UNIVERSITY OF CALIFORNIA,
IRVINE

Bioinformatics Software Systems for the Study of Circadian Rhythms

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Computer Science

by

Yu Liu

Dissertation Committee:
Professor Pierre Baldi, Chair
Professor Paolo Sassone-Corsi
Professor Richard H. Lathrop

2018

DEDICATION

To

My parents, for their love and support.

Pierre, for his mentorship and guidance.

My colleagues and friends, especially Nicholas Ceglia,

for his contribution to part of the works.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my advisor, Prof. Pierre Baldi, for providing guidance on many of these projects.

I would like to thank my co-authors, including but not limited to, Dr. Nicholas Ceglia, Dr. Paola Tognini, Dr. Mari Murakami , Dr. Selma Masri and Dr. Paolo Sassone-Corsi for their major contribution across many projects that are cited in this thesis.

I would like the thank the publishers of my published papers, which I have cited in whole or in part, for giving me permission to reprint those papers under fair use for the sole purpose of writing this thesis.

I would like to thank my colleagues, including but not limited to Christophe Magnan, Yuzo Kanomata, Peter Sadowski and Forest Agostinelli, for their contributions to work mentioned in this thesis.

CURRICULUM VITAE

Yu Liu

EDUCATION

Doctor of Philosophy in Computer Science **2018**
University of California, Irvine *Irvine, California*

Bachelor of Science in Biochemistry **2012**
Rice University *Houston, Texas*

RESEARCH EXPERIENCE

Graduate Research Assistant, Baldi Group, Institute for Genomics and Bioinformatics
University of California, Irvine *Irvi*

TEACHING EXPERIENCE

Teaching Assistant **2017–2018**
University of California, Irvine *Irvine, California*

REFEREED JOURNAL PUBLICATIONS

Lung Adenocarcinoma Distally Rewires Hepatic Circadian Homeostasis. **2016**
Cell [37]

Gut microbiota directs PPAR driven reprogramming of the liver circadian clock by nutritional challenge. **2016**
EMBO Reports [41]

Distinct Circadian Signatures in Liver and Gut Clocks Revealed by Ketogenic Diet. **2017**
Cell Metabolism [49]

MotifMap-RNA: a genome-wide map of RBP binding sites. **2017**
Bioinformatics [35]

CircadiOmics: Circadian Omic Web Portal. **2018**
Nucleic Acids Research [42]

SOFTWARE

mafslice/motifmap-core `https://gitlab.ics.uci.edu/yul13/mafslice`
C++ low level IO and compute library for MotifMap and MotifMap-RNA. Referenced in Chapter 2. Contribution: sole contributor.

pyga `https://gitlab.ics.uci.edu/yul13/pyga`
High performance genome annotation library for Python. Referenced in Chapter 2. Contribution: sole contributor.

MotifMap-RNA `https://gitlab.ics.uci.edu/yul13/motifmap_revamp`
Shared Python codebase revamped for MotifMap and MotifMap-RNA. Referenced in Chapter 2. Contribution: sole contributor.

ChIPSeq Enrichment Pipeline `https://gitlab.ics.uci.edu/uci-igb/chipseq_enrichment/com`
Python codebase for importing and analyzing ENCODE ChIPSeq datasets. Referenced in Chapter 2. Contribution: major contributor.

igb-pipelines: Circadian Statistics `https://gitlab.ics.uci.edu/uci-igb/igb-pipelines/tree/`
Python codebase performing statistical analysis on circadian time series data. Referenced in Chapter 3. Contribution: major contributor.

igb-pipelines: Circadian Profiling https://gitlab.ics.uci.edu/uci-igb/igb-pipelines/tree/
Python codebase performing machine learning techniques on circadian data. Referenced in Chapter 3. Contribution: sole contributor.

igb-pipelines: Factor Enrichment https://gitlab.ics.uci.edu/uci-igb/factor_enrichment
Python codebase performing functional annotation and enrichment analysis based on transcription factor binding sites, RNA binding sites or ChIPSeq experimental peaks. Referenced in Chapter 4. Contribution: sole contributor.

igb-pipelines: Pathway Enrichment https://gitlab.ics.uci.edu/uci-igb/mpfe
Python codebase performing functional annotation and enrichment analysis based on functional pathways based on data from CPDB or PathwayCommons. Referenced in Chapter 4. Contribution: sole contributor.

igb-pipelines: Other https://gitlab.ics.uci.edu/uci-igb/igb-pipelines/tree/master/
Rest of IGB Pipelines. Referenced in Chapter 3,4,5. Contribution: major contributor.

pyfuncgraph https://gitlab.ics.uci.edu/uci-igb/pyfuncgraph
Python library for creating regulatory graphs based on functional annotation derived from binding sites or functional pathways. Mentioned in Chapter 4. Contribution: sole contributor.

viz https://gitlab.ics.uci.edu/uci-igb/viz
Python templating and visualization library. Mentioned in Chapter 5. Contribution: major contributor.

goworg https://gitlab.ics.uci.edu/uci-igb/goworg
Go web back end for hosting circadian project websites and analysis results. Mentioned in Chapter 5. Contribution: sole contributor.

circadiomics https://gitlab.ics.uci.edu/uci-igb/public_circadiomics
Web portal for circadian omic data. Mentioned in Chapter 6. Contribution: major contributor.

PyCircadiOmics https://gitlab.ics.uci.edu/uci-igb/PyCircadiomics
Integrated Python software library for comprehensive circadian study. Mentioned in Chapter 7. Contribution: major contributor.

ABSTRACT OF THE DISSERTATION

Bioinformatics Software Systems for the Study of Circadian Rhythms

By

Yu Liu

Doctor of Philosophy in Computer Science

University of California, Irvine, 2018

Professor Pierre Baldi, Chair

The study of circadian rhythms increasingly relies on high throughput circadian omic data. Processing, annotating and analyzing these data require an efficient and powerful bioinformatics software system. In this thesis, I highlight the implementation of such a system. Major components of this system include genome-wide binding site predictors, statistical and machine learning pipelines, functional enrichment pipelines, visualization tools and a data management server. Using this system, we have produced high impact publications, a circadian web portal (CircadiOmics, http://circadiomics.ics.uci.edu/), and advanced our understanding of the organization of the circadian transcriptome.

Chapter 1

Introduction

Circadian rhythms are biological processes that display an oscillatory behavior, in an endogenous and entrainable manner, following a 24-hour clock. Circadian rhythms are ubiquitous in life and deeply rooted in evolution–they can be observed in life forms ranging from cyanobacteria to human [43]. On a macroscopic scale, circadian rhythms are vital for many physiological processes, including the sleep/wake cycle, hormone secretion, diet related metabolism, and neural function [9, 14, 19, 39]. As a consequence, loss or disruption of circadian rhythms can result in diseases and health issues such as diabetes, obesity, and premature aging [44, 18, 6, 1, 47]. On a microscopic scale, circadian rhythms manifest as oscillations of molecular species such as transcripts, proteins, and metabolites that interact with one another [24]. One of the most important and well studied types of molecular oscillations is that of the transcripts, which is the focus of this thesis. It is known that transcriptomic oscillations are pervasive and well organized [45, 14, 32]. Transcripts do not oscillate in isolation–they form regulatory feedback loops that oscillate in a network of coupled oscillators [38, 46, 45]. At the center of this transcriptomic network is a set of highly coordinated regulators in the form of transcription factors (TF), including CLOCK, BMAL, REV-ERB, CRYs, PERs and DBP [31]. This is known as the *core clock*. Moreover, oscilla-

tory metabolites and proteins can interact with oscillatory transcripts and their regulatory system [17, 13, 16, 33]. The resulting landscape of molecular oscillations is astoundingly complex and robust.

One of the most important findings in the study of the circadian transcriptome is that almost all transcripts are capable of oscillation [45]. In a typical transcriptomic dataset, $\sim 10\%$ of measured transcripts are found to oscillate [16, 15, 38, 37, 41]. However, the intersection of oscillating transcripts between any two conditions is only $\sim 2\%$ [45]. As a result, the circadian landscape in a cell can be drastically different depending on genetic and epigenetic conditions [32, 16, 53, 45]. 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, 34, 7, 41]. Logically, it follows that the union of all oscillating transcripts is large–indeed, it is estimated that over 90% of transcripts can become circadian through reprogramming. The underlying mechanism guiding reprogramming and global organization of circadian rhythms is of great interest for the circadian field.

Due to the pervasiveness, complexity and importance of circadian rhythms, the circadian field is emerging as a frontier of biological research. Coincidentally, it is also a field that is increasingly reliant on "big data" and high-throughput bioinformatics analysis [24]. As such, there is a strong need for a complete software system tailored for the analysis of high throughput circadian omic data. In this thesis, I discuss our contribution [1] in this area. Our work focuses on three main areas: first, generating "big data" as resources for circadian study; second, performing novel analyses of circadian omic data; third and finally, engineering of software systems that integrate said analyses. As a culmination of our work, we also try to tackle one of the most fundamental and difficult questions in this field, namely the aforementioned transcriptomic organization of circadian rhythms. Substantial results,

[1] Unless otherwise stated, I am the main contributor of most of the projects listed. Some of the projects were contributed equally by Nicholas Ceglia & me. A detailed breakdown of contributions is included in Appendix 2 of this chapter.

generated by applying all the tools that are described in this thesis, are shown in the last chapter (publication pending).

1.1 Circadian Omic Data and Challenges

The circadian field is not the only field in biology that utilizes more and more high throughput omic data. Indeed, this trend is observed generally in biology, where the number of publication involving omic data has multiplied by 10X from 2002 to 2016 [24]. In the same period of time, the number of circadian omic papers has also increased by more than 6X, with a mix of micro-array, RNA-Seq, ChIP-Seq proteomics and metabolomic papers [24]. High throughput omic data differs from traditional biological data in quantity, complexity and nature of variability [20]. This in turn raises new challenges for the field [24, 22, 29, 20, 3, 48]:

1. **Challenge 1:** High throughput circadian omic data must be processed from their raw form (e.g. RNASeq reads) to usable data (e.g. RPKM values) [2] and to meaningful interpretations (e.g., enriched pathways). The speed by which this process happens is the new bottleneck for circadian study.

2. **Challenge 2:** Large scale data requires the usage of increasingly complex models. This produces a classic bias-variance trade-off situation. As a result, a combination of statistical and machine learning techniques must be employed to account for noise and achieve robustness.

3. **Challenge 3:** Different forms of omic data (e.g., genomics, proteomics, metabolomics) need to be integrated with each other. Additionally, experiment-specific data need to

[2] This thesis largely omits discussion of this particular processing step (from raw data to quantitative input data). This is because, for each type of high throughput omic data, there already exist mature and reliable software systems for this processing process. For example, RNASeq data can be readily processed by the Tuxedo protocol [50], or kallisto [10].

be integrated with existing knowledge bases such as Gene Ontology (GO), KEGG pathways and transcription factor binding sites (TFBS).

4. **Challenge 4:** Results need to be properly managed and stored. The management system also needs to deliver results in a rapid, user-friendly way for biologists who may not be versed in quantitative methods (i.e., via visualization). Such results should also be publicly available if possible.

Given these challenges, it is not hard to arrive at the conclusion that a robust bioinformatics approach with efficient software systems is the proper solution for the analysis of circadian omic data. Indeed, in this thesis I present our work in this area, which is of a pioneering nature.

1.2 Overview: Bioinformatics Software Systems for Circadian Rhythms

Most of our work originated as individual projects that were designed to be inter-operable with each other. However, by now most code or logic of the individual projects has been incorporated into three main systems: **IGB-pipelines** integrates multiple data processing and analysis pipelines which are used for individual experimental datasets. **CircadiOmics** is a major web portal for circadian data that integrates data storage and visualization with a focus on providing the most comprehensive set of circadian omic data for the public. **PyCircadiomics** integrates data mining, machine learning and bioinformatics tools to conduct global analysis of the circadian transcriptome.

Nevertheless, to focus on individual components with clarity, in this thesis I will first present them in their original form as subsystems in the last three chapters. How and where they are

incorporated into one of the larger systems is mentioned in chapters 2-6. Novel algorithms, statistical and machine learning methods are noted as well. Each of the three main systems also has one chapter summarizing its components and other novel designs. Application of any of the systems is also discussed in each chapter.

Additionally, particularly worth mentioning is BIO_CYCLE [2], which is a deep learning based approach to predict rhythmicity. Obviously, to conduct large scale analysis of biological oscillations, it is vital to have a highly reliable and tunable method for predicting such oscillations, which forms the basis for all subsequent analyses. BIO_CYCLE provides such baseline data for most of the components of our system, which can also function using another older method for detection (JTK_CYCLE [25]). That said, BIO_CYCLE is otherwise outside of the scope of this thesis.

A detailed breakdown of the chapters is listed below:

- **Chapter 2** describes `MotifMap-RNA` and `ChIPSeq-pipeline`. These general purpose genomics projects generate global functional data which can be vital for interpretation of circadian signals (Challenge 1).
- **Chapters 3 & 4** describe novel statistical tools, machine learning tools and functional analysis tools for individual circadian datasets. These chapters also describe various databases and software infrastructures for combining quantitative data and functional annotation data. They address Challenges 1-3 and are also components of `igb-pipelines`.
- **Chapter 5** describes visualization and rapid client side delivery solutions. They constitute the front end and integration layer of `igb-pipelines` and address Challenges 3 and 4.
- **Chapter 6** describes `CircadiOmics`, a web portal for ciradian data. CircadiOmics

provides the largest repository of high throughput circadian omic data for the general public (Challenge 4).

- **Chapter 7** describes `PyCircadiomics`, which integrates all of the data and analyses, applying them globally to understand the organization of the circadian transcriptome.
- **Chapter 8** gives a short summary.

By designing individual pipelines and integrating them into larger systems, we have produced a powerful yet fully extensible software system for the study of circadian rhythms, focusing on transcriptomic organizations. By combining specialized pipelines and libraries, we have formulated a data-driven bioinformatics framework which can competently address questions in circadian biology. Moreover, results generated using this framework are readily understandable and usable by biologists. As a result, our work and collaborations have produced high impact publications. Finally, we have advanced the understanding of one of the most fundamental problems in circadian biology, namely the organization of the circadian transcriptome.

Chapter 2

MotifMap-RNA

2.1 Introduction: Genome-wide Computational RNA Binding Sites Prediction

The primary mode for transcriptomic regulation is regulation by transcription factors (TFs). An important extension to TF regulation is post-transcriptional control, induced by RNA binding proteins (RBPs). Therefore, knowledge of where and how TFs and RBPs bind to and thus regulate their targets is fundamental to transcriptomic studies at all levels.

Circadian transcriptomic datasets typically consist of micro-array, RNASeq and ChIPSeq datasets. While ChIPSeq datasets themselves measure TF or histone bindings, the vast majority of other datasets do not intrinsically contain such information. Therefore, augmenting such datasets with functional information in the form of TF and RBP binding sites can be very important.

In terms of TF binding sites (TFBS), MotifMap [12, 52] provides a systematic computational prediction throughout the genome. Importantly, the quality of MotifMap predictions are

generally consistent and tunable by adjusting filtering parameters. These TFBS results, although purely computational, have been proven to be highly applicable for tasks such as identifying key TFs within a dataset, identifying enrichment of targets of certain TFs and identifying key pathways in a dataset [12].

A natural extension of TFBS prediction is the prediction of RBP binding sites, which accounts for post-transcriptional regulation such as alternative splicing. Therefore, I developed **MotifMap-RNA** for generating system-wide computational predictions of RBP binding sites. While MotifMap-RNA is not directly related to circadian rhythms, it serves as a knowledge base of possible post-transcriptional control. Along with the original MotifMap, these two methods constitute an important building block for other projects.

Fundamentally, MotifMap and MotifMap-RNA employ the same principle to predict binding sites, with two major components: first, binding motifs, in the form of positional weight matrices (PWMs), are used to generate binding scores for all genomic sequences in a species. Sites with the highest scores are considered to have the highest probability for potential binding by the factor, quantified as a form of normalized z-scores (NLOD). Second, phylogenetic evidence, in the form of multiple species alignments (MSA), is used to estimate how conservative each of the top binding sites is. Conservation rates are used as weights for calculation of a Bayesian Branch Length Score (BBLS). This approach has been proven to be effective in reducing false positive rates [52]. In addition, MotifMap-RNA uses a few novel statistical tools to improve performance relative to RBP, which is detailed later in this chapter. Due to the fact that the results are systematic and completely scalable by tuning parameters, MotifMap and MotifMap-RNA can often behave better than experimental ChIPSeq results when applied to high throughput omic data. That is to say, by applying a stricter filtering parameter such as the BBLS, predicted binding sites will reliably become less prone to false positives, and the scaling of the false discovery rate (FDR) is consistent across different predicted motifs [35, 12].

However, both methods face new challenges raised by the rapidly increasing data quantity in the field of genomics. More specifically, both methods rely on generating knowledge based on multiple sequence alignment files (MAF) for the relevant species, which is provided by UCSC (http://hgdownload.soe.ucsc.edu/downloads.html). For mouse alone, the latest genome build (mm10) contains alignment to 59 other species, doubling the amount of data from the previous iteration (mm9). As such, the old MotifMap codebase, which was developed in pure python, struggles to perform well for large scale predictions of newer genomes. This is also an example of Challenge 1 mentioned in the introduction. To address the performance needs of these systems and to ensure that they can scale to future genome builds which may further grow in size, I engineered a low level software library named `motifMap-core` which provides high performance IO and compute functionalities for both projects.

2.2 Overview of the MotifMap-Core Infrastructure

`motifmap-core` is a C++ library that handles low level algorithms for both MotifMap and MotifMap-RNA. It consists of three main components: `mafslice` handles IO related to genomic sequences and sequence alignments. `motifmap-compute` is a C++ wrapper for MOODS (https://www.cs.helsinki.fi/group/pssmfind/) which handles sequence scoring against PWMs. `pyga` is a high performance genome annotation tool for generating basic annotations accompanying each predicted binding site.

A dependency graph showcasing the major classes is generated by the software doxygen (https://www.stack.nl/~dimitri/doxygen/) and is shown in Figure 2.1.

