.org/content/early/2017/06/18/151555). (**D**) Bar heights show the ordered number of oscillating protein coding transcripts with a $P \leq 0.05$ in each mouse transcriptomic experiment in the repository. The trend is the cumulative union of oscillating transcripts. Over 93% of possible protein coding transcripts are found to oscillate in at least one tissue or condition across all mouse datasets.

datasets confirms and amplifies the notion that circadian oscillations are pervasiveness: 93.5% of all possible protein coding transcripts exhibit circadian oscillations in at least one tissue or experiment (up from about 67% in (2)) (Figure 4 D). The large number of datasets in CircadiOmics facilitates these kinds of integrative analyses. Additional analysis of the 1275 protein coding transcripts that are not found to oscillate in any condition or tissue is provided in Supplementary Table S2.

The latest release of CircadiOmics is the largest single repository of circadian omic data available. Updates in server architecture and data mining ensure that CircadiOmics will continue to maintain and grow as new data is published. Improvement in features for search and visualization expand the possibilities for study of circadian rhythms in omic datasets. These possibilities include generating specific hypothesis for individual experiments and answering larger questions about the organization of oscillation within a cell.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Yuzo Kanomata for computing support.

FUNDING

National Institute of General Medical Sciences [GM123558 to P.B.]; Defense Advanced Research Projects Agency [D17AP00002 to P.B.]. Funding for open access charge: UCI Institute for Genomics and Bioinformatics.

Conflict of interest statement. None declared.