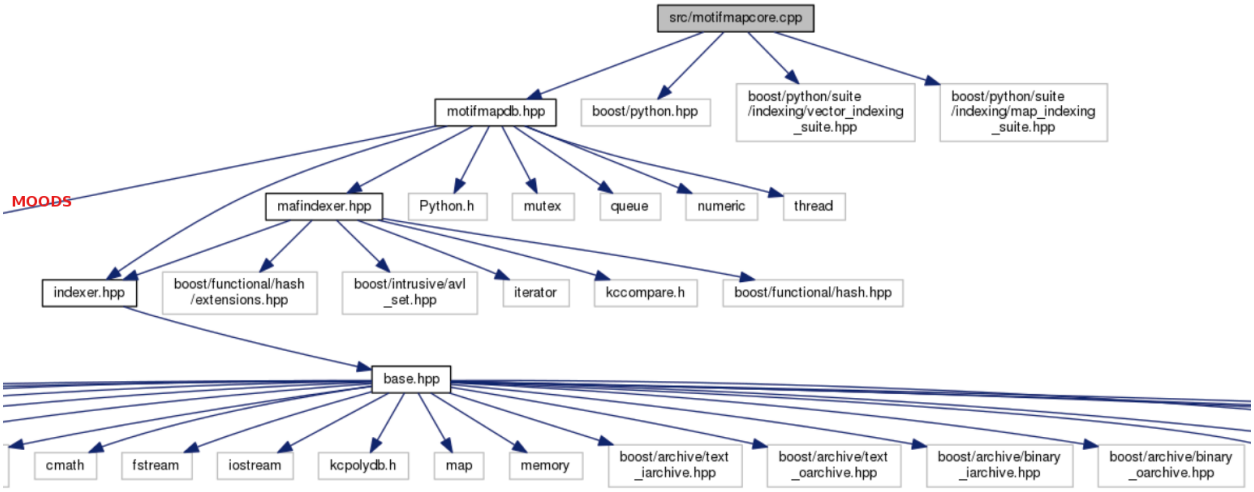


Figure 2.1: Class dependency graph for motifmap-core

2.2.1 `mafslice`

One major performance bottleneck for both MotifMap and MotifMap-RNA is the retrieval of the MAF hits, which are plain text records that describe instances where sequences of a reference genome (e.g., human) intersect with its phylogentic neighbors (e.g., chimp, mouse etc). Below is an example of a typical MAF hit:

```
a score=1500.000000
s hg38.chr1             11470 1 + 248956422 G
s panTro4.chr15         14084 1 -  99548318 G
i panTro4.chr15     C 0 C 0
s calJac3.chr9      112497732 1 - 124281992 G
i calJac3.chr9      C 0 C 0
s saiBol1.JH378170   10178365 1 -  12910020 G
i saiBol1.JH378170 C 0 C 0
s bosTau8.chr5      107494275 1 + 121191424 g
i bosTau8.chr5      C 0 C 0
```

```
s myoLuc2.GL429865     46543 1 -   4619595 g
i myoLuc2.GL429865 N 0 C 0
```

This record shows, in addition to other information, that there is an alignment of evolutionary neighbors around the sequence at human chromosome 1, position 11470. The intersection is only for one one base pair. A whole human genome (hg38) can contain more than 100,000,000 similar MAF hits, totaling a size of 800GB in plain text format. A random range based retrieval task appears when one calculates the likely binding sites, which may be found in widespread locations in the reference genome. The range of a single binding site can also correspond to multiple MAF hits. The algorithm of MotifMap and MotifMap-RNA requires that one retrieve all intersecting sequences, align them or extend them from their original host genomes, and then recalculate BBLS based on scores on all individual aligning sequences [52].

`mafslice` is the IO library designed for such a task. It utilizes a heterogeneous data structure for efficient retrieval and processing of MAF hits. Briefly, records are stored on an on-disk hashtable (one hastable or database per chromosome, implemented by Kyoto-Cabinet http://fallabs.com/kyotocabinet/), whose keys correspond to the genomic starting locations plus an offset (encoded in two bytes). Hashing is done via murmurHash3 (https://sites.google.com/site/murmurhash/) because of its excellent performance. The keys are then extracted and stored in an index tree, which resides fully in memory at runtime. The index tree is based on boost-avl-tree (https://www.boost.org) which is sorted by genomic coordinates. Since the index tree allows for range query in log n time (in memory) and the retrieval of records of the on-disk hashtable is constant time (on-disk), this constitutes a dramatic performance improvement over previous solutions [1] where the query takes up to log n time on-disk. Indeed, profiling done in the mouse genome mm10 shows an IO performance gain up to 20X.

[1] pygr, https://omictools.com/pygr-tool

Additionally, a novel hit merging algorithm was developed to efficiently combine records and reduce fetching of original sequences when it is unnecessary. A brief excerpt of the code is below:

```
 auto inner = [](const char &c, std::string &out, int &cnt, int &gap) {
   if (c != '-') {
     out.push_back(c);
     ++cnt;
     return true;
   } else
     ++gap;
   return false;
 };
...
for (auto &it : hits.second) {
   int t_lshift = 0, t_dist = 0, t_gap1 = 0, t_gap2 = 0;
   std::string seq1(""), seq2("");
   for (int i = 0; i < lshift + gap1; ++i) {
     char c = it.second.seq[i];
     inner(c, seq1, t_lshift, t_gap1);
   }
   for (int i = lshift + gap1; i < lshift + gap1 + dist + rs.first; ++i) {
     char c = it.second.seq[i];
     inner(c, seq2, t_dist, t_gap2);
   }
   it.second.seq = routine(it.second.seq);
   interval itt;
```

```
    itt.seq = seq2;
    itt.ref = it.second.ref;
    itt.chr = it.second.chr;
    itt.strand = it.second.strand;
    itt.score = it.second.score;
    itt.l = t_lshift;
    itt.r = t_dist; // local change
    out.second[itt.ref] = itt;
 ...
 }
```

In this code snippet, `hits.second` contains all of the MAF hits for a particular binding site prediction. Each of these sequences partially match the range of the prediction, but there are typically gaps in alignment. In order to combine all of them, the helper function `inner` is used to scan their alignments and remove gaps (annotated as `-`) while incrementing a counter for gaps. This gap counter is then used both on the left end and the right end of the alignment to match the aligning sequence back to the reference. Finally, the correct range for the aligning sequence is updated in place while the sequence is stitched together from the original input. As long as the updated range falls into an interval suitable for matching the whole of the predicted sequences, there is no need to further retrieve it from the original genome. This is a significant improvement over the old algorithm, where only coordinates, not matching sequences, were utilized.

2.2.2 motifmap-compute

`motifmap-compute` is yet another C++ library wrapping the sequence aligning algorithm MOODS (https://www.cs.helsinki.fi/group/pssmfind/). It utilizes the advanced scan function provided in the MOODS C++ library over simple PWM multiplication, which is an improvement over the old MotifMap. The advanced scan takes into consideration of background distribution of base pairs over the whole genome when calculating PWM scores and is thus more accurate in species where there is a strong C/G bias in the genome. Another improvement of `motifmap-compute` is its use of parallelism, which cannot be achieved in similar efficiency in pure python. Briefly, it generates compute threads at the C++ level and then releases the python global interpretor lock (GIL) to run all of them as real processes. A code snippet demonstrating this functionality is attached below:

```
void compute(const int &cnt, const std::string &motif, const std::vector<matrix> &mats,
             const std::vector<double> &bg, const std::vector<double> &ths) {
  try {
    if (threads.size() > 100) {
      std::vector<std::thread> ths;
      std::cerr << "Releasing GIL" << std::endl;
      releaseGIL unlock = releaseGIL();
      std::cerr << "Too many threads, joining 100" << std::endl;
      for (auto &c : threads) {
        ths.push_back(c.spawn());
      }
      for (auto &th : ths)
        th.join();
      threads.clear();
    }
```

```
      std::cerr << " Creating compute object...";
      threads.push_back(motifmapcompute(inp, motif, mats,bg, ths, cnt));
    } catch (...) {
      std::cerr << "Compute failed on " << print_interval(inp.first)
                << std::endl;
    }
    return;
  };
```

2.2.3 pyga

Once predicted binding sites are produced by MotifMap or MotifMap-RNA, an important task is to annotate them with biological meanings. Sites are predicted with genomic coordinates (i.e., which chromosome, starting point and strand of DNA/RNA). Annotations include the closest gene, the distance to said gene, and if the binding site is within the location of the gene, whether or not it is within a functional part of this location (e.g. in the promoter region). The old MotifMap utilizes MySQL to match hits in database, which has proven to be suboptimal for performance. Other tools for annotation, such as bedtools (https://bedtools.readthedocs.io/en/latest/), lack a convenient or scalable python interface.

pyga is designed for such high performance genome annotation tasks. It contains a C++ layer and a set of python interfaces for ease of use. Importantly, it has a very low memory profile ($< 1Mb$ when loaded with a single genome annotation) and fast performance, which means it is well suited for parallel annotation tasks or on-demand annotation tasks that are performed multiple times during a MotifMap or MotifMap-RNA run. Additionally it also contains functionalities for generating random sequences and scanning base pair distributions, which are again used by both MotifMap and MotifMap-RNA.

`motifmap-core` combines the aforementioned components in a boost-python templated class, which can be conveniently used in python once compiled. Compilation is done via cmake. All MotifMap and MotifMap-RNA results are generated via motifmap-core and stored in MySQL databases, with rich querying and filtering functionalities (detailed in the paper attached later in this chapter).

2.3 Novel Methods in MotifMap-RNA

In addition to the improvement in software infrastructure and low level routines in terms of IO and compute, MotifMap-RNA improves upon the original MotifMap in its statistical methods. These new statistical methods were inspired by the specific properties of RBP binding sites. For example, it is known that RBP binding sites tend to cluster over a larger range than TFBS [51]. As such, two new forms of scores were devised for MotifMap-RNA. The following is an excerpt from the supplementary material of the MotifMap-RNA paper [35]:

" Two forms of meta z-scores were generated by aggregating z-scores in the vicinity of each hit, extending up to 150 bps on each side. This should have the effect of reflecting the clustering of local high scoring binding sites and was shown in the validation to improve performance when validated by eCLIP peaks. Briefly, the two meta z-scores are defined as: $\mathbf{ExpZ} = \sum_i \frac{1^{r_i-1}}{2} Z_i$ and $\mathbf{StoufferZ} = \frac{\sum_i r_i^{-1} Z_i}{\sqrt{\sum_i r_i^{-2}}}$, where is the index of the hit, is the z-score and is the 1-based local rank. Both meta z-scores are weighted sum of the local z-scores, whose weights are a function of ranks. In ExpZ the weight decays exponentially, while StoufferZ uses a variant of Stouffers method with linear weight decay. In calculating these scores, the flanking sequence was allowed to extend beyond the original boundary to obtain even flanking distances everywhere. "

Additionally, MotifMap-RNA also utilizes eCLIP peaks, which is a very new class of experimental data for validation of predicted results. eCLIP experiments produce broad peak signals that inform of RBP bindings in the transcriptome. The processing of such experimental data and extraction of samples for validation tests (e.g., generating ROC curve) is a relatively novel subject. MotifMap-RNA utilizes an approach based on filtering peak signals and sampling background information to produce sufficient samples from limited experimental data (e.g., 2000 samples from 500 peaks). The following is an excerpt from supplementary material of the MotifMap-RNA paper [35]:

" Because of the fact that the binding may be tissue specific and are also sensitive to the amount of RNAs expressed in the cell, it is almost certainly not globally comparable to MotifMap results. The fact that a certain high scoring hit is not found to be within a positively enriched peak is not necessarily indicative of the hit being false, but may be indicative of a variety of biological and experimental processes resulting in the lack of binding shown in the eCLIP experiment. As such, we restricted our validation to peaks with high degree of certainty (highly positively enriched and negatively enriched peaks). For positively enriched peaks, a log-p value threshold of $>= 2$ was used, and up to 1000 top positive peaks were selected as ground truth positives. For negative peaks, due to the lack of highly negative peaks, a log-p threshold of $<= 0$ was used. To even out the number of positive and negative samples, random sequences from the genome were also introduced as negative samples if they were not found to be overlapping any positive peaks. Enough random sequences were taken such that the total number of positive and negative samples were equal. "

2.4 The MotifMap-RNA Web Portal

As mentioned in Challenge 4 in Chapter 1, availability of results and ease-of-use is crucial for bioinformatics projects in the age of big data. MotifMap-RNA addresses this by providing

a modern, fast and intuitive web portal (http://motifmap-rna.ics.uci.edu). The user can query the full results of RBP hits from all motifs using a variety of parameters to filter and sort the data. The user can further filter the results using distances of the hits from closest genome markers. Moreover, all of the data is presented in a fully interactive table allowing the user to further search, query and download. A brief user guide is available at http://motifmap-rna.ics.uci.edu/help.

2.5 MotifMap-RNA

A full paper of MotifMap-RNA is appended here to fully demonstrate the motivation, results and impact of the project.

Bioinformatics, 33(13), 2017, 2029–2031
doi: 10.1093/bioinformatics/btx087
Advance Access Publication Date: 22 February 2017
Applications Note

OXFORD

Genome analysis

MotifMap-RNA: a genome-wide map of RBP binding sites

Yu Liu[1], Sha Sun[2], Timothy Bredy[3], Marcelo Wood[3], Robert C. Spitale[4,*] and Pierre Baldi[1,*]

[1]Department of Computer Science and Institute for Genomics and Bioinformatics, [2]Department of Developmental and Cell Biology, [3]Department of Neurobiology and Behavior and [4]Department of Pharmaceutical Sciences and the Chao Family Comprehensive Cancer Center, University of California, Irvine, CA 92697, USA

*To whom correspondence should be addressed.
Associate Editor: Inanc Birol

Received on November 24, 2016; revised on January 20, 2017; editorial decision on February 6, 2017; accepted on February 7, 2017

Abstract

Motivation: RNA plays a critical role in gene expression and its regulation. RNA binding proteins (RBPs), in turn, are important regulators of RNA. Thanks to the availability of large scale data for RBP binding motifs and *in vivo* binding sites results in the form of eCLIP experiments, it is now possible to computationally predict RBP binding sites across the whole genome.
Results: We describe MotifMap-RNA, an extension of MotifMap which predicts binding sites for RBP motifs across human and mouse genomes and allows large scale querying of predicted binding sites.
Availability and Implementation: The data and corresponding web server are available from: http://motifmap-rna.ics.uci.edu/ as part of the MotifMap web portal.
Contact: rspitale@uci.edu or pfbaldi@uci.edu
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

RNA serves not only as a messenger between DNA and protein, but also as a regulator of important processes such as genome organization and gene expression (Morris and Mattick, 2014). RNA itself is regulated by a diverse collection of RNA binding proteins (RBPs), which are responsible for an array of functions such as alternative splicing, RNA modification, polyadenylation, mRNA transport and translational regulation (Glisovic *et al.*, 2008; Mercer *et al.*, 2009). RBPs typically bind to their targets via one or more RNA binding domains (RBDs) which are thought to have specific binding motifs (Lunde *et al.*, 2007). Due to the large number of known and predicted RBPs and their important role in RNA regulation, there has been much interest in systematically understanding their behavior.

Recently, large scale *in vivo* surveys have been carried out to discover the binding motifs of a large number of RBPs (Cook *et al.*, 2011; Ray *et al.*, 2013). At the same time, high throughput *in vivo* eCLIP experiments have been effective in identifying the RBP bindings to RNAs in human immortalized and primary cells (Tollervey *et al.*, 2011; Van Nostrand *et al.*, 2016). Together, these biological data provide a foundation for systematically predicting RBP binding sites across the whole genome and validating them. Previously computational methods have been described to predict motif specific RBP binding sites for a given sequence or a range of sequences (Paz *et al.*, 2014; Zhang *et al.*, 2013). However, to our knowledge, there is no service that allows systematic, genome-wide binding site querying.

Here we describe MotifMap-RNA, a novel extension of MotifMap (Daily *et al.*, 2011; Xie *et al.*, 2009), a system for transcription factors binding site prediction, to RBP binding sites. MotifMap-RNA predicts *z*-score based binding sites specific to RBP motifs across the human and mouse genomes. It also allows the user to filter and sort the results based on clustering of local binding sites, represented by weighted *z*-scores, or evolutionary conservation, quantified by Bayesian branch length scores (BBLS). Furthermore, we organized genomic sequences into 4 major classes: UTRs, intronic regions, lncRNAs and miRNAs, for all of which we generated class specific model parameters. Finally, we implemented a web server which allows the user to interact with MotifMap-RNA results through a friendly interface.

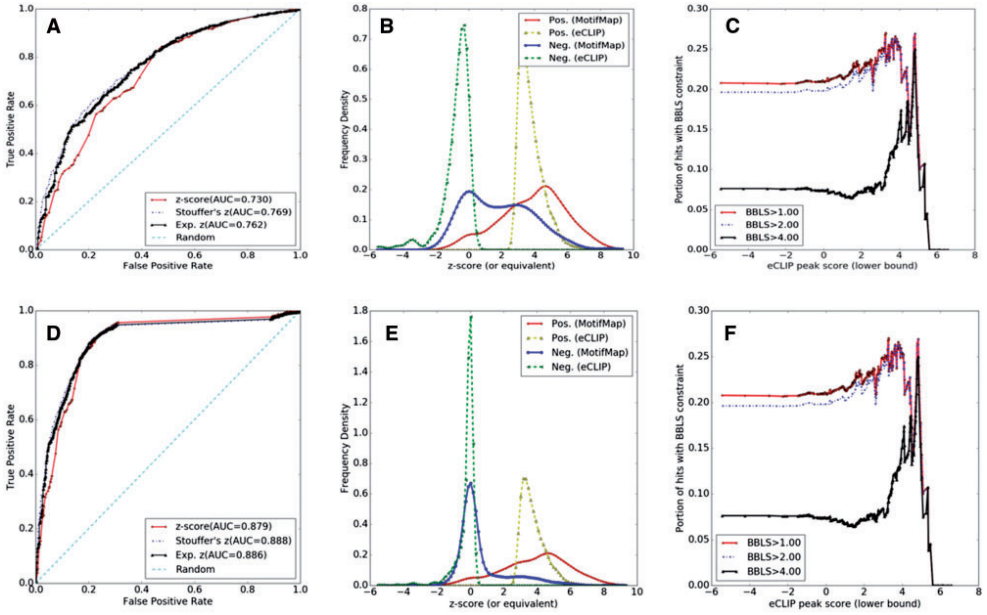


Fig. 1. (**A**) ROC curve of a representative RBP (HNRNPK) from 3 types of *z*-scores: raw, exponentially weighted *z*-score and Stouffer's *z*-scores using rank as weights (details in supplement). Results show that aggregating *z*-scores improves the AUC performance. (**B**) Distribution of ground truth peak scores for both negative and positive examples used in ROC calculation, and their corresponding aggregate *z*-scores from MotifMap (exponentially weighted, corresponds to black solid line in A). (**C**) Portion of MotifMap hits with BBLS scores greater or equal to the marked thresholds from hits corresponding to (positive) peaks at different peak score cutoffs. This shows the relative amount of highly conserved hits increases as the peak score increases, i.e. the most positively enriched peaks tend to overlap highly conserved MotifMap hits. However, the total portion of conserved hits is low (< 0.3). (**D**). Same as (A) but with random sequences added as negative examples. This shows that random sequences improves the AUC performance but retains the same trend. (**E**). Same as (B) but with random sequences as negatives. MotifMap-RNA is able to identity random negatives effectively. (**F**). Same as (C) but with random sequences as negatives. BBLS score is not sensitive to the addition of random negatives, since high BBLS scores are concentrated on a portion of the most positive peaks (Color version of this figure is available at *Bioinformatics* online.)

2 Materials and methods

2.1 Motif and genomic data collection

We obtained experimental and computational RBP binding motif data, in the form of positional weight matrices (PWMs), from RBPDB and CISBP (Cook *et al.*, 2011; Ray *et al.*, 2013). In total, we curated 371 PWMs, 266 of which are from human, 22 from mouse and 83 from other sources. We estimated that these motifs correspond to approximately 235 unique RBPs. We also downloaded the latest human and mouse genome assemblies (hg38, mm10) and their multiple species alignments from the UCSC genomic browser (http://genome.ucsc.edu/). We filtered the genomic sequences into 4 classes by annotations: untranslated regions (UTRs), intronic regions, long non-coding RNA (lncRNA) and miRNA (sources in supplement).

2.2 Scoring

We scanned each class of filtered genomic sequences using curated motifs, calculating *z*-scores with class-specific mean and variance to achieve better specificity. For each sequence in a class (e.g. 5'UTR of a particular gene), we filtered the top scoring motif binding sites in terms of *z*-scores (only positive *z*-scores were considered), up to 3 on each strand. These were considered hits from the motif. Per sequence hits were chosen over hits with the highest absolute *z*-scores to maximize the coverage across the entire genome.

We also incorporated additional metrics to measure the hits. Some RBP bindings tend to be locally clustered (Ule *et al.*, 2006). As such, Two forms of weighted *z*-scores were used to reflect local clustering of high *z*-score hits from the same motif. In addition, RBP binding sites can be less conserved than TF binding sites (Gerstberger *et al.*, 2014; Vaquerizas *et al.*, 2009). Conservation scores in the form of BBLS were also generated using method described in the original MotifMap (Xie *et al.*, 2009). Details about scoring and filtering of hits are described in the Supplementary Material.

3 Results

Overall we generated binding predictions for 371 motifs in 4 classes of human and mouse genomic sequences. The total number of hits is typically between 100 000 and 200 000. While the amount of hits can be enormous due to the short and degenerate nature of some motifs, which may produce lower quality hits, the user can effectively filter out a small set of hits of interest using a combination of aforementioned metrics through the web portal.

3.1 Validating the quality of *z*-score and BBLS hits

In order to validate the quality of the predictions, we downloaded eCLIP results for 12 RBPs from the ENCODE project (https://www.encodeproject.org/; Van Nostrand *et al.*, 2016) and generated ROC

curves for matching MotifMap-RNA results. As an example, UTR results from the RBP HNRNPK are shown in Figure 1A–C (from HepG2 tissue, Replicate 1). Notably, aggregating *z*-scores improves the AUC performance while high BBLS scores tend to concentrate on highly positive peaks.

Due to the fact that many eCLIP results lack sufficient negative examples for ROC curve estimation, we included random sequences not overlapping any positive eCLIP peaks as extra negatives. Their effect on HNRNPK results are shown in Figure 1D–F. AUC performance generally improves while BBLS performance remains consistent.

Overall, with random sequences added, we obtained an average AUC of 0.76 for *z*-scores in the UTR region, and 0.68 for lncRNA region. For details on the validation method, see the supplement.

Additionally, Fisher's exact test was applied to MotifMap-RNA results which overlap positive or negative peaks. In all tested cases, MotifMap-RNA hits significantly overlap more positive peaks. Comparison to existing method (RBPmap) also shows favorable results (details in Supplementary Tables S1 and S2).

3.2 Web server

We constructed a database to host the results and implemented the MotifMap-RNA web portal, which provides the user a friendly interface to effectively find, filter, sort and navigate the binding site results in two different modes: motif search and gene search. In motif search, the user can obtain an interactive table containing results from all of the hits of the selected motif, filtered and sorted by a variety of parameters. In gene search, instead of selecting one motif, the user can input a gene symbol or an annotation ID (e.g. miRNA accession), and search for hits from all motifs to that target (details in Supplementary Material).

4 Conclusion

In conclusion, MotifMap-RNA is a novel system for genome-wide querying of RBP binding sites. Together with its friendly interface, it will assist users in their investigations of RBPs and RNA regulation, and the fundamental roles they play across multiple biological processes.

Conflict of Interest: none declared.

References

Cook,K.B. *et al.* (2011) RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res.*, **39**, 1–8.

Daily,K. *et al.* (2011) MotifMap: integrative genome-wide maps of regulatory motif sites for model species. *BMC Bioinformatics*, **12**, 495.

Gerstberger,S. *et al.* (2014) Evolutionary Conservation and Expression of Human RNA-Binding Proteins and Their Role in Human Genetic Disease. In: Yeo,G.W. (ed.), *Advances in Experimental Medicine and Biology*, Vol. **825**. Springer, New York, NY. pp. 1–55.

Glisovic,T. *et al.* (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.*, **582**, 1977–1986.

Lunde,B.M. *et al.* (2007) RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell Biol.*, **8**, 479–490.

Mercer,T.R. *et al.* (2009) Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.*, **10**, 155–159.

Morris,K.V. and Mattick,J.S. (2014) The rise of regulatory RNA. *Nat. Rev.. Genet.*, **15**, 423–437.

Paz,I. *et al.* (2014) RBPmap: A web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res.*, **42**, 1–7.

Ray,D. *et al.* (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, **499**, 172–177.

Tollervey,J.R. *et al.* (2011) Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat. Neurosci.*, **14**, 452–458.

Ule,J. *et al.* (2006) An RNA map predicting Nova-dependent splicing regulation. *Nature*, **444**, 580–586.

Van Nostrand,E.L. *et al.* (2016) Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods*, **13**, 1–9.

Vaquerizas,J.M. *et al.* (2009) A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.*, **10**, 252–263.

Xie,X. *et al.* (2009) MotifMap: A human genome-wide map of candidate regulatory motif sites. *Bioinformatics*, **25**, 167–174.

Zhang,C. *et al.* (2013) Prediction of clustered RNA-binding protein motif sites in the mammalian genome. *Nucleic Acids Res.*, **41**, 6793–6807.

2.6 The ChIPSeq pipeline

While MotifMap and MotifMap-RNA provide comprehensive predictive binding sites for functional studies, experimental data is often times vital for specific studies. To augment MotifMap and MotifMap-RNA, we developed the ChIPSeq pipeline. We then imported more than 140 experimental results, most from the ENCODE project (https://www.encodeproject.org). Data is stored in a MySQL database with similar interfaces to MotifMap and MotifMap-RNA databases. A novel rank based score is introduced in an attempt to make filtered peaks across different experiments more consistent. Briefly, for each experiment, where peak score of some sort is available (e.g., peak pileup value), peaks are ranked by this score and the ranks are used as a non-parametric score. When filtering peaks from multiple experiments, this rank is optionally used as a filtering parameters similar to equivalent parameter in computational results such as MotifMap. Empirical testing shows that this approach improves the results of multi-experimental functional analysis, although only to a limited extent. An example of this application is used in a study of mouse microbiome [41], which is appended at the end of Chapter 3.

2.7 Integration into PyCircadiOmics

MotifMap, MotifMap-RNA and ChIPSeq data are used in PyCircadiOmics to conduct genome wide functional study of circadian regulation. Details of PyCircadiOmics and the functional results are included in Chapter 7.

Chapter 3

Circadian Statistical and Machine Learning Pipelines

3.1 Introduction

One of the most common forms of circadian omic data is measurements of concentrations of molecular species, such as transcripts, proteins or metabolites. Such measurements are usually taken over the course of a day (24hrs) and are collected over at least 4 time points. As a result, a typical dataset contains hundreds or thousands of time series of individual molecular species. Processing this time series data is an important first step of any circadian analysis. Furthermore, machine learning techniques can be applied to individual time series data or collections of time series, typically grouped by circadian behavior under different experimental conditions (i.e., circadian groups).

However, as is typical of high-throughput biological data, circadian omic data introduces a series of challenges for its processing and analysis. First of all, there is intrinsic noise from the biological signals, accompanied by experimental noise particular to each type of experimental

technique. Furthermore, the signals may contain numerous missing data, outlier data or unannotated data. Last but not least, different experiments may produce different intensities of signals, which makes cross-experimental analysis challenging.

3.2 Overview of the Pipelines

The circadian statistical pipeline (`igb-pipelines:circadian_statistics`) was designed to process noisy and variable circadian time series signals. At the very core, it employs either BIO_CYCLE [2] or JTK_CYCLE [25] to predict circadian properties of time series data. The user can choose parameters specific to either software through the interface of the statistical pipeline. Predicted phases, amplitudes, periods and circadian p-values/q-values are collected automatically by the pipeline and used for downstream processing. In addition, the statistical pipeline utilizes a few common methods to improve the robustness of the analysis. First, a Dixon's test is performed, at each timepoint for multi-replicate data, to determine any outlier replicate data to exclude. The test can be performed at either the 90th, 95th or 99th percentile. Second, at each timepoint, where pairwise conditional data is available (e.g., the experiment is done in two conditions, WT vs KO), Cyber-T ([8, 30]), a t-test with Bayesian regularization is used to perform differential analysis at that particular timepoint with multiple testing corrections. Finally, the statistics generated by the pipeline (e.g., circadian p-value, phases, differential p-values where applicable) are collected into a well-formatted table, to be incorporated into the downstream pipelines including the machine learning pipeline.

Given the output table of the statistical pipeline, we essentially obtain a matrix of circadian features for each molecular species in a given experiment. The `circadian_profiling` pipeline takes this feature matrix as input and perform a series of basic machine learning and visualization techniques.

The most commonly used visualization for such circadian data is a time series heatmap showing average concentrations at each time point, sorted by the predicted phase. This is demonstrated in the analysis of mouse ketogenic data ([49]) attached later in Chapter 4. A novel technique is also invented to utilize Dynamic Time Warping (DTW, [40]) to order time series. Briefly, a random time series with phase close to ZT0 [1] is used as a "bootstrap" time series and other time series are ordered by the DTW distance away from this series. This processing is useful for cases where the predicted phases do not separate distinct time series well, typically in experiments with high noise.

In addition, a collection of machine learning techniques are also available from the profiling pipeline. These include: dimension reduction (PCA, t-SNE [36]), clustering of timeseries signals or features derived from such signals via K-Means, Agglomerative Clustering (AC), and DBSCAN. The goal of these techniques is, in general, to identify for further analysis, groups of molecular species (e.g., genes) that behave similarly in their circadian oscillations. Implementation is done via scikit-learn in python.

Moreover, novel visualization techniques are also available, such as superimposed Andrews plot [4], phase-amplitude hexagonal correlation plot and radial phase distribution plot. Select examples of the aforementioned techniques are shown in the following figures. A published example of the radial phase distribution plot is shown in a diet study ([49]) attached at the end of Chapter 4.

[1] ZT indicates time point, thus ZT0 is time point 0 or hour 0

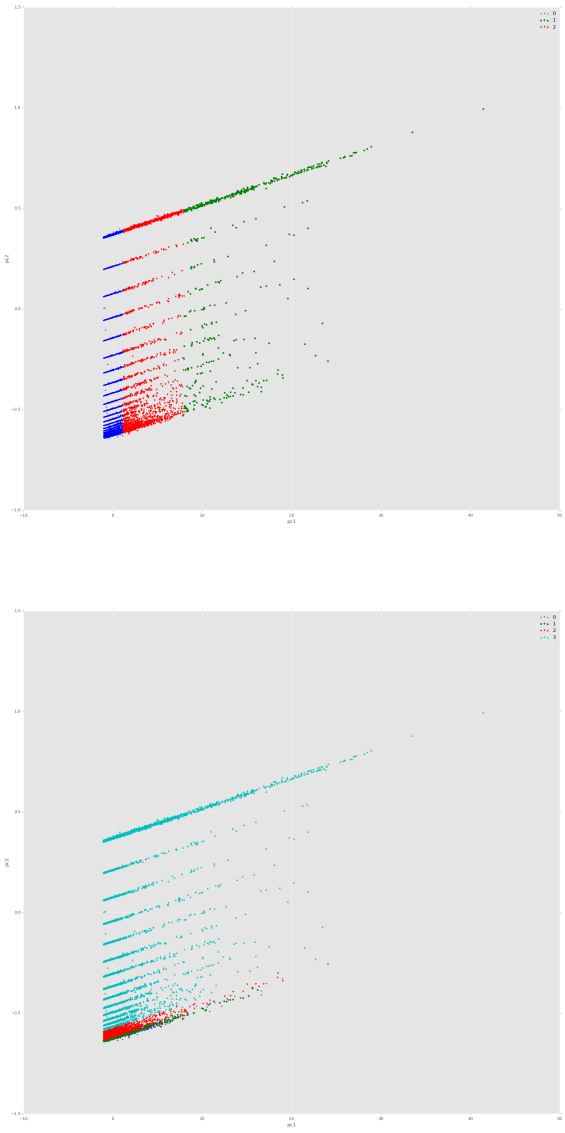


Figure 3.1: Dimension reduction (PCA) of circadian features and clustering by K-Means and Agglomerative Clustering. Data taken from a mouse liver diet experiment [16]. Features include phases, amplitudes and circadian p-values. Above shows K-Means while below shows AC. X, Y axis are principal components. Color indicates assigned clusters.

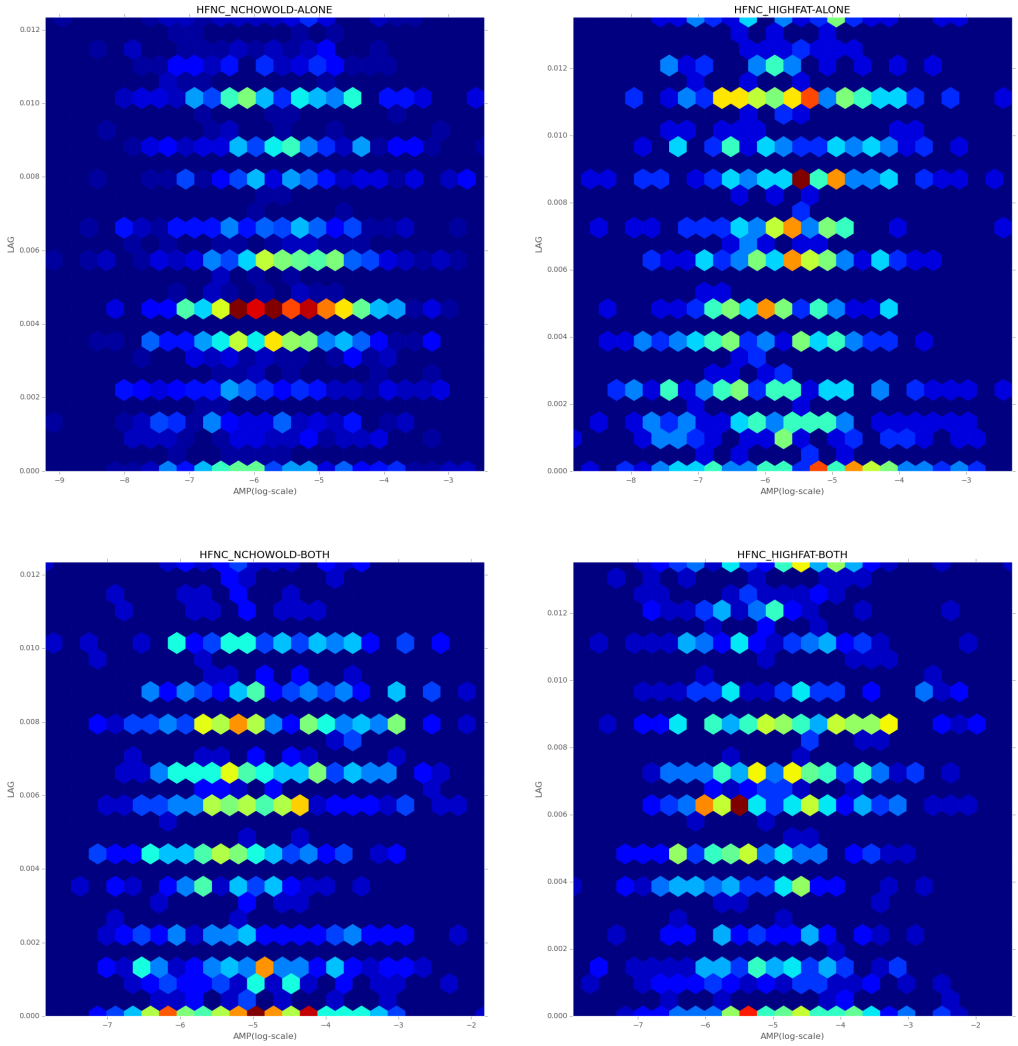


Figure 3.2: Hexagonal correlation plot of circadian data. Data taken from a mouse liver diet experiment [16]. LAG (y) indicates predicted phases while AMP (x) indicates predicted amplitudes. Color indicates frequency.

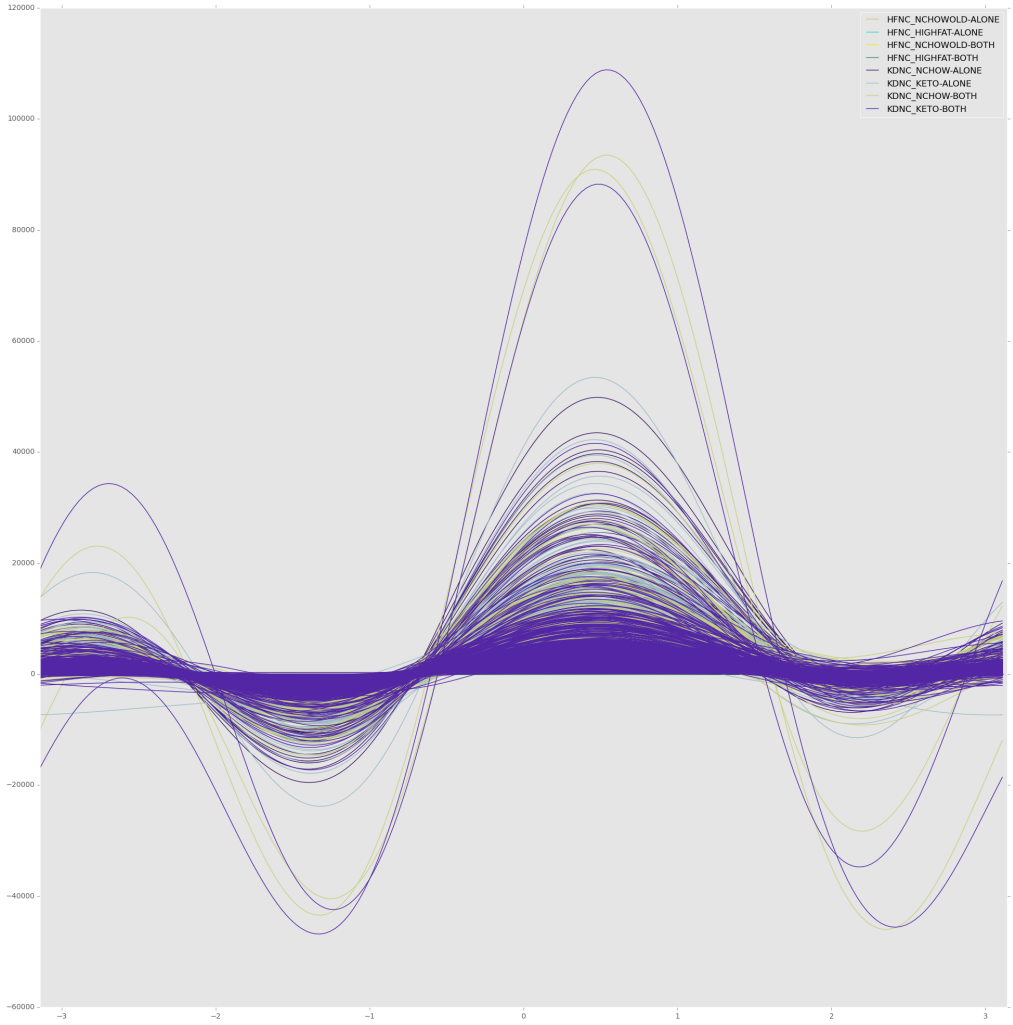


Figure 3.3: Superimposed Andrews Plot of circadian time series from several circadian groups. Data taken from a mouse liver diet experiment [16].

3.3 Integration into igb-pipelines

Through statistical analyses, we can convert high throughput circadian omic data from their raw, quantitative form, to a set of condensed, higher level features. These features include circadian p-values, phases, amplitudes, etc. Furthermore, through machine learning techniques or traditional threshold based filtering techniques, we can produce sets of circadian species that are of specific interest or significance. A very commonly used technique is to separate transcriptomic data into those that are considered significantly oscillating in one experimental condition vs. another condition. Similar sets can also be obtained by analyzing the phase and amplitude of the time series, through differential analysis at certain time points, or via dimension reduction of circadian features and clustering of such features.

Most of our projects rely on these outputs from the circadian statistical and machine learning pipelines. As such, these pipelines are fully integrated into the worg publication system (Chapter 5), CircadiOmics (Chapter 6) and PyCircadiOmics (Chapter 7).

3.4 Application: Mouse Gut Microbiome Paper

An example publication on the study of mouse microbiome, for which I am a contributing author, is attached below [41].

In particular, the aforementioned pipelines contributed to Figure 3 and the accompanying results section. Differential analysis was performed at fixed time points (ZT0 and ZT12) between control mice and mice that underwent microbial transplant from a high fat diet host. The analysis revealed a significant number of up-regulated and down-regulated genes due to the transplant. Furthermore, it revealed that at ZT12 a large number of Pparγ target genes are up-regulated. This finding is further strengthened by a metabolomic study using

similar differential analysis techniques. The full analysis and discovery is shown in the paper below.

Published online: July 14, 2016

Scientific Report

SOURCE DATA | TRANSPARENT PROCESS

EMBO reports

Gut microbiota directs PPARγ-driven reprogramming of the liver circadian clock by nutritional challenge

Mari Murakami[1,†], Paola Tognini[1,†], Yu Liu[2], Kristin L Eckel-Mahan[1,3], Pierre Baldi[2] & Paolo Sassone-Corsi[1,*]

Abstract

The liver circadian clock is reprogrammed by nutritional challenge through the rewiring of specific transcriptional pathways. As the gut microbiota is tightly connected to host metabolism, whose coordination is governed by the circadian clock, we explored whether gut microbes influence circadian homeostasis and how they distally control the peripheral clock in the liver. Using fecal transplant procedures we reveal that, in response to high-fat diet, the gut microbiota drives PPARγ-mediated activation of newly oscillatory transcriptional programs in the liver. Moreover, antibiotics treatment prevents PPARγ-driven transcription in the liver, underscoring the essential role of gut microbes in clock reprogramming and hepatic circadian homeostasis. Thus, a specific molecular signature characterizes the influence of the gut microbiome in the liver, leading to the transcriptional rewiring of hepatic metabolism.

Keywords circadian clock; liver; microbiota; PPARγ
Subject Categories Metabolism; Transcription; Microbiology, Virology & Host Pathogen Interaction
DOI 10.15252/embr.201642463 | Received 30 March 2016 | Revised 3 June 2016 | Accepted 22 June 2016

Introduction

Circadian rhythms are intimately linked to a large array of physiological processes, metabolic control, immune responses, hormonal regulations, and behavior [1,2]. The mammalian central pacemaker is localized in the hypothalamus, in a paired neuronal structure called the suprachiasmatic nucleus (SCN). The discovery that all tissues and virtually all cells contain an intrinsic circadian clock revolutionized the field, providing a conceptual framework toward the understanding of organismal homeostasis and physiological tissue-to-tissue communication [3–10]. It has been recently demonstrated that misalignment of the peripheral clocks with the SCN may contribute to a variety of pathological conditions [11]. Indeed, disruption of the circadian clock has been shown to lead to imbalance in metabolic homeostasis, contributing to a number of pathologies [12–15].

Accumulating evidence reveals that changes in nutritional regimes, including time-restricted feeding and challenge by high-fat diet (HFD), extensively influence liver circadian metabolism [16–18]. Importantly, the clock system undergoes a metabolic and transcriptional reprogramming in response to nutritional challenge, which involves the cyclic activation of otherwise non-circadian transcription factors. Specifically, oscillation of peroxisome proliferative activated receptor-γ (PPARγ) and its recruitment to chromatin drives a significant fraction of the HFD-induced liver clock reprogramming [19].

As dietary intake drastically affects the microbial community structure residing in the gut [20–23], we sought to explore whether gut microbes might be responsible for the reprogramming of hepatic circadian rhythmicity. Indeed, while it has been demonstrated that HFD-induced adiposity is transmissible via microbiota transfer [20,24] and that the gut microbiota contributes to circadian clock function [25,26], little is known about the molecular signatures through which gut microbes regulate clock function in a distal tissue. This question has unique relevance with respect to circadian biology as gut bacteria are considered to constitute an additional host organ, and it has been shown that they have important influences on host developmental and physiological processes [27–30]. Here, we show that a significant fraction of HFD-induced transcriptional reprogramming of the liver clock by PPARγ [19] is mediated by the gut microbial communities. This finding provides new insights into the connection between gut microbiome and the host's circadian metabolism. Indeed, we describe a specific molecular mechanism within the liver clock that interprets gut microbiota-driven signals to serve the interplay between the gut and hepatic metabolism.

1 Center for Epigenetics and Metabolism, Department of Biological Chemistry, U904 INSERM, University of California Irvine, Irvine, CA, USA
2 Institute for Genomics and Bioinformatics, School of Information and Computer Sciences, University of California Irvine, Irvine, CA, USA
3 Center for Metabolic and Degenerative Diseases, Institute of Molecular Medicine, University of Texas Health Sciences Center, Houston, TX, USA
*Corresponding author. Tel: +1 949 824 4540; E-mail: psc@uci.edu
†These authors contributed equally to this work

Results and Discussion

Microbial transfer mimics HFD-induced changes in liver metabolism

To determine whether the HFD-induced metabolic phenotype is transmissible through microbial transfer, we colonized control chow (CC)-fed recipient mice with microbial communities freshly harvested from donors fed CC or HFD (weighed 25.82 ± 0.36 g and 40.15 ± 1.55 g, respectively; for details see Materials and Methods). Consistent with previous reports [20,31], the gut microbiome derived from HFD-fed donors (HF-D) contained an increased fraction of *Firmicutes* and lower levels of *Bacteroidetes* as compared to CC-fed donors (CC-D) (Fig EV1A).

Ten days following the final fecal transplantation, HF-D feces recipient (HF-R) had significantly more epididymal (2.52 ± 0.08% versus 1.97 ± 0.10%) and mesenteric (1.32 ± 0.04% versus 1.19 ± 0.04%) fat pad mass than CC-D feces recipient (CC-R) (Fig 1A), whereas body weight and blood glucose levels were not affected (Fig EV1B and C). Also, an increase in lipid abundance in the liver was evident in HF-R compared to CC-R, which paralleled hepatic histology of HFD-fed animals (Fig 1B and C). This metabolic profile strongly suggests that transmission of the HFD-induced phenotype depends directly on microbiota transplantation. As expected, CC-R and HF-R mice displayed a significant difference in their microbiota composition (Fig 1D). There was an increase in the levels of *Actinobacteria* in HF-R compared with CC-R as well as in *Coriobacteriaceae*, which have been previously associated with obesity [32]. Moreover, HF-R microbiota was enriched in *Mollicutes*, in line with observations indicating that this class of bacteria is involved in diet-induced obesity in mice [24,33].

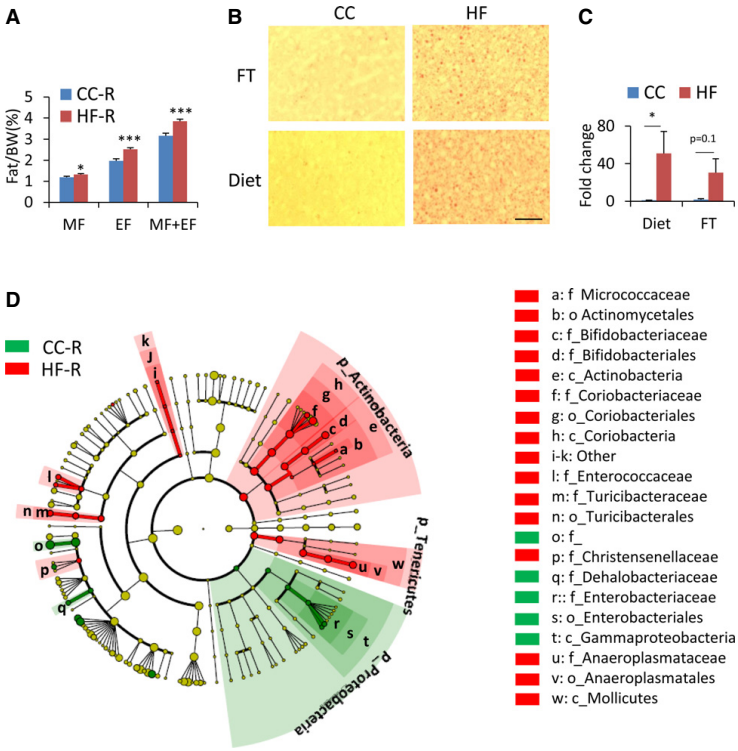


Figure 1. Metabolic phenotype induced by fecal transplantation.

A % fat mass/body weight of recipient mice (n = 34 per group, Student's t-test, $^*P < 0.05$, $^{***}P < 0.001$). Error bars represent SEM. MF: mesenteric fat, EF: epididymal fat. CC-R: control chow-feces recipients, HF-R: high-fat diet-feces recipients.

B Oil Red O staining of livers after 2 weeks of diet challenge (Diet) or fecal transplantation (FT). Fat vesicles were photographed at 100× magnification. CC: control chow, HF: high-fat diet. Scale bar, 0.1 mm.

C Fold change in lipid droplet area (n = 3–5 each group, diet: Mann–Whitney rank-sum test $^*P < 0.05$; fecal transplantation: unpaired, two-tailed Student's t-test $P = 0.1$). Error bars represent SEM.

D Cladogram generated from LEfSe analysis showing the differentially abundant microbiota from CC-R and HF-R.

Interestingly, in accordance with the notion that the gut microbiota elicits a profound effect on bile acid metabolism and on gut Farnesoid X receptor (FXR) signaling [34,35], we found that expression of *Fxr* and FXR target genes fatty acid binding protein-6 (*Fabp6*), organic solute transporter-α (*Osta*) and organic solute transporter-β (*Ostb*) to be significantly upregulated in the HF-R ileum (Fig EV1D), confirming the remodeling of the gut microbiome by fecal transplantation in recipient mice. As previous studies demonstrated that gut microbiota is able to promote an obesity phenotype through FXR [36], it is conceivable that phenotypic differences between HF-R and CC-R were at least in part attributable to FXR signaling alteration.

Rewiring of the hepatic circadian clock by HFD and HFD-derived gut microbiome

Circadian regulation plays an important role in liver functions, as bile acids, lipids, cholesterol, and glucose are all subject to diurnal control [37]. In order to investigate how gut microbial remodeling affects the circadian clock machinery in the liver, we examined hepatic gene expression in mice receiving fecal transplants from CC- or HFD-fed mice. Livers were collected at 4-h intervals along the circadian cycle, starting at 10 days following the final fecal transplantation (FT). It has been previously described that HFD, while inducing an extensive reprogramming of alternative transcription pathways, does not significantly influence the rhythmicity of the core clock genes, causing modest changes in the phase and amplitude of transcript oscillation [17–19]. Livers from CC-R and HF-R followed the same pattern, with clock genes such as *Bmal1, Per2, Rev-erba,* and the CLOCK:BMAL1-driven gene albumin D-box binding protein (*Dbp*) robustly rhythmic and slightly phase advanced in HF-R (Fig 2A). These profiles perfectly mirrored the ones of the corresponding transcripts in HFD-fed mice, as shown by the difference between HF and CC expression profiles (delta (HF-CC)) in each condition (diet and FT) (Fig 2B). This analysis underscores that the effect of HF fecal transplant on the liver clock parallels the one of HFD. Finally, BMAL1 phosphorylation, known to contribute to BMAL1-driven circadian activity [38,39], was virtually unaltered between CC-R and HF-R (Fig 2C), showing a mild phase advance in BMAL1 phosphorylation in HF-R compared to CC-R, in keeping with the phase advance of BMAL1

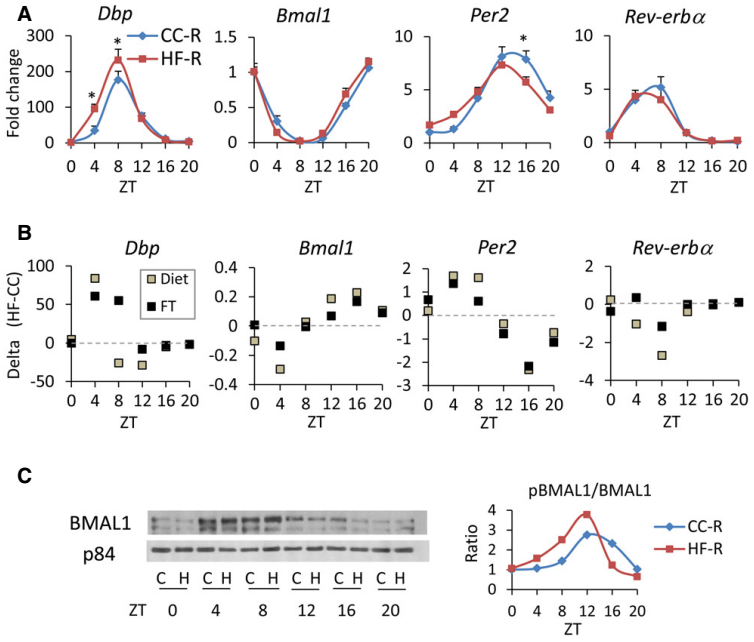


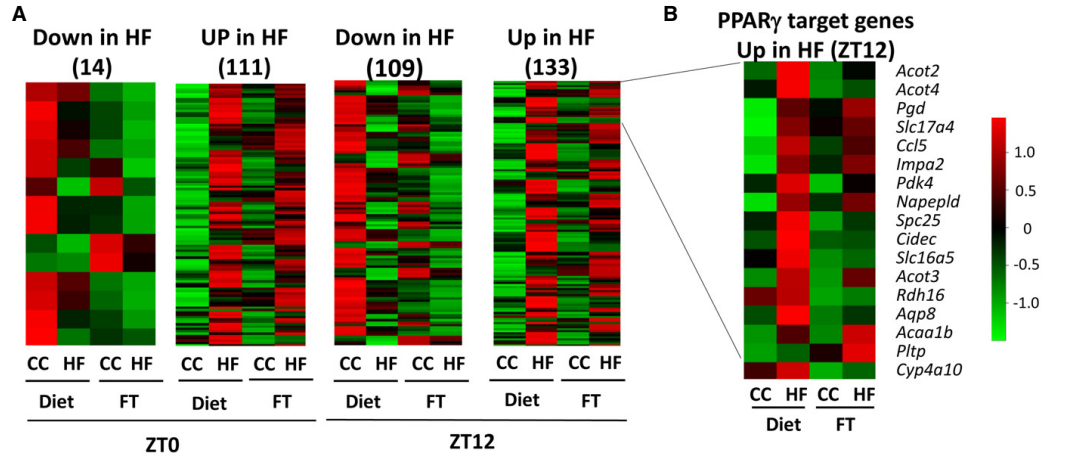
Figure 2. Core clock genes and BMAL1 protein expression.

A qPCR of core clock genes (n = 5–6 each group, two-way ANOVA, *post hoc* Holm–Sidak comparisons, *P < 0.05). Error bars represent SEM.

B Clock gene expression comparison between HFD-fed and HFD-R mice. Every square in the graphs represents the difference between HF and CC qPCR values (delta (HF-CC)) in each condition (Diet or FT) at each time point.

C Representative image of immunoblot analysis of BMAL1 in nuclear fraction throughout the circadian cycle. C: CC-R, H: HF-R. On the right, ratio between phosphorylated BMAL1 and non-phosphorylated BMAL1 is shown.

targets observed in HF-R mice (Fig 2A). Our data demonstrate that the liver core clock is only slightly modulated by HFD-dependent alteration in gut microbiota composition. On the other hand, metabolic changes triggered by the microbiota affect liver functions that in turn can influence clock-driven circadian physiology.



C

ZT12 Down in HF

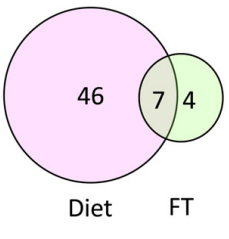
KEGG pathway	No. of genes
Protein processing in endoplasmic reticulum	6
RNA transport	5
Complement and coagulation cascades	5
Steroid hormone biosynthesis	3
Protein export	3
Alanine, aspartate and glutamate metabolism	2

D

ZT12 Up in HF

KEGG pathway	No. of genes
Biosynthesis of unsaturated fatty acids	6
Retinol metabolism	6
PPAR signaling pathway	5
Arachidonic acid metabolism	5
Drug metabolism - cytochrome P450	5
Fatty acid biosynthesis	2

E



palmitoleate (16:1n7)
cis-vaccenate (18:1n7)
eicosenoate (20:1n9 or 11)
glycerol
dihomo-linoleate (20:2n6)
oleate (18:1n9)
10-nonadecenoate (19:1n9)

Figure 3. PPARγ-driven transcriptional reprogramming induced by microbial transplantation.

A Heat maps representing the genes significantly downregulated or upregulated in HF feeding and HF fecal transplantation at ZT0 and ZT12 (n = 4, Cyber t-test $P < 0.05$).

B Heat map of PPARγ target genes upregulated in both HFD feeding and HF-R at ZT12 (n = 4, Cyber t-test $P < 0.05$).

C KEGG pathway analysis of genes downregulated at ZT12 in both HF feeding and HF-R.

D KEGG pathway analysis of genes upregulated at ZT12 in both HFD feeding and HF-R.

E Venn diagram and heat map of metabolites whose expression increased in both HFD feeding and HF-R at ZT12 (n = 5, Cyber t-test $P < 0.05$).

Source data are available online for this figure.

HFD and HFD-induced dysbiosis share a specific transcriptional program

To assess the impact of dysbiosis on the liver of fecal-transplanted animals, we performed transcriptome analysis by studying genes differentially expressed at ZT0 and ZT12 (the end and the beginning of the mouse active phase). At ZT0, 367 transcripts were downregulated and 871 were upregulated in HF-R compared to CC-R. On the other hand, 416 transcripts were downregulated and 2,071 were upregulated in HF-R compared to CC-R at ZT12. These results were then compared to our previous liver transcriptome data from 10-weeks HFD-fed animals [19] to identify common regulatory pathways. We performed pathway analyses both for genes shared by HF feeding and FT (Fig 3C and D) and for genes exclusively regulated by either HFD feeding or HF-FT with respect to their CC group (Fig EV2). Of all transcripts found to be upregulated in both HFD-fed mice and HF-R compared to CC-fed and CC-R respectively at ZT12, 133 transcripts were shared (Fig 3A, Source Data). Singular enrichment analysis of KEGG pathways of the overlapping genes revealed several categories related to lipid metabolism, such as "biosynthesis of unsaturated fatty acids" and "PPAR signaling pathway" (Fig 3D). On the other hand, 109 genes appeared in the group of transcripts that were downregulated in both HFD-fed animals and HF-R at ZT12 (Fig 3A). In contrast to "up in both high-fat groups" at ZT12, we found unique annotations including "protein processing in endoplasmic reticulum", "RNA transport", and "protein transport" (Fig 3C). At ZT0 (Fig 3A, Source Data), "steroid hormone biosynthesis pathway" was found "down in both high-fat groups" while no liver-associated annotations were enriched in "up in both high-fat groups" (see Table EV1 for details). The comparative analyses between high-fat feeding and HF-R transcriptome data revealed that lipid metabolism is the principal biological process, which is influenced by both high-fat feeding and HFD-induced dysbiosis. Moreover, the categories related to lipid metabolism present in our gene ontology analysis suggest ZT12 as the most relevant ZT involved in the transcriptional modifications induced by both HFD and HF-R conditions.

Gut microbial remodeling induces circadian PPARγ recruitment in HF-R

A prominent pathway through which high-fat feeding induces remodeling of the liver clock is the *de novo* cyclic activation of PPARγ [19]. PPARγ is abundantly expressed in adipose tissue and it is less abundant in the liver under physiological conditions, however it is induced in hepatic steatosis or in obesity [40,41]. Based on the signatures of transcriptome analyses, we speculated that PPARγ-driven reprogramming induced by high-fat feeding might be mediated by microbial alteration. Thus, to address this issue, we crossed 133 genes "up in both high-fat groups" with previously established PPARγ-ChIP Seq analysis data [42,43]. Notably, we found that 17 out of the 133 (12.8%; $P = 0.006$) genes, including DFFA-like effector c (*Cidec*), acyl-CoA thioesterase 2 (*Acot2*), and phospholipid transfer protein (*Pltp*) (Fig 3B), were PPARγ targets. Notably, comparative analysis of the liver metabolome at ZT12 (Source Data) revealed that 6 out of 7 metabolites, which were significantly increased in both HFD-fed animals and HF-R at ZT12, were long-chain fatty acids, and some of them, such as palmitoleate and oleate, are possible ligands of PPARγ [44,45] (Figs 3E and EV3C). Indeed, gene expression of lipogenic enzymes catalyzing the synthesis of these long-chain fatty acids such as acetyl-CoA carboxylase 1 (*Acc1*), fatty acid elongase 6 (*Elovl6*), and stearoyl-CoA desaturase 1 (*Scd1*) and its transcription factor SREBP1 were significantly upregulated in HF-R compared to CC-R (Fig EV3A–C). Furthermore, it is reported that PPARγ enhances lipogenic gene expression in the liver and induces hepatic lipid accumulation [40,41]. Taken together, our comprehensive analyses confirm that microbial remodeling induces ZT-dependent PPARγ pathway activation at a global transcriptional level in the liver accompanied by hepatic metabolites alteration.

Based on these results, we next analyzed the PPARγ transcriptional pathway along the circadian cycle. Gene expression of *Pparg* and its target genes *Cidec*, *Pltp*, *Acot2* and pyruvate carboxylase (*Pcx*) all displayed rhythmic profiles, which peaked at ZT12 in the HF-R mice (Fig 4A). The expression of *Cidec* and *Pcx* genes has been previously demonstrated to become cyclic upon high-fat feeding [19]. Interestingly, *Cidec* is a lipid-binding protein induced in the steatotic liver and contributes to lipid accumulation [46,47]. Also, the expression of both *Acot2*, the enzyme that hydrolyses long-chain fatty acyl-CoA and facilitates fatty acid oxidation [48], and *Pltp*, a lipid transfer protein that drives lipoprotein production [49], showed rhythmic expression with a zenith at ZT12 in HFD-fed animals, in contrast to the expression of these genes in the CC-fed animals (Fig EV3D). Again, *Pparg* and its target genes oscillated with striking amplitude in HF-R as compared to CC-R, peaking at ZT12 (Fig 4A). Similar results were confirmed in germ-free (GF) mice transplanted with feces from HF-D or CC-D (Fig EV3E).

Figure 4. Microbial transplantation induces PPARγ pathway activation in HF-R mice liver.

A PPARγ target genes expression measured by qPCR (n = 5–12 each group, two-way ANOVA, *post hoc* Holm–Sidak comparisons, $*P < 0.05$, $**P < 0.01$). Error bars represent SEM.

B Immunoblot analysis of PPARγ in nuclear fraction. On the right, quantification of PPARγ normalized to p84 is shown. Average of donor mice (ref. [19]) and four experimental replicates per time point, per group of recipient mice (two-way ANOVA, *post hoc* Holm–Sidak comparisons, $*P < 0.05$). Error bars represent SEM.

C Immunoblot analysis of PPARγ in chromatin fraction. On the right, quantification of PPARγ normalized to the TBP is shown. Average of 4 experimental replicates, per time point, per group (two-way ANOVA, no significant difference). Error bars represent SEM.

D Chromatin recruitment of PPARγ at PPAR binding site contained in *Cidec* and *Pcx* promoters (n = 3–5 per group, per ZT, two-way ANOVA, *post hoc* Holm–Sidak comparisons, $*P < 0.05$). H3K4me3 abundance at *Cidec* and *Pcx* promoters (n = 4–7, per group, per ZT, unpaired, two-tailed Student's t-test, $*P < 0.05$). Error bars represent SEM.

E *Pparg* and *Cidec* gene expression at ZT12 in the liver of fecal-transplanted animals treated with GW9662 or vehicle (n = 6–9 per group, two-way ANOVA, *post hoc* Holm–Sidak comparisons, $*P < 0.05$). Error bars represent SEM.

F % epididymal fat mass/body weight of fecal-transplanted animals treated with GW9662 or vehicle (n = 6–9 per group, unpaired, two-tailed Student's t-test, $*P < 0.05$). Error bars represent SEM.

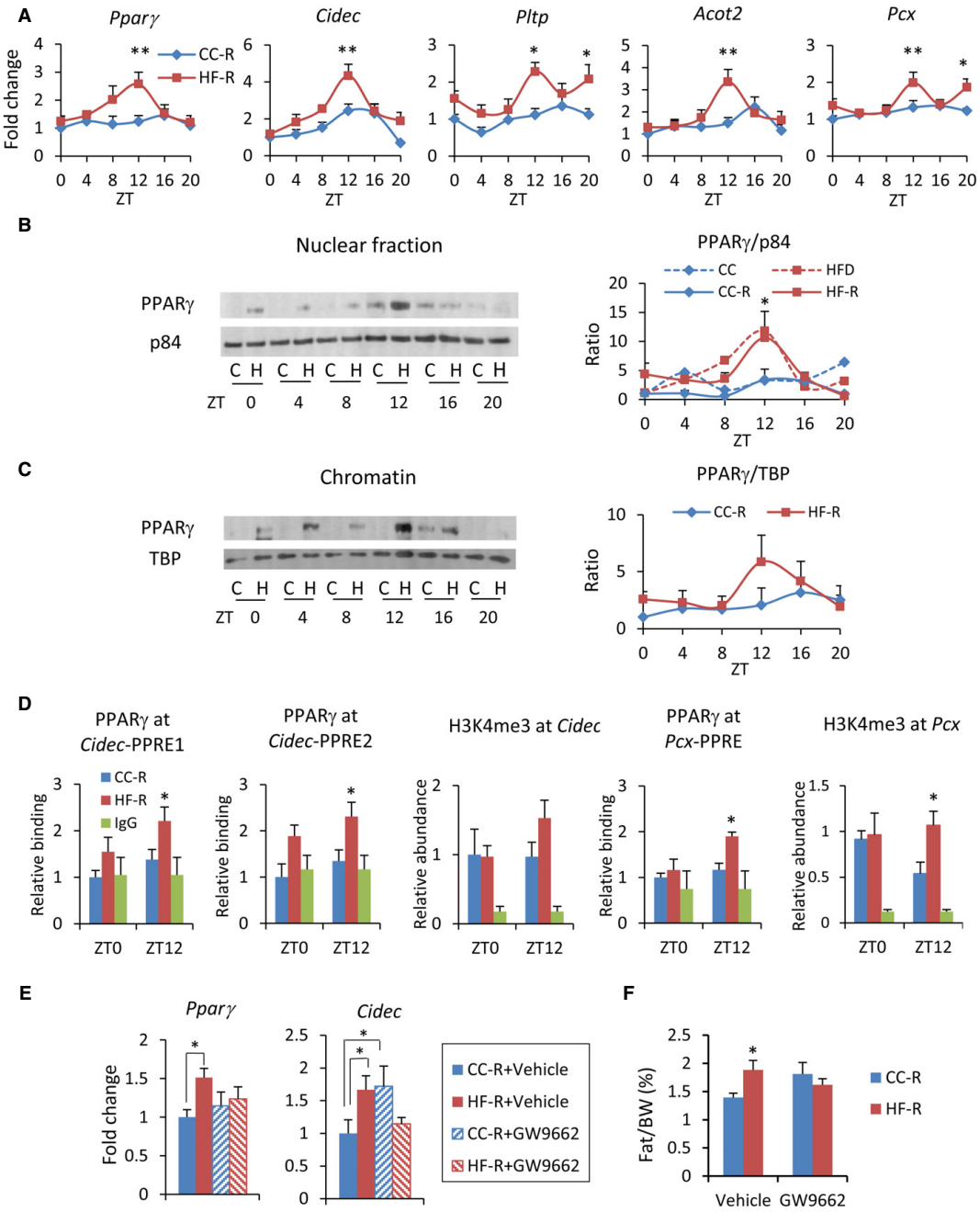


Figure 4.

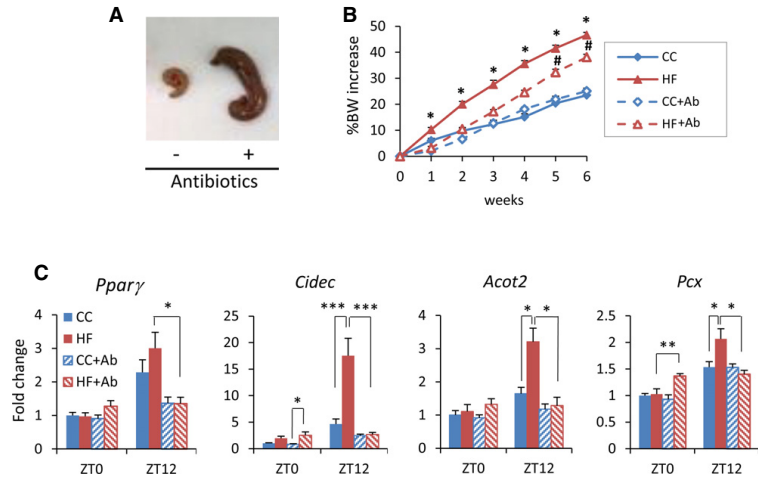


Figure 5. HFD-induced reprogramming is counteracted by antibiotic treatment.

A Cecum of mice treated with or without antibiotics for 6 weeks.

B Body weight of animals fed CC or HFD and treated with or without antibiotics for 6 weeks. Animals were fed CC or HFD during the experiment (n = 8 HF and CC control, n = 18 HF and CC + antibiotic; HFD control versus HF antibiotic: Student's t-test, $^{*}P < 0.05$; HF antibiotic versus CC antibiotic: unpaired, two-tailed Student's t-test, $^{\#}P < 0.05$). Error bars represent SEM.

C Gene expression of *Pparg* and PPARγ target genes in the liver of mice fed CC or HFD for 6 weeks and simultaneously treated with an antibiotic cocktail (n = 4 for each antibiotic-untreated group, n = 9 for each antibiotic-treated group, per ZT; ZT0 and ZT12 two-way ANOVA, *post hoc* Holm–Sidak comparisons, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$). Error bars represent SEM.

Importantly, nuclear- and chromatin-bound PPARγ (Fig 4B and C) drastically oscillated in HF-R with a zenith at ZT12. This profile is remarkably reminiscent of the previously reported effect of HFD on liver PPARγ nuclear protein [19] (Fig 4B). Moreover, PPARγ recruitment to PPARγ-binding sites (PPRE) in *Cidec* and *Pcx* promoters was significantly increased at ZT12, in line with gene and protein expression profiles (Fig 4D) and paralleled by a tendency to increase in trimethylation of Lys4 on the histone H3 (H3K4Me3) at the same genomic regions (Fig 4D). Finally, to verify that PPARγ activation is one of the central molecular mechanisms underlying HF-R liver reprogramming, we treated fecal-transplanted mice with the selective PPARγ antagonist GW9662 [50]. As expected, HF-R injected with the vehicle (+ vehicle) displayed an increase in the epididymal fat with respect to CC-R + vehicle. On the contrary, HF-R injected with GW9662 (+ GW9662) showed no difference with respect to CC-R mice + vehicle (Fig 4F). Strikingly, the increase in *Pparg* and *Cidec* liver gene expression at ZT12 was completely prevented in HF-R + GW9662 (Fig 4E). Thus, a significant fraction of the HF microbiota-dependent liver diurnal reprogramming relies on the induction and activation of the transcription factor PPARγ.

HFD-induced reprogramming is neutralized by antibiotic treatment

To validate that gut microbiota contributes to the HFD-induced reprogramming, we treated HFD-fed mice with an antibiotic cocktail for 6 weeks [51]. Following antibiotic treatment, a remarkable enlargement of the cecum was observed (Fig 5A), a distinctive feature of germ-free mice. The body weight of antibiotic-treated animals on HFD showed a significantly smaller increase with respect to HFD-fed control mice, but it was still significantly higher than that of CC-fed control animals. On the contrary, there was no difference in body weight between CC-fed animals treated or not treated with antibiotics (Fig 5B). Moreover, the high levels of serum glucose induced by high-fat feeding were significantly decreased to control levels by antibiotic treatment (Fig EV4). The expression of *Pparg* and its targets were significantly higher in HFD-fed control mice compared to CC-fed control mice at ZT12, whereas there was no difference between the groups at ZT0. Strikingly, antibiotic treatment reversed the effect of HFD on *Pparg* and its target genes, significantly decreasing their expression at ZT12, the peak time point of *Pparg* in HFD-fed animals (Fig 5C). It is worth noting that HFD-induced reprogramming is dissociable from obesity [19]. In fact, the expression of PPARγ target genes was not different between antibiotic-treated HFD-fed and CC-fed animals even though the former weighed significantly more than the latter (Fig 5B). Thus, the reversible effects shown in the antibiotic-treated HFD-fed animals were induced by the depletion of microbiota, rather than the loss of body weight. These results confirm that the microbial alteration plays an important role in PPARγ-driven diurnal transcriptional remodeling in HFD-fed animals.

Accumulating evidence reveals that the gut microbiota is involved in maintaining the host metabolic homeostasis. Since metabolic networks and the circadian clock are intimately intertwined, the impact of the gut microbiota in regulating host peripheral clocks is significant [25,52]. Evidence is emerging that feeding

rhythms induce diurnal fluctuations of gut microbiome [53–55] while disruption of the rhythmic microbial population in HFD-fed animals leads to an alteration of hepatic circadian transcripts [25]. However, the molecular mechanisms through which gut microbes regulate liver homeostasis by rewiring the circadian transcriptome have been unexplored.

Nuclear receptors play crucial roles in connecting the circadian network and metabolism [56] by sensing oscillatory metabolites and integrating them into host energy balance [57,58]. Among them, PPARγ, a master player in adipogenesis and a direct regulator of genes involved in lipid and glucose metabolism [59], is at the heart of this network [60–62]. Notably, high-fat feeding was shown to amplify the expression of *Pparg* [60] and its target genes by inducing *de novo* cyclic recruitment of PPARγ to chromatin [19]. Our results show that gut microbial remodeling under high-fat feeding induces rhythmic activation of PPARγ that in turn leads to transcriptional reprogramming in the liver. Circadian activation of PPARγ and SREBP1 expression coordinately contributes to the regulation of hepatic lipid metabolism in HF-R mice and previous evidence demonstrates that metabolites produced by gut microbes regulate host liver lipogenesis [63–65]. Indeed, we observed an increase in the levels of hepatic long-chain fatty acids that are involved in both signaling pathways and in lipid accumulation in the liver. Furthermore, short-chain fatty acids (SCFA) produced by bacterial fermentation are modulators of PPARγ [66], and PPARγ signaling might be altered by different SCFA profiles depending on dietary changes. Thus, we speculate that gut microbe-derived metabolites modulate PPARγ activity in HF-R liver either through direct activation or indirectly via SREBP1. Interestingly, HFD-driven dysbiosis is implicated in hematopoietic stem cell differentiation through PPARγ activation [67], further supporting a PPARγ-mediated effect of gut microbiota on distal tissues.

The circadian clock is fundamental to maintain liver tissue physiology and contributes to whole-body metabolic homeostasis. Our results demonstrate that HFD-induced gut dysbiosis affected hepatic diurnal rhythmicity and induced hepatosteatosis and an increase in fat depots. GW9662 treatment, blocking the microbiota-driven PPARγ diurnal reprogramming, reverted the fat depot phenotype in HF-R mice. The change in circadian gene expression subsequent to HFD microbiome remodeling might exacerbate more severe metabolic diseases.

In conclusion, we have demonstrated that the HFD-induced remodeling in gut microbiome mediates PPARγ-driven reprogramming of the host liver clock, leading to specific rewiring of circadian transcription. This effect was reversible by depletion of gut microbiota, confirming the importance of intestinal microbes' contribution to cellular plasticity in response to nutritional stress. The elucidation of the molecular mechanisms underlying the powerful effect of diet-induced gut microbial alteration on the host liver circadian clock paves the way for pharmaceutical strategies to target metabolic diseases.

Materials and Methods

Animals

Age-matched male C57BL/6J mice (Jackson Laboratory) were maintained on a 12-h light/12-h dark cycle with *ad libitum* access of food and water. Animals' care and use was in accordance with guidelines of the Institutional Animal Care and Use Committee at the University of California, Irvine. Male C57BL/6J germ-free mice were kindly provided by Dr. Sarkis Mazmanian (California Institute of Technology).

Microbiota transplantation experiments

Six-week-old donor mice were fed CC or HFD for 6–11 weeks before their feces were harvested. Recipient mice were fed CC from the age of 6 weeks old, and they were pretreated with a combination of vancomycin 0.5 g/l and neomycin 1 g/l in drinking water for 1 week followed by intraperitoneal injection of clindamycin at 10 mg/kg body weight for two consecutive days before fecal transplantation. At the time of fecal transplantation, freshly harvested donor feces were suspended in PBS and mixed with a vortex for 10 min. Recipients at 12 weeks of age were colonized with microbiota by gavaging freshly prepared donor fecal suspension (0.8–1.0 mg feces/g BW/day) for 5 days, once a day. All recipients were kept on CC throughout the experimental period and sacrificed 10 days after the final transplantation. For fecal transplantation into germ-free mice, 8-week-old male germ-free mice were colonized only once. Recipient mice were sacrificed 14 days after fecal transplantation.

Microarray analysis

RNA was extracted using TRIzol reagent (GIBCO BRL Life Technologies) and cleaned up with RNeasy spin columns (Qiagen). Quality of purified RNA was checked with an Agilent Bioanalyzer (Agilent Technologies). Microarrays were performed at the UCI Genomics High-throughput Facility, University of California, Irvine, as described in a previous study [19]. Data sets can be found in the NCBI gene expression omnibus (GEO), GSE 82250 for FT and GSE 52333 for Diet experiment.

GW9662 treatment

GW9662 (Cayman Chemical) dissolved in DMSO was diluted in PBS and intraperitoneally injected to fecal-transplanted mice (CC-R and HF-R) at a dose of 4 mg/kg at ZT12 for 15 days, once a day, from the second day of microbial transplantation. Animals were sacrificed at ZT12, and liver tissue was harvested for gene expression 24 h after the final injection.

Antibiotic treatment

Six-week-old mice were given vancomycin 0.5 g/l, ampicillin 1 g/l, neomycin 1 g/l, and metronidazole 1 g/l in drinking water *ad libitum* following previously published protocol to deplete commensals [51] and they were fed CC or HFD during this period. Mice were sacrificed after 6 weeks of antibiotic treatment.

Diets

Mice were fed a control chow (CC) (Harlan Laboratories, TD. 110870) or HFD (60% kcal fat, Research Diets, D12492) with *ad libitum* access to both diets. 2020X (Harlan Laboratories) was

used as a CC for antibiotics treatment experiment. All diets used in this study were irradiated or autoclaved to avoid contamination.

Histological analysis

Freshly isolated liver was embedded in OCT, immediately snap frozen in liquid nitrogen and stored in −80°C until they were used for the analysis. Sections were sliced into 10 μm and stained with Oil Red O. Area of lipid droplets was calculated by ImageJ.

16S rRNA gene sequencing analysis

Feces collected from donors and recipients were immediately frozen in liquid nitrogen. Fecal DNA was extracted with QIAamp Fast DNA Stool Mini kit (Qiagen) according to the manufacturer's protocol. 16S rRNA gene sequencing was performed by Zymo Research (Irvine, CA, USA). DNA samples were quantified, followed by PCR amplification of V3 and V4 regions of the 16S rRNA gene. A second PCR was performed to add barcodes to each sample. The final amplicon libraries were quantified and then sequenced on MiSeq. Paired-end raw reads in 2 × 250 bp length went through quality check steps including adaptor removal and trimming 3′ bases with quality score $q < 20$. Reads shorter than 20 bp were discarded. The data were analyzed using the program QIIME release 1.8.0 (www.qiime.org) and the latest Greengenes 16S rRNA gene database (gg_13_8).

Quantitative RT–PCR

Liver and intestine samples were homogenized in TRIzol lysis reagent. Chloroform was added and the samples were shaken for 15 s. The samples were left at RT for 3 min and then centrifuged (12,000 *g*, 15 min, 4°C). The upper phase aqueous solution, containing RNA, was collected in a fresh tube, and the RNA was precipitated by adding isopropanol. Samples were mixed by vortexing, left at RT for 10 min and then centrifuged (12,000 *g*, 10 min, 4°C). Supernatant was discarded and the RNA pellet was washed in 75% ethanol by centrifugation (7,500 *g*, 5 min, 4°C). Supernatant was discarded and the pellet was resuspended in DEPC water. Total RNA concentrations were determined by Nanodrop spectrophotometer (ThermoScientific). Total RNA was reverse transcribed using iScript Reverse Transcription Supermix (Bio-Rad). Gene expression was analyzed by Real-Time PCR (Bio-Rad) using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad). See Table EV2 for the primer sequence list.

Nuclear and chromatin fraction extraction

Approximately 250 mg of liver was homogenized in 4 ml buffer A (10 mM HEPES, pH 7.8, 25 mM KCl, 0.5 mM spermidine, 1 mM EGTA, 1 mM EDTA, 0.32 M sucrose, and 0.3% Triton) with protease inhibitors. Samples were centrifuged at 1,000 *g* for 10 min at 4°C. Pellets were resuspended in 4 ml buffer A and centrifuged again at 1,000 *g* for 10 min at 4°C. Pellets were then resuspended in 4 ml low-salt buffer (10 mM HEPES, pH 7.8, 25 mM KCl, 0.5 mM spermidine, 1 mM EGTA, 1 mM EDTA, and 20% glycerol) and then centrifuged again at 1,000 *g* for 10 min at 4°C. Nuclear pellets were resuspended in 1 ml low-salt buffer, centrifuged, and resuspended in 1× volume low-salt buffer and 2× high-salt buffer (10 mM HEPES, pH 7.8, 25 mM KCl, 0.5 mM spermidine, 1 mM EGTA, 1 mM EDTA, 20% glycerol, and 0.5 M KCl). Suspensions were nutated for 1 h at 4°C and then centrifuged at 12,000 *g* for 20 min. The resulting supernatant is used as the soluble nuclear fraction. The pellet was resuspended in modified RIPA buffer, sonicated, centrifuged at 12,000 *g* for 20 min at 4°C, and used as chromatin fraction for subsequent immunoblot analysis.

Immunoblot analysis

Approximately 3–10 μg of nuclear or chromatin extract from liver samples was loaded on 8% polyacrylamide gels. The blots were blocked in 5% milk in TBST and incubated overnight at 4°C with a specific primary antibody. BMAL1 antibody (abcam # 93806) was diluted 1:2,000, PPARγ (1+2) antibody (abcam # 41928) 1:1,000, SREBP1 antibody (Santa Cruz # 13551) 1:100, and p84 antibody (GeneTex # 70220) 1:3,000. Blots were then washed three times in TBST for 20 min, incubated in HRP-conjugated anti-mouse or anti-rabbit diluted (1:8,000) in 2.5% milk in TBST 1 h at RT. The membranes were then rinsed three times in TBST and incubated in enhanced chemiluminescent substrate and exposed to films. Films were scanned and densitometry was analyzed through ImageJ software.

Chromatin immunoprecipitation

Approximately 100 mg of liver tissue was minced and double crosslinked with DSG for 40 min and 1% formaldehyde for 10 min followed by adding glycine (0.125 M final concentration) at room temperature for 10 min. After homogenizing tissue pellet in PBS, 1 ml of lysis buffer was added. Samples were sonicated by Bioruptor to generate 200–500-bp fragments and centrifuged at 10,000 *g* at 4°C. Supernatant was diluted in dilution buffer (1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris–HCl, 167 mM NaCl), precleared with protein-G beads blocked with salmon sperm DNA and BSA for 2 h. Precleared supernatant was incubated with 2 μg of PPARγ (1+2) antibody (abcam # 41928) or 3 μg of histone H3K4me3 (Active Motif #39159) primary antibodies overnight at 4°C. Protein-G beads were added to the supernatant and incubated for 2 h at 4°C and centrifuged. Beads were recovered and washed in low-salt buffer, high-salt buffer, and LiCl buffer, followed by washing in TE for three times. Elution buffer (300 mM NaCl, 0.5% SDS, 10 mM Tris–HCl, 5 mM EDTA) was added to the washed beads and treated with RNase at 37°C for 2 h and proteinase K at 65°C overnight. Equal amount of PCI was added to the samples and the aqueous phase was recovered. DNA was precipitated by adding 100% ethanol, NaOAc, and glycogen and kept at −80°C overnight. Samples were centrifuged at 12,000 *g* for 30 min at 4°C and washed with 70% ethanol followed by centrifugation at 12,000 *g* for 30 min at 4°C. Quantitative PCR was done using SYBRGreen reagent.

Metabolite analysis

Liver metabolite analysis was performed by Metabolon, Inc. (Durham, NC, USA) based on ultrahigh-performance liquid

chromatography–tandem mass spectroscopy (UPLC-MS/MS) and gas chromatography–mass spectroscopy (GC-MS).

Statistical analysis

Data are expressed as mean ± SEM. The significance of differences was analyzed by Student's *t*-test or ANOVA and *post hoc* analysis for multiple group comparison. When the data were not distributed normally, we used Mann–Whitney rank-sum test. For microarray data, at each time point of interest, CyberT, a differential analysis program using a Bayesian-regularized *t*-test [68,69] was performed between experimental group and control group (i.e. HFD versus CC, HF-R versus CC-R) and *P*-values were calculated. Transcripts/metabolites passing P-value < 0.05 were considered differential at that time point. Sets of differential transcripts/metabolites from diet or fecal transplantation data were further intersected to produce the list of commonly differential transcripts/metabolites. To cross our data with ChIP-Seq databases, ChIP-Seq peaks were extracted from previously published data [42,43]. Genes with peaks that are close to transcription start sites (from 10,000 bp upstream to 2,000 bp downstream) from any of the data sets were considered potential targets.

Gene annotation analysis

Gene annotation was performed using Genecodis [70]. Data presented in the manuscript was obtained using KEGG as knowledge base and pathways were ranked by number of genes (hits) with pathway annotations.

Expanded View for this article is available online.

Acknowledgements
We thank members of the Sassone-Corsi laboratory for constructive comments and Sherry Dilag-Penilla, Manuela Raffatellu, Stefan Jellbauer, Fei Teng, Christina Klinger, Sara McBride, Taren Thron, and Sarkis Mazmanian. We thank Melanie Oakes, Seung-Ah Chung, and Valentina Ciobanu at the UCI Genomics High-Throughput Facility. This work was supported by NIH grant DA036408 and the Institut National de la Sante et de la Recherche Medicale (INSERM, France) (P.S-C.) and the Mishima Kaiun Memorial Foundation (M.M.). M.M was supported by Sumitomo Life Welfare and Culture Foundation; P. T. is supported by EMBO ALTF 1510-2012 and Human Frontier Science Program LT 000576/2013.

Author contributions
MM and PT performed and analyzed all the experiments. YL and PB performed bioinformatics analysis. KE-M provided HFD samples and data. MM, PT, and PS-C designed the experiments and wrote the manuscript.

Conflict of interest
The authors declare that they have no conflict of interest.

References

1. Bass J, Takahashi JS (2010) Circadian integration of metabolism and energetics. *Science* 330: 1349 – 1354
2. Eckel-Mahan K, Sassone-Corsi P (2013) Metabolism and the circadian clock converge. *Physiol Rev* 93: 107 – 135
3. Balsalobre A, Damiola F, Schibler U (1998) A serum shock induces circadian gene expression in mammalian tissue culture cells. *Cell* 93: 929 – 937
4. Giebultowicz JM, Stanewsky R, Hall JC, Hege DM (2000) Transplanted *Drosophila* excretory tubules maintain circadian clock cycling out of phase with the host. *Curr Biol* 10: 107 – 110
5. Krishnan B, Dryer SE, Hardin PE (1999) Circadian rhythms in olfactory responses of *Drosophila melanogaster*. *Nature* 400: 375 – 378
6. Schibler U, Sassone-Corsi P (2002) A web of circadian pacemakers. *Cell* 111: 919 – 922
7. Stokkan KA, Yamazaki S, Tei H, Sakaki Y, Menaker M (2001) Entrainment of the circadian clock in the liver by feeding. *Science* 291: 490 – 493
8. Whitmore D, Foulkes NS, Strahle U, Sassone-Corsi P (1998) Zebrafish clock rhythmic expression reveals independent peripheral circadian oscillators. *Nat Neurosci* 1: 701 – 707
9. Crane BR, Young MW (2014) Interactive features of proteins composing eukaryotic circadian clocks. *Annu Rev Biochem* 83: 191 – 219
10. Dunlap JC (1999) Molecular bases for circadian clocks. *Cell* 96: 271 – 290
11. Mukherji A, Kobiita A, Damara M, Misra N, Meziane H, Champy MF, Chambon P (2015) Shifting eating to the circadian rest phase misaligns the peripheral clocks with the master SCN clock and leads to a metabolic syndrome. *Proc Natl Acad Sci USA* 112: E6691 – E6698
12. Marcheva B, Ramsey KM, Buhr ED, Kobayashi Y, Su H, Ko CH, Ivanova G, Omura C, Mo S, Vitaterna MH *et al* (2010) Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes. *Nature* 466: 627 – 631
13. Paschos GK, Ibrahim S, Song WL, Kunieda T, Grant G, Reyes TM, Bradfield CA, Vaughan CH, Eiden M, Masoodi M *et al* (2012) Obesity in mice with adipocyte-specific deletion of clock component Arntl. *Nat Med* 18: 1768 – 1777
14. Turek FW, Joshu C, Kohsaka A, Lin E, Ivanova G, McDearmon E, Laposky A, Losee-Olson S, Easton A, Jensen DR *et al* (2005) Obesity and metabolic syndrome in circadian Clock mutant mice. *Science* 308: 1043 – 1045
15. Sahar S, Sassone-Corsi P (2009) Metabolism and cancer: the circadian clock connection. *Nat Rev Cancer* 9: 886 – 896
16. Asher G, Sassone-Corsi P (2015) Time for food: the intimate interplay between nutrition, metabolism, and the circadian clock. *Cell* 161: 84 – 92
17. Hatori M, Vollmers C, Zarrinpar A, DiTacchio L, Bushong EA, Gill S, Leblanc M, Chaix A, Joens M, Fitzpatrick JA *et al* (2012) Time-restricted feeding without reducing caloric intake prevents metabolic diseases in mice fed a high-fat diet. *Cell Metab* 15: 848 – 860
18. Kohsaka A, Laposky AD, Ramsey KM, Estrada C, Joshu C, Kobayashi Y, Turek FW, Bass J (2007) High-fat diet disrupts behavioral and molecular circadian rhythms in mice. *Cell Metab* 6: 414 – 421
19. Eckel-Mahan KL, Patel VR, de Mateo S, Orozco-Solis R, Ceglia NJ, Sahar S, Dilag-Penilla SA, Dyar KA, Baldi P, Sassone-Corsi P (2013) Reprogramming of the circadian clock by nutritional challenge. *Cell* 155: 1464 – 1478
20. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI (2009) The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med* 1: 6ra14
21. Carmody RN, Gerber GK, Luevano JM Jr, Gatti DM, Somes L, Svenson KL, Turnbaugh PJ (2015) Diet dominates host genotype in shaping the murine gut microbiota. *Cell Host Microbe* 17: 72 – 84

22. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA *et al* (2014) Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505: 559 – 563
23. Wu M, McNulty NP, Rodionov DA, Khoroshkin MS, Griffin NW, Cheng J, Latreille P, Kerstetter RA, Terrapon N, Henrissat B *et al* (2015) Genetic determinants of *in vivo* fitness and diet responsiveness in multiple human gut Bacteroides. *Science* 350: aac5992
24. Turnbaugh PJ, Backhed F, Fulton L, Gordon JI (2008) Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* 3: 213 – 223
25. Leone V, Gibbons SM, Martinez K, Hutchison AL, Huang EY, Cham CM, Pierre JF, Heneghan AF, Nadimpalli A, Hubert N *et al* (2015) Effects of diurnal variation of gut microbes and high-fat feeding on host circadian clock function and metabolism. *Cell Host Microbe* 17: 681 – 689
26. Montagner A, Korecka A, Polizzi A, Lippi Y, Blum Y, Canlet C, Tremblay-Franco M, Gautier-Stein A, Burcelin R, Yen YC *et al* (2016) Hepatic circadian clock oscillators and nuclear receptors integrate microbiome-derived signals. *Sci Rep* 6: 20127
27. Garidou L, Pomie C, Klopp P, Waget A, Charpentier J, Aloulou M, Giry A, Serino M, Stenman L, Lahtinen S *et al* (2015) The gut microbiota regulates intestinal CD4 T cells expressing RORgammat and controls metabolic disease. *Cell Metab* 22: 100 – 112
28. Selkrig J, Wong P, Zhang X, Pettersson S (2014) Metabolic tinkering by the gut microbiome: implications for brain development and function. *Gut Microbes* 5: 369 – 380
29. Sharon G, Garg N, Debelius J, Knight R, Dorrestein PC, Mazmanian SK (2014) Specialized metabolites from the microbiome in health and disease. *Cell Metab* 20: 719 – 730
30. Ohnmacht C, Park JH, Cording S, Wing JB, Atarashi K, Obata Y, Gaboriau-Routhiau V, Marques R, Dulauroy S, Fedoseeva M *et al* (2015) Mucosal Immunology. The microbiota regulates type 2 immunity through RORgammat(+) T cells. *Science* 349: 989 – 993
31. Hildebrandt MA, Hoffmann C, Sherrill-Mix SA, Keilbaugh SA, Hamady M, Chen YY, Knight R, Ahima RS, Bushman F, Wu GD (2009) High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology* 137: 1716 – 1724.e2
32. Zhang H, DiBaise JK, Zuccolo A, Kudrna D, Braidotti M, Yu Y, Parameswaran P, Crowell MD, Wing R, Rittmann BE *et al* (2009) Human gut microbiota in obesity and after gastric bypass. *Proc Natl Acad Sci USA* 106: 2365 – 2370
33. Clarke SF, Murphy EF, Nilaweera K, Ross PR, Shanahan F, O'Toole PW, Cotter PD (2012) The gut microbiota and its relationship to diet and obesity: new insights. *Gut Microbes* 3: 186 – 202
34. Li F, Jiang C, Krausz KW, Li Y, Albert I, Hao H, Fabre KM, Mitchell JB, Patterson AD, Gonzalez FJ (2013) Microbiome remodelling leads to inhibition of intestinal farnesoid X receptor signalling and decreased obesity. *Nat Commun* 4: 2384
35. Sayin SI, Wahlstrom A, Felin J, Jantti S, Marschall HU, Bamberg K, Angelin B, Hyotylainen T, Oresic M, Backhed F (2013) Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist. *Cell Metab* 17: 225 – 235
36. Parseus A, Sommer N, Sommer F, Caesar R, Molinaro A, Stahlman M, Greiner TU, Perkins R, Backhed F (2016) Microbiota-induced obesity requires farnesoid X receptor. *Gut* doi:10.1136/gutjnl-2015-310283
37. Reinke H, Asher G (2016) Circadian clock control of liver metabolic functions. *Gastroenterology* 150: 574 – 580
38. Tamaru T, Hirayama J, Isojima Y, Nagai K, Norioka S, Takamatsu K, Sassone-Corsi P (2009) CK2alpha phosphorylates BMAL1 to regulate the mammalian clock. *Nat Struct Mol Biol* 16: 446 – 448
39. Sahar S, Zocchi L, Kinoshita C, Borrelli E, Sassone-Corsi P (2010) Regulation of BMAL1 protein stability and circadian function by GSK3beta-mediated phosphorylation. *PLoS ONE* 5: e8561
40. Schadinger SE, Bucher NL, Schreiber BM, Farmer SR (2005) PPARgamma2 regulates lipogenesis and lipid accumulation in steatotic hepatocytes. *Am J Physiol Endocrinol Metab* 288: E1195 – E1205
41. Uno K, Katagiri H, Yamada T, Ishigaki Y, Ogihara T, Imai J, Hasegawa Y, Gao J, Kaneko K, Iwasaki H *et al* (2006) Neuronal pathway from the liver modulates energy expenditure and systemic insulin sensitivity. *Science* 312: 1656 – 1659
42. Nielsen R, Pedersen TA, Hagenbeek D, Moulos P, Siersbaek R, Megens E, Denissov S, Borgesen M, Francoijs KJ, Mandrup S *et al* (2008) Genome-wide profiling of PPARgamma:RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis. *Genes Dev* 22: 2953 – 2967
43. Siersbaek MS, Loft A, Aagaard MM, Nielsen R, Schmidt SF, Petrovic N, Nedergaard J, Mandrup S (2012) Genome-wide profiling of peroxisome proliferator-activated receptor gamma in primary epididymal, inguinal, and brown adipocytes reveals depot-selective binding correlated with gene expression. *Mol Cell Biol* 32: 3452 – 3463
44. Kliewer SA, Sundseth SS, Jones SA, Brown PJ, Wisely GB, Koble CS, Devchand P, Wahli W, Willson TM, Lenhard JM *et al* (1997) Fatty acids and eicosanoids regulate gene expression through direct interactions with peroxisome proliferator-activated receptors alpha and gamma. *Proc Natl Acad Sci USA* 94: 4318 – 4323
45. Sauma L, Stenkula KG, Kjolhede P, Stralfors P, Soderstrom M, Nystrom FH (2006) PPAR-gamma response element activity in intact primary human adipocytes: effects of fatty acids. *Nutrition* 22: 60 – 68
46. Matsusue K, Kusakabe T, Noguchi T, Takiguchi S, Suzuki T, Yamano S, Gonzalez FJ (2008) Hepatic steatosis in leptin-deficient mice is promoted by the PPARgamma target gene Fsp27. *Cell Metab* 7: 302 – 311
47. Xu X, Park JG, So JS, Lee AH (2015) Transcriptional activation of Fsp27 by the liver-enriched transcription factor CREBH promotes lipid droplet growth and hepatic steatosis. *Hepatology* 61: 857 – 869
48. Moffat C, Bhatia L, Nguyen T, Lynch P, Wang M, Wang D, Ilkayeva OR, Han X, Hirschey MD, Claypool SM *et al* (2014) Acyl-CoA thioesterase-2 facilitates mitochondrial fatty acid oxidation in the liver. *J Lipid Res* 55: 2458 – 2470
49. Yazdanyar A, Jiang XC (2012) Liver phospholipid transfer protein (PLTP) expression with a PLTP-null background promotes very low-density lipoprotein production in mice. *Hepatology* 56: 576 – 584
50. Leesnitzer LM, Parks DJ, Bledsoe RK, Cobb JE, Collins JL, Consler TG, Davis RG, Hull-Ryde EA, Lenhard JM, Patel L *et al* (2002) Functional consequences of cysteine modification in the ligand binding sites of peroxisome proliferator activated receptors by GW9662. *Biochemistry* 41: 6640 – 6650
51. Rakoff-Nahoum S, Paglino J, Eslami-Varzaneh F, Edberg S, Medzhitov R (2004) Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell* 118: 229 – 241
52. Mukherji A, Kobiita A, Ye T, Chambon P (2013) Homeostasis in intestinal epithelium is orchestrated by the circadian clock and microbiota cues transduced by TLRs. *Cell* 153: 812 – 827

53. Thaiss CA, Zeevi D, Levy M, Zilberman-Schapira G, Suez J, Tengeler AC, Abramson L, Katz MN, Korem T, Zmora N *et al* (2014) Transkingdom control of microbiota diurnal oscillations promotes metabolic homeostasis. *Cell* 159: 514 – 529
54. Voigt RM, Forsyth CB, Green SJ, Mutlu E, Engen P, Vitaterna MH, Turek FW, Keshavarzian A (2014) Circadian disorganization alters intestinal microbiota. *PLoS ONE* 9: e97500
55. Zarrinpar A, Chaix A, Yooseph S, Panda S (2014) Diet and feeding pattern affect the diurnal dynamics of the gut microbiome. *Cell Metab* 20: 1006 – 1017
56. Yang X, Downes M, Yu RT, Bookout AL, He W, Straume M, Mangelsdorf DJ, Evans RM (2006) Nuclear receptor expression links the circadian clock to metabolism. *Cell* 126: 801 – 810
57. Eckel-Mahan KL, Patel VR, Mohney RP, Vignola KS, Baldi P, Sassone-Corsi P (2012) Coordination of the transcriptome and metabolome by the circadian clock. *Proc Natl Acad Sci USA* 109: 5541 – 5546
58. Sahar S, Sassone-Corsi P (2012) Regulation of metabolism: the circadian clock dictates the time. *Trends Endocrinol Metab* 23: 1 – 8
59. Tontonoz P, Spiegelman BM (2008) Fat and beyond: the diverse biology of PPARgamma. *Annu Rev Biochem* 77: 289 – 312
60. Green CB, Douris N, Kojima S, Strayer CA, Fogerty J, Lourim D, Keller SR, Besharse JC (2007) Loss of Nocturnin, a circadian deadenylase, confers resistance to hepatic steatosis and diet-induced obesity. *Proc Natl Acad Sci USA* 104: 9888 – 9893
61. Grimaldi B, Bellet MM, Katada S, Astarita G, Hirayama J, Amin RH, Granneman JG, Piomelli D, Leff T, Sassone-Corsi P (2010) PER2 controls lipid metabolism by direct regulation of PPARgamma. *Cell Metab* 12: 509 – 520
62. Kawai M, Rosen CJ (2010) PPARgamma: a circadian transcription factor in adipogenesis and osteogenesis. *Nat Rev Endocrinol* 6: 629 – 636
63. Backhed F, Ding H, Wang T, Hooper LV, Koh GY, Nagy A, Semenkovich CF, Gordon JI (2004) The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci USA* 101: 15718 – 15723
64. Singh V, Chassaing B, Zhang L, San Yeoh B, Xiao X, Kumar M, Baker MT, Cai J, Walker R, Borkowski K *et al* (2015) Microbiota-dependent hepatic lipogenesis mediated by stearoyl CoA desaturase 1 (SCD1) promotes metabolic syndrome in TLR5-deficient mice. *Cell Metab* 22: 983 – 996
65. den Besten G, Lange K, Havinga R, van Dijk TH, Gerding A, van Eunen K, Muller M, Groen AK, Hooiveld GJ, Bakker BM *et al* (2013) Gut-derived short-chain fatty acids are vividly assimilated into host carbohydrates and lipids. *Am J Physiol Gastrointest Liver Physiol* 305: G900 – G910
66. den Besten G, Bleeker A, Gerding A, van Eunen K, Havinga R, van Dijk TH, Oosterveer MH, Jonker JW, Groen AK, Reijngoud DJ *et al* (2015) Short-chain fatty acids protect against high-fat diet-induced obesity via a PPARgamma-dependent switch from lipogenesis to fat oxidation. *Diabetes* 64: 2398 – 2408
67. Luo Y, Chen GL, Hannemann N, Ipseiz N, Kronke G, Bauerle T, Munos L, Wirtz S, Schett G, Bozec A (2015) Microbiota from obese mice regulate hematopoietic stem cell differentiation by altering the bone niche. *Cell Metab* 22: 886 – 894
68. Baldi P, Long AD (2001) A Bayesian framework for the analysis of microarray expression data: regularized *t*-test and statistical inferences of gene changes. *Bioinformatics* 17: 509 – 519
69. Kayala MA, Baldi P (2012) Cyber-T web server: differential analysis of high-throughput data. *Nucleic Acids Res* 40: W553 – W559
70. Carmona-Saez P, Chagoyen M, Tirado F, Carazo JM, Pascual-Montano A (2007) GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. *Genome Biol* 8: R3

Chapter 4

Functional Enrichment Pipelines

4.1 Introduction

Regardless of how one processes the time series signals through statistical or machine learning methods, there is still a gap between those results and meaningful biological discoveries. This gap requires the annotation and interpretation of signals.

In a typical transcriptomic experiment, one analyzes more than 20,000 transcripts under one experimental condition. Grouping these transcripts results in an astronomically large number of possible sets. It is naturally the task of bioinfomatics softwares to annotate each transcript and discover meanings from sets of transcripts which may have particularly high number of annotations of a certain kind. This is referred to as a *functional enrichment*.

The circadian functional enrichment pipelines were designed specifically for this analysis. Again we focus on the transcriptome, but extend also to the metabolome: `factor_enrichment` is an annotation and enrichment analysis pipeline centered around predicted binding sites produced by MotifMap and MotifMap-RNA, with additional results produced by the ChIPSeq

pipeline as well. `mpfe` is a pathway based annotation and enrichment analysis pipline centered around KEGG transcriptomic pathways, with additional information from KEGG metabolomic pipelines and GO term association. Together, these two pipelines produce robust and detailed annotation and enrichment reports based on transcriptomic or metabolomic inputs in the form of sets of circadian species. Additionally, they can be readily used on non-circadian data such as differential analysis data.

While set based enrichment analyses have been proven to be effective in interpreting biological meanings from high-throughput data, they are an over simplification. Real mechanisms cannot be explained simply based on functional groups. A more sophisticated approach is to study circadian species as interacting nodes in a complex regulatory network. In this network, functional information can be used to infer directed regulatory edges. `pyfuncgraph` is a software library designed for supporting such analyses. It constructs meta-functional graphs based on predicted binding sites, experimental binding peaks, or pathway and annotation association. Such graphs can be intersected with circadian data to produce a *circadian functional graph*, which serves as an automatically generated, data-driven model, or a knowledge base, for explaining the organization of circadian species under certain experimental conditions.

4.2 Overview of the Enrichment Pipelines

The `factor_enrichment` pipeline takes a set of enriched gene or transcript symbols as input. The user can conduct set intersection or union operations and use any of the resulting new sets as inputs. To perform set enrichment analysis, the pipeline utilizes results from MotifMap, MotifMap-RNA and ChIPSeq databases previously described in Chapter 2. Binding site or ChIPSeq peak results are queried from the MySQL database with a fixed set filtering thresholds for each type of data.

Category	Source:Input Set	Source:Background
Binding Site Hits	Hits from the Input Set (exact)	Total Hits (exact)
Missing Binding Sites	Missing from Input (est.)	Total Missing (est.)

Table 4.1: Demonstration of how a contingency table is set up for Fisher's Exact Test.

For MotifMap, binding sites are predicted at the promoter region of genes (-10,000 $\sim$ +2,000 bps relative to the transcription start site), using filtering thresholds $BBLS \geq 1, NLOD \geq 0.9, FDR \leq 0.25$. Results from more than 1,300 transcription factors are available. For MotifMap-RNA, binding sites are predicted for more than 350 RNA binding proteins at the UTRs or introns (with 200 bps flanking distance) of transcripts, with filtering thresholds $NLOD \geq 0.9, BBLS > 0$. For ChIPSeq peaks, experimental results from more than 100 TFs or histones are filtered close to the promoter regions (defined the same as in MotifMap), with peak scores percentile $\geq 50\%$. These binding sites or peak results are gathered for each gene or transcript in the input set and are compared to a background set or background genome for enrichment analysis. The background set can be provided by the user (e.g., a set of all transcripts measured in a micro-array experiment) while a background genome is provided by the pipeline for mouse (mm10).

Enrichment analysis is performed via Fisher's exact test on a contingency table using the input set and the background set/genome. A novel approach is adopted to estimate more precise out-of-group numbers than simply counting the number of genes without binding sites. Instead, average binding sites or peaks per target is estimated using the background set or background genome and an expected value of binding sites or peaks is used for the out-of-group numbers. A demonstration of this contingency table is shown in Table 4.1. Furthermore, Benjamini-Hochberg q values (BHQ) are generated to correct for multiple testing between different source factors (e.g. MotifMap TFs, RBPs). A novel invention of this pipeline is the *meta-factor enrichment analysis*, which is performed on multiple input sets generating set-specific enrichment results. Factors common or unique to these enriched sets are then extracted from each enrichment analysis and their shared enrichment status is

analyzed. Multiple testing correction via BHQ is also performed.

The `mpfe` (multiple pathways functional enrichment) pipeline performs a similar set based enrichment analysis. However, instead of binding sites or ChIPSeq peaks, results are generated by pathway association and GO term association. Two publicly available databases, CPDB and PathwayCommons [26, 11], are used as data sources. These databases incorporate pathway and annotation information from KEGG [27, 28] and Gene Ontology [5]. Genes and metabolites are organized by their shared pathways or level 2 GO terms. A Fisher's exact test is performed per pathway or GO term, where the contingency table is built based on gene symbols or metabolite names. A novel aspect of this pipeline is the *metabolite-transcript association enrichment analysis*. This analysis associates metabolites with transcripts (and vice versa) and performs enrichment analysis on those associated metabolites or transcripts. This was done by querying the KEGG enzymatic reaction database (via PathwayCommons). Figures 4.1 and Figure 4.2 showcase a set of sample output from the this novel analysis.

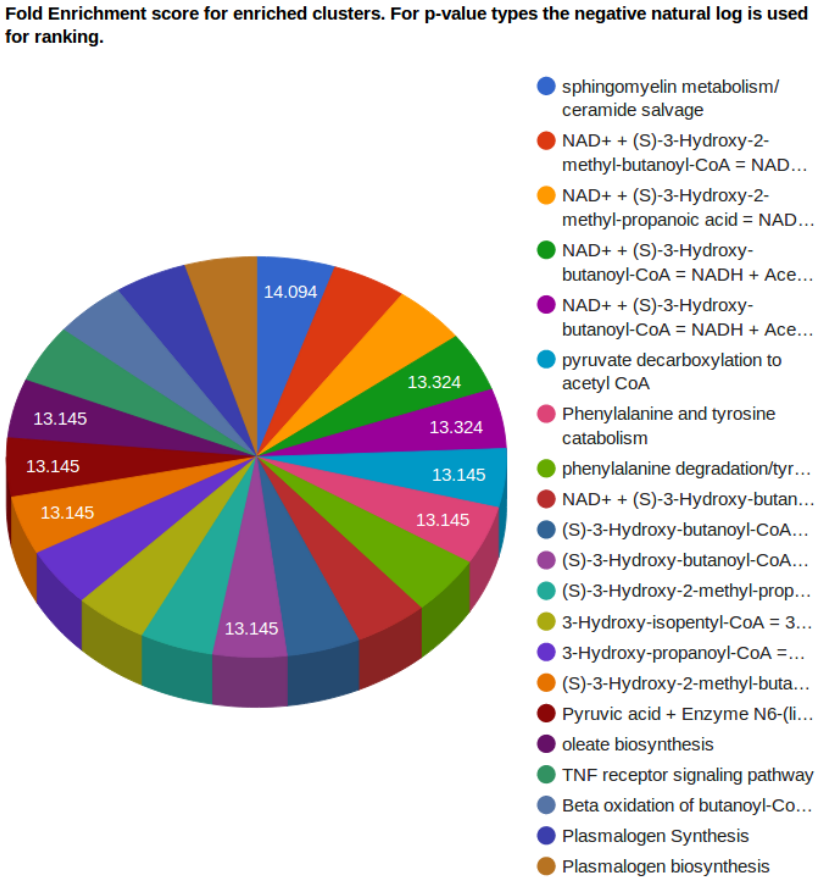
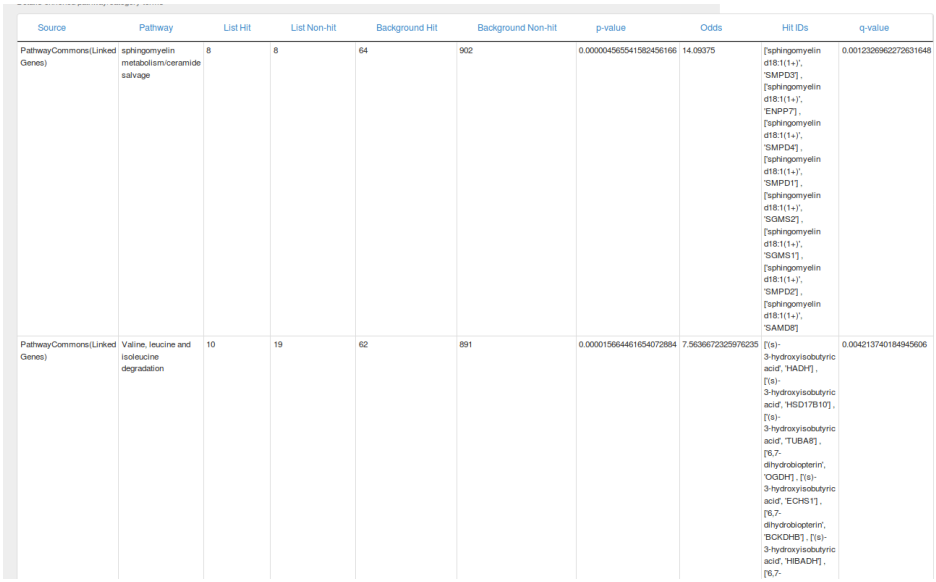


Figure 4.1: Enriched reaction pathways, ranked by fold changes (odds ratio in Fisher's exact test), generated from genes (transcripts) associated with an input set of metabolites. Data taken from a mouse brain metabolome study, Tognini et al 2018, publication pending.



Source	Pathway	List Hit	List Non-hit	Background Hit	Background Non-hit	p-value	Odds	Hit IDs	q-value
PathwayCommons(Linked Genes)	sphingomyelin metabolism/ceramide salvage	8	8	64	902	0.00000456554158245616 6	14.09375	['sphingomyelin d18:1(1+)', 'SMPD3'] , ['sphingomyelin d18:1(1+)', 'ENPP7'] , ['sphingomyelin d18:1(1+)', 'SMPD4'] , ['sphingomyelin d18:1(1+)', 'SMPD1'] , ['sphingomyelin d18:1(1+)', 'SGMS2'] , ['sphingomyelin d18:1(1+)', 'SGMS1'] , ['sphingomyelin d18:1(1+)', 'SMPD2'] , ['sphingomyelin d18:1(1+)', 'SAMD8']	0.00123269622726316 48
PathwayCommons(Linked Genes)	Valine, leucine and isoleucine degradation	10	19	62	891	0.00001566446165407288 4	7.5636672325976235	['(s)-3-hydroxyisobutyric acid', 'HADH'] , ['(s)-3-hydroxyisobutyric acid', 'HSD17B10'] , ['(s)-3-hydroxyisobutyric acid', 'TUBA8'] , ['6,7-dihydrobiopterin', 'OGDH'] , ['(s)-3-hydroxyisobutyric acid', 'ECHS1'] , ['6,7-dihydrobiopterin', 'BCKDHB'] , ['(s)-3-hydroxyisobutyric acid', 'HIBADH'] , ['6,7-	0.00421374018494560 6

Figure 4.2: Detail of select enriched reaction pathways generated from genes (transcripts) associated with an input set of metabolites. Data taken from a mouse brain metabolome study, Tognini et al 2018, publication pending.

Both enrichment pipelines connect to the `viz` rendering engine. Outputs are integrated into the `goworg` project websites. Compared to existing enrichment analysis tools such as DAVID [23], these pipelines utilize novel data sources (MotifMap and MotifMap-RNA), perform novel analyses (metabolome enrichment, metabolite-transcript association enrichment) and run much faster on local environments (most similar services such as DAVID require using a web interface). Example application of the generated results are shown in a mouse microbiome study ([41], Chapter 3), in a tumor bearing mouse study ([37], Chapter 5), and a mouse ketogenic study ([49], attached below).

4.3 Application: Mouse Ketogenic Diet Study Paper

The functional enrichment pipelines, as well as other methods mentioned in this thesis, are utilized extensively to conduct bioinformatics analysis of the mouse liver and gut under different diet treatments by Tognini et al., detailed in the appended paper below [49], for which I am a first co-author.

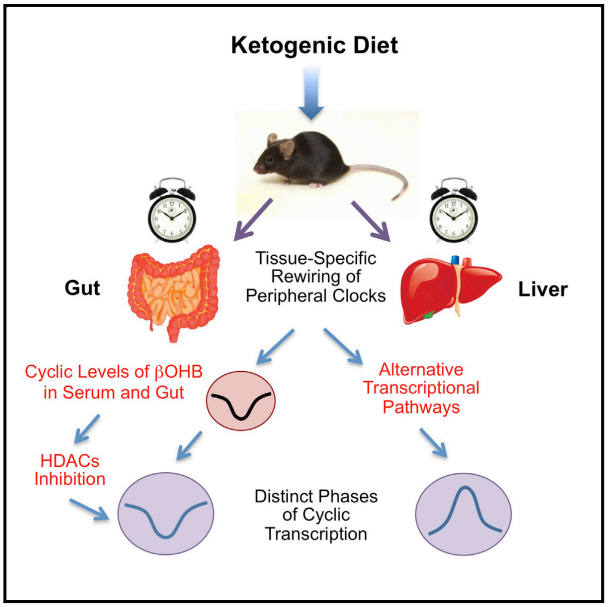
In particular, Figure 1 C,G, Figure 2 H and accompanying texts utilize the pathway enrichment analysis and meta-factor enrichment analysis. Functional enrichment analysis of circadian transcripts under specific diet conditions reveals that distinct functional pathways are activated by the same diet treatment in different tissues. In particular, in the liver, under ketogenic diet (KD), genes associated with endoplasmic reticulum, TNF signaling pathways and glucagon signaling pathways are enriched. In contrast, in the gut, under the same treatment, genes are enriched for pathways involved in amino acid degradation and fatty acid metabolism. A meta-factor enrichment analysis was performed to elucidate the underlying mechanisms for this tissue specificity. Results identified key transcription factors such as Pparα and Stat5b. The full discovery and analysis is available in the appended paper below.

Article

Cell Metabolism

Distinct Circadian Signatures in Liver and Gut Clocks Revealed by Ketogenic Diet

Graphical Abstract



Highlights

- KD induces tissue-specific reprogramming of the circadian clock in liver and gut
- KD induces an increase in liver BMAL1 chromatin recruitment and amplitude of CCGs
- KD drives tissue-specific oscillation of PPARα and its target genes
- Oscillation of βOHB in gut and serum parallels gut-specific cycling of H3 acetylation

Authors

Paola Tognini, Mari Murakami, Yu Liu, ..., Eric Verdin, Pierre Baldi, Paolo Sassone-Corsi

Correspondence

psc@uci.edu

In Brief

Tognini et al. reveal how a ketogenic diet (KD) differently affects liver and intestine circadian clocks and drives tissue-specific oscillation of PPARα and its target genes. Serum and intestine βOHB shows a unique diurnal rhythmicity, associated with daily epigenetic changes exclusively in the gut.

Tognini et al., 2017, Cell Metabolism *26*, 523–538
September 5, 2017
http://dx.doi.org/10.1016/j.cmet.2017.08.015

CellPress

Cell Metabolism
Article

Distinct Circadian Signatures in Liver and Gut Clocks Revealed by Ketogenic Diet

Paola Tognini,[1,5] Mari Murakami,[1,5] Yu Liu,[2,5] Kristin L. Eckel-Mahan,[1,3] John C. Newman,[4] Eric Verdin,[4] Pierre Baldi,[2] and Paolo Sassone-Corsi[1,6,*]
[1]Center for Epigenetics and Metabolism, Department of Biological Chemistry, U1233 INSERM
[2]Institute for Genomics and Bioinformatics, School of Information and Computer Sciences
University of California, Irvine, Irvine, CA, USA
[3]Center for Metabolic and Degenerative Diseases, Institute of Molecular Medicine, University of Texas Health Sciences Center, Houston, TX, USA
[4]Gladstone Institutes, University of California, San Francisco, 1650 Owens Street, San Francisco, CA 94158, USA
[5]These authors contributed equally
[6]Lead Contact
*Correspondence: psc@uci.edu
http://dx.doi.org/10.1016/j.cmet.2017.08.015

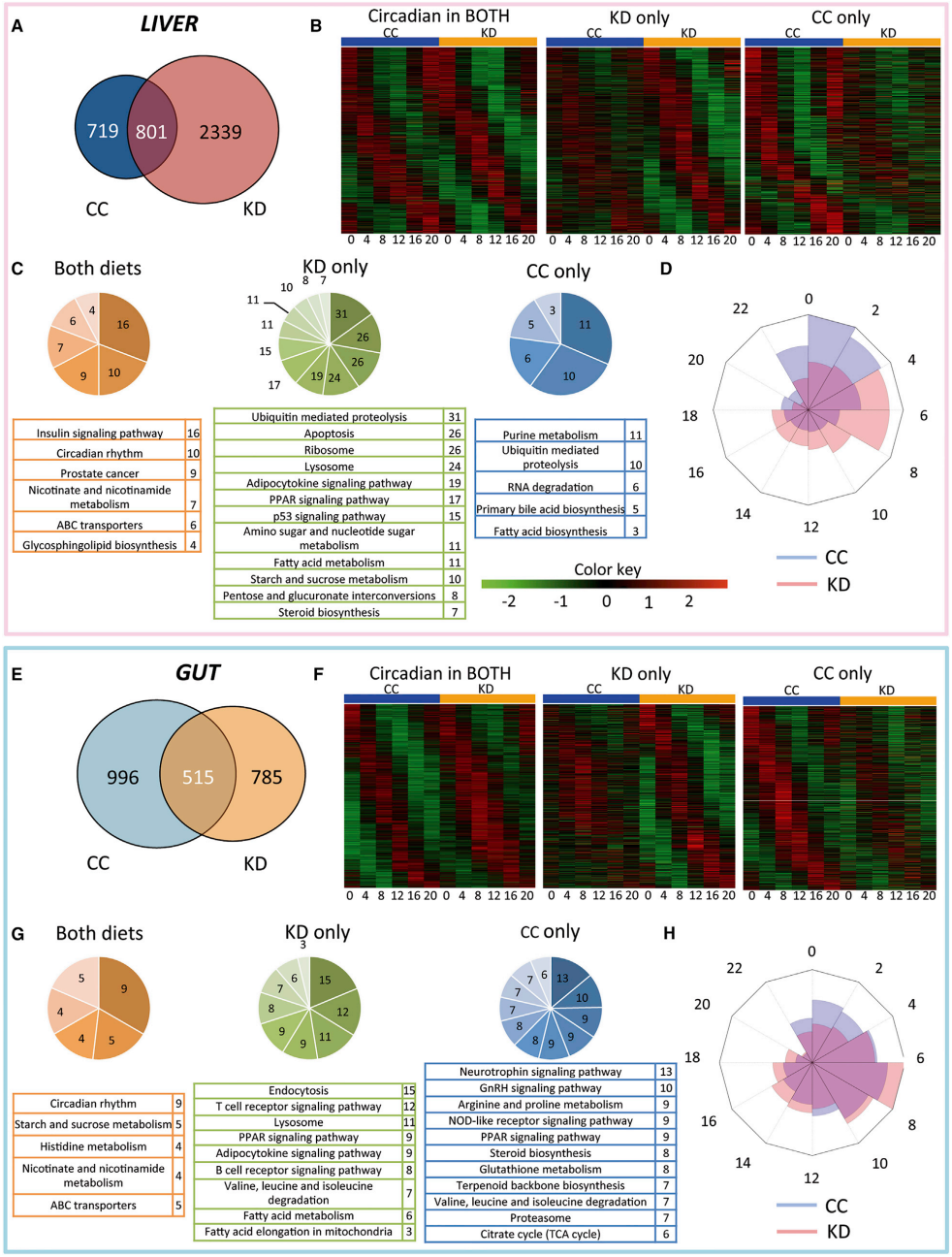
SUMMARY

The circadian clock orchestrates rhythms in physiology and behavior, allowing organismal adaptation to daily environmental changes. While food intake profoundly influences diurnal rhythms in the liver, how nutritional challenges are differentially interpreted by distinct tissue-specific clocks remains poorly explored. Ketogenic diet (KD) is considered to have metabolic and therapeutic value, though its impact on circadian homeostasis is virtually unknown. We show that KD has profound and differential effects on liver and intestine clocks. Specifically, the amplitude of clock-controlled genes and BMAL1 chromatin recruitment are drastically altered by KD in the liver, but not in the intestine. KD induces nuclear accumulation of PPARα in both tissues but with different circadian phase. Also, gut and liver clocks respond differently to carbohydrate supplementation to KD. Importantly, KD induces serum and intestinal β-hydroxyl-butyrate levels to robustly oscillate in a circadian manner, an event coupled to tissue-specific cyclic histone deacetylase (HDAC) activity and histone acetylation.

INTRODUCTION

The circadian clock is an endogenous timekeeper that provides organisms with the ability to anticipate daily fluctuations in the environment, thus allowing an appropriate physiological adaptation (Bass and Takahashi, 2010; Eckel-Mahan and Sassone-Corsi, 2013). The molecular clock is characterized by complex transcriptional-translational feedback loops involving both transcriptional activators and repressors. The positive limb of the mammalian clock machinery comprises CLOCK and BMAL1, which heterodimerize and induce the expression of clock-controlled genes (CCGs) by binding E-boxes on target promoters. Cryptochrome (*Cry1* and *Cry2*) and Period (*Per1*, *Per2*, and *Per3*) genes encode proteins that form the negative limb, inhibiting CLOCK:BMAL1-mediated transcription (Asher and Sassone-Corsi, 2015; Tamanini et al., 2007). Another important diurnal loop involves the transcriptional activator retinoid-related orphan receptors (RORs) and the repressor REV-ERBα/REV-ERBβ, which induces *Bmal1* rhythmic expression and delays *Cry1* expression (Everett and Lazar, 2014; Partch et al., 2014). A number of studies have shown that chromatin remodeling plays a central role in the harmonic oscillation of circadian gene expression (Koike et al., 2012). Among the chromatin remodelers linked to the molecular clock are the nuclear sirtuins, specifically SIRT1 (Asher et al., 2008; Nakahata et al., 2008) and SIRT6 (Masri et al., 2014). These are deacetylases that use NAD^+ as a co-enzyme, thus linking cellular energy metabolism with epigenetic control. Remarkably, SIRT1 activity and the levels of NAD^+ oscillate in a 24 hr manner since the molecular clock directly controls the circadian transcription of the gene encoding the enzyme nicotinamide phosphorybosiltransferase (NAMPT) (Nakahata et al., 2009; Ramsey et al., 2009). NAMPT operates as the rate-limiting step enzyme in the NAD-salvage pathway. Notably, NAD^+ oscillation is abolished in the liver under nutritional challenge by high-fat diet because of the lack of chromatin recruitment of the CLOCK:BMAL1 complex to the *Nampt* gene promoter (Eckel-Mahan et al., 2013).

The mammalian circadian system is a hierarchical network in which the central oscillator, located in the suprachiasmatic nuclei (SCN) of the hypothalamus, acts in concert with peripheral clocks (Mohawk et al., 2012; Okamura, 2004; Schibler and Sassone-Corsi, 2002). Importantly, both organismal and cellular metabolism are tightly interlocked with circadian rhythms and food challenges, such as high-fat diet or restricted feeding, that are able to profoundly remodel the liver circadian transcriptome and metabolome (Eckel-Mahan et al., 2013; Hatori et al., 2012; Kohsaka et al., 2007; Stokkan et al., 2001; Vollmers et al., 2009). The gut has a powerful endogenous clock that is thought to be implicated in time-specific food intake. The disruption of the clock causes higher risk of gut inflammation (Bellet et al., 2013; Tognini et al., 2017) and increased intestinal permeability (Summa et al., 2013). Interestingly, the crosstalk between gut microbiota and intestinal epithelial cells is mediated by the host clock, and clock impairment due to the absence of



(legend on next page)

microbial signaling induces a pre-diabetic status (Mukherji et al., 2013). In addition, food challenge-driven remodeling in the gut microbiota can transcriptionally rewire the liver clock (Leone et al., 2015; Murakami et al., 2016).

Both liver and intestine clocks play a key role in sustaining metabolic homeostasis. However, little is known about how the gut and liver clocks can specifically interpret a nutritional challenge and how diet variations might uniquely affect the circadian physiology of distinct peripheral organs. To address this conceptually critical question, we reasoned that the ketogenic diet (KD) would provide the unique opportunity to explore how specific diet-generated metabolites would differentially impact distinct peripheral clocks. Indeed, KD is a high-fat, adequate protein, very-low-carbohydrate diet that induces a switch to fatty acid oxidation as an energy source. The result is an overproduction of acetyl-CoA, which leads to the synthesis of ketone bodies via the activation of the ketogenic pathway. Ketone bodies refer to three distinct molecules, acetone, acetoacetic acid, and β- hydroxyl-butyrate (βOHB), that are produced principally in the mitochondrial matrix in the liver, during fasting or prolonged exercise (Hawley et al., 2014; Newman and Verdin, 2014). Mice fed a KD present a low level of glucose but increased blood level of βOHB, which is used as an energy source by the brain and other tissues (Douris et al., 2015; Kennedy et al., 2007). Interestingly, a central role for rhythmic release of βOHB from the liver appears to be implicated in driving food anticipation via feedback to the hypothalamus (Chavan et al., 2016). Furthermore, βOHB seems to function as a cofactor for a recently defined epigenetic mark, histone β-hydroxyl-butyrylation, involved in the control of gene expression under specific metabolic states (Xie et al., 2016).

Another relevant facet of KD is its use in therapeutics to treat refractory epilepsy in children as well as obesity and metabolic diseases. Moreover, KD has been under evaluation for alternative indications such as polycystic ovarian syndrome, cancer, and neurodegenerative diseases (Paoli et al., 2013). Despite this biomedical use, little is known about how KD impacts tissue-specific gene expression and how it may influence circadian homeostasis. Here we show that KD triggers distinct transcriptional responses by activating unique tissue-specific pathways involved in rewiring cyclic gene expression. Finally, we reveal that serum and intestinal βOHB levels display diurnal rhythmicity, which accompany a time-dependent HDAC activity and histone acetylation primarily in the gut. Our findings support the view that ketogenic-based nutrition contributes to diurnal transcriptional rewiring via metabolite-driven chromatin remodeling.

RESULTS

KD Abolishes Rhythmicity in Respiratory Metabolism

To explore how KD influences circadian physiology, we submitted 8-week-old C57BL/6 mice to KD dietary regimen during 4 weeks and compared them to control chow (CC)-fed mice. KD-fed mice lost weight in the first 2 weeks of the new dietary regimen and then stabilized to the same weight as CC-fed animals (Figure S1A), in line with previous studies (Kennedy et al., 2007). As expected, serum βOHB concentration significantly increased in KD-fed animals (Figure S1B). KD induced a significant decrease in serum glucose with respect to CC, although it remained in a physiological range (Figure S1C). To further understand the metabolic state induced by KD, animals were analyzed by indirect calorimetry for CO_2 emission, O_2 consumption, and energy intake. Notably, while the respiratory exchange ratio (RER) of normally fed mice oscillated along the circadian cycle, RER rhythmicity was abolished in KD-fed mice and remained flat throughout the circadian cycle with an average value of 0.7 (Figures S1D and S1E). This indicated that fat was the only fuel metabolized by KD-fed mice throughout the circadian cycle. Caloric and water intake (Figures S1F and S1G), as well as the feeding pattern (Figure S1F), were equivalent in CC- and KD-fed mice, showing that rhythmicity in food intake was not altered under KD.

Tissue-Specific Remodeling of Diurnal Transcription by KD

To gain further insight into how KD influences diurnal metabolism, we performed high-throughput profiling of hepatic tissue and ileal intestinal epithelial cells (IECs) by transcriptome microarrays. The analysis of rhythmic transcripts was performed using the non-parametric test JTK_CYCLE (Hughes et al., 2010), incorporating a window of 20–28 hr for the determination of circadian periodicity.

A first notable finding is that, while the number of cyclic genes in CC-fed mice was comparable between the liver and the gut, this ratio was drastically different in KD-fed mice, with a much larger number of genes oscillating in the liver (Figures 1A and 1E). Our analysis showed that 2,339 genes started to oscillate de novo in the liver upon KD, while cyclic transcripts under CC were 719. Genes displaying rhythmicity under both nutritional conditions were 801 in the liver (Figures 1A and 1B). Gene annotation analysis revealed specific categories for all groups (Figures 1C and S2A; Table S1). Remarkably, the effect of KD on the number of cycling transcripts in IECs was the opposite, with a decrease in oscillating genes (785, KD-only rhythmic

Figure 1. Heatmaps, KEGG Pathway, and Phase Lag Analyses of the Diurnal Transcriptome in Liver and Gut upon KD

(A) Venn diagram representing the number of genes rhythmic only in KD, only in CC, and in both conditions in the liver (n = 3 per time point, per group, $p < 0.01$, false discovery rate [FDR] < 0.1).

(B) Heatmaps representing the genes diurnal in both conditions, KD only, and CC only in liver (n = 3 per time point, per group, $p < 0.01$, FDR < 0.1).

(C) KEGG pathway analysis in liver. The numbers in the pie charts and tables represent the number of genes enriched in the specific KEGG pathway.

(D) Radar plots representing the phase lag of genes exclusively diurnal in KD or CC in hepatic tissue (Anderson-Darling test; phase distributions significance, $p < 0.0001$).

(E) Venn diagram representing the number of genes oscillating only in KD, only in CC, and in both conditions in the intestine (n = 3 per time point, per group, $p < 0.01$, FDR < 0.1).

(F) Heatmaps representing the genes diurnal in both conditions, KD only, and CC only in the gut (n = 3 per time point, per group, $p < 0.01$, FDR < 0.1).

(G) KEGG pathway analysis in the intestine. The numbers in the pie charts and tables represent the number of genes enriched in the specific KEGG pathway.

(H) Radar plots representing the phase lag of genes exclusively diurnal in KD or CC in ileal cells (Anderson-Darling test; phase distribution significance in IECs, KD versus CC, $p < 0.0001$; comparisons not shown; phase distribution difference upon KD, liver versus IECs, $p < 0.0001$; phase distribution difference upon CC, liver versus IECs, $p < 0.0001$).

transcripts) with respect to CC condition (996, CC-only rhythmic transcripts). Cyclic genes in both CC and KD in IECs were 515 (Figures 1E and 1F), and the most overrepresented Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways among these genes, such as "circadian rhythm" and "nicotinate and nicotinamide metabolism," were the same of the liver "both" genes (Figures 1C, 1G, S2A, and S2B; Tables S1 and S2). The heatmaps of the oscillating transcripts showed the loss of rhythmicity in CC (Figures 1B and 1F, second map) and in KD (Figures 1B and 1F, third map). Disruption of diurnal oscillations in gene expression was particularly evident with an increased p value threshold ($p < 0.05$) in both liver (Figure S3A) and intestine (Figure S3B). For the complete list of genes oscillating in liver and IECs, see Tables S3 and S4.

It is worth noting that the transcripts rhythmic under either KD or CC (KD only and CC only) were enriched in a variety of different pathways in a tissue-specific manner (Figures 1C, 1G, S2A, and S2B; Tables S1 and S2), underscoring the unique metabolic states induced by KD in the gut and liver clocks.

To explore the physiological relevance of the diurnal alterations in the transcriptome under KD, we analyzed the levels of specific metabolites in the liver and intestine of KD-fed mice. Our KEGG pathway analysis indicated that "steroid biosynthesis" was a metabolic pathway circadian in KD-only liver (Table S1, sheet 2) and in CC-only IECs (Table S2, sheet 1). Indeed, we found a significant increase in total cholesterol levels in KD liver with respect to CC and a significant difference between zeitgeber time 0 (ZT0) and ZT12 (Figure 2A). On the other hand, we could not monitor convincing time-of-the-day-dependent changes in gut cholesterol levels or a significant difference between CC and KD (Figure 2C). The KEGG pathway analysis also indicated "fatty acid metabolism" as diurnal in both liver and IECs upon KD (Tables S1 and S2, sheet 2). In the liver, there was an increase in free fatty acid concentration upon KD, although no rhythmic alterations between the ZTs tested (Figure 2B). Strikingly, we found a strong increase in free fatty acid concentration in the gut at nighttime (ZT20), while in CC the levels were unchanged during night and day (Figure 2D). As fatty acids are ligands of the peroxisome proliferator-activated (PPAR) family of nuclear receptors (Grygiel-Górniak, 2014), these data nicely match the rhythmic nuclear accumulation of PPARα (Figure 4D) and its target gene expression (Figure 4E) in the intestine of mice fed to KD. In keeping with higher, but not diurnal, fatty acid levels, there was a non-circadian increase in the expression of PPARα target genes in the liver (Figure 4B) (see Cyclic Activation of PPARα by KD).

Interestingly, the KD-induced diurnal genes oscillated in a coordinated manner, in phase with a sharp peak at ZT6–ZT8 in the liver and ZT8 in the intestine, as shown by time lag analysis (Figures 1D and 1H). The genes oscillating in CC only were still coordinated in phase with an earlier peak around ZT2 in the liver and ZT8 in the gut (Figures 1D and 1H). Moreover, the amplitude analysis in "both diets" condition underlined another tissue-specific feature of the KD challenge: of all common oscillators, 54.7% of the 801 transcripts displayed an increase in amplitude in the liver upon KD feeding (Figure S2C). In contrast, in the gut 52% of the genes showed a decrease in amplitude and only 43% an increase (Figure S2D). Altogether, these data indicate that distinct diets control the phase and amplitude of oscillation of the circadian transcriptome via distinct mechanisms in different tissues.

To further determine the diurnal transcriptional signature induced by KD in the two distinct tissues, we analyzed the KD-driven rhythmic transcripts in liver and gut (Figure 2E; Table S5). The genes cycling exclusively upon KD in the liver were 2,724 and they clustered in the KEGG pathways "protein processing in endoplasmic reticulum," "TNF signaling pathway," and "glucagon signaling pathway" (see Table S6, sheet 1, for a complete list). Only 884 genes were rhythmic exclusively in the gut (Figure 2E) and they were prevalently enriched in "metabolic pathways"; "valine, leucine, and isoleucine degradation"; and "fatty acid degradation and elongation" (Table S6, sheet 2). The shared rhythmic genes were 416 and, as expected, comprised the core clock genes (Figure 2E; Table S6, sheet 3). The heatmaps of the oscillating transcripts showed a total disruption in diurnal rhythms of the exclusive cycling genes