REFERENCES

1. Panda,S., Hogenesch,J.B. and Kay,S.A. (2002) Circadian rhythms from flies to human. *Nature*, **417**, 329–335.
2. Patel,V.R., Ceglia,N., Zeller,M., Eckel-Mahan,K., Sassone-Corsi,P. and Baldi,P. (2015) The pervasiveness and plasticity of circadian oscillations: the coupled circadian-oscillators framework. *Bioinformatics*, **31**, 3181–3188.
3. Bass,J. (2012) Circadian topology of metabolism. *Nature*, **491**, 348–356.
4. Dibner,C., Schibler,U. and Albrecht,U. (2010) The mammalian circadian timing system: organization and coordination of central and peripheral clocks. *Annu. Rev. Physiol.*, **72**, 517–549.
5. Gerstner,J.R., Lyons,L.C., Wright,K.P., Loh,D.H., Rawashdeh,O., Eckel-Mahan,K.L. and Roman,G.W. (2009) Cycling behavior and memory formation. *J. Neurosci.*, **29**, 12824–12830.
6. Menet,J.S. and Rosbash,M. (2011) When brain clocks lose track of time: cause or consequence of neuropsychiatric disorders. *Curr. Opin. Neurobiol.*, **21**, 849–857.
7. Adser,H., Wojtaszewski,J. F.P., Jakobsen,A.H., Kiilerich,K., Hidalgo,J. and Pilegaard,H. (2011) Interleukin-6 modifies mRNA expression in mouse skeletal muscle. *Acta Physiol.*, **202**, 165–173.
8. Asher,G. and Sassone-Corsi,P. (2015) Time for food: the intimate interplay between nutrition, metabolism, and the circadian clock. *Cell*, **161**, 84–92.
9. Fu,L. and Lee,C.C. (2003) The circadian clock: pacemaker and tumour suppressor. *Nat. Rev. Cancer*, **3**, 350–361.
10. Partch,C.L., Green,C.B. and Takahashi,J.S. (2014) Molecular architecture of the mammalian circadian clock. *Trends Cell Biol.*, **24**, 90–99.
11. Roenneberg,T. and Merrow,M. (2016) The circadian clock and human health. *Curr. Biol.*, **26**, R432–R443.
12. Koike,N., Yoo,S.-H., Huang,H.-C., Kumar,V., Lee,C., Kim,T.-K. and Takahashi,J.S. (2012) Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science*, **338**, 349–354.
13. Ko,C.H. and Takahashi,J.S. (2006) Molecular components of the mammalian circadian clock. *Hum. Mol. Genet.*, **15**, 271–277.
14. Masri,S., Patel,V.R., Eckel-Mahan,K.L., Peleg,S., Forne,I., Ladurner,A.G., Baldi,P., Imhof,A. and Sassone-Corsi,P. (2013) Circadian acetylome reveals regulation of mitochondrial metabolic pathways. *Proc. Natl. Acad. Sci.*, **110**, 3339–3344.
15. Robles,M.S., Cox,J. and Mann,M. (2014) In-vivo quantitative proteomics reveals a key contribution of post-transcriptional mechanisms to the circadian regulation of liver metabolism. *PLoS Genet.*, **10**, e1004047.
16. Dallmann,R., Viola,A.U., Tarokh,L., Cajochen,C. and Brown,S.A. (2012) The human circadian metabolome. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 2625–2629.
17. Eckel-Mahan,K.L., Patel,V.R., Mohney,R.P., Vignola,K.S., Baldi,P. and Sassone-Corsi,P. (2012) Coordination of the transcriptome and metabolome by the circadian clock. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 5541–5546.
18. Feng,D. and Lazar,M.A. (2012) Clocks, metabolism, and the epigenome. *Mol. Cell*, **47**, 158–167.
19. Krishnaiah,S.Y., Wu,G., Altman,B.J., Growe,J., Rhoades,S.D., Coldren,F., Venkataraman,A., Olarerin-George,A.O., Francey,L.J., Mukherjee,S. *et al.* (2017) Clock regulation of metabolites reveals coupling between transcription and metabolism. *Cell Metabol.*, **25**, 961–974.
20. Yagita,K., Horie,K., Koinuma,S., Nakamura,W., Yamanaka,I., Urasaki,A., Shigeyoshi,Y., Kawakami,K., Shimada,S., Takeda,J. *et al.* (2010) Development of the circadian oscillator during differentiation of mouse embryonic stem cells in vitro. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 3846–3851.
21. Azzi,A., Dallmann,R., Casserly,A., Rehrauer,H., Patrignani,A., Maier,B., Kramer,A. and Brown,S.A. (2014) Circadian behavior is light-reprogrammed by plastic DNA methylation. *Nat. Neurosci.*, **17**, 377–382.
22. Haspel,J.A., Chettimada,S., Shaik,R.S., Chu,J.H., Raby,B.A., Cernadas,M., Carey,V., Process,V., Hunninghake,G.M., Ifedigbo,E. *et al.* (2014) Circadian rhythm reprogramming during lung inflammation. *Nat. Commun.*, **5**, 1–15.
23. Li,X.M., Delaunay,F., Dulong,S., Claustrat,B., Zampera,S., Fujii,Y., Teboul,M., Beau,J. and Lévi,F. (2010) Cancer inhibition through circadian reprogramming of tumor transcriptome with meal timing. *Cancer Res.*, **70**, 3351–3360.
24. Murakami,M., Tognini,P., Liu,Y., EckelMahan,K. L., Baldi,P. and SassoneCorsi,P. (2016) Gut microbiota directs PPARγdriven reprogramming of the liver circadian clock by nutritional challenge. *EMBO Rep.*, **17**, 1292–1303.
25. Pizarro,A., Hayer,K., Lahens,N.F. and Hogenesch,J.B. (2013) CircaDB: a database of mammalian circadian gene expression profiles. *Nucleic Acids Res.*, **41**, 1009–1013.
26. Gutiérrez-Monreal,M.A., Treviño,V., Moreno-Cuevas,J.E. and Scott,S.-P. (2016) Identification of circadian-related gene expression profiles in entrained breast cancer cell lines. *Chronobiol. Int.*, **33**, 392–405.
27. Masri,S., Papagiannakopoulos,T., Kinouchi,K., Liu,Y., Cervantes,M., Baldi,P., Jacks,T. and Sassone-Corsi,P. (2016) Lung adenocarcinoma distally rewires hepatic circadian homeostasis. *Cell*, **165**, 896–909.
28. Agostinelli,F., Ceglia,N., Shahbaba,B., Sassone-Corsi,P. and Baldi,P. (2016) What time is it? Deep learning approaches for circadian rhythms. *Bioinformatics*, **32**, i8–i17.
29. Hughes,M.E., Hogenesch,J.B. and Kornacker,K. (2010) JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *J. Biol. Rhythms*, **25**, 372–380.
30. Kanehisa,M. and Goto,S. (2000) Kegg: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.
31. Hughes,M.E., Abruzzi,K.C., Allada,R., Anafi,R., Arpat,A.B., Asher,G., Baldi,P., De Bekker,C., Bell-Pedersen,D., Blau,J. *et al.* , (2017) Guidelines for genome-scale analysis of biological rhythms. *J. Biol. Rhythms*, **32**, 380–393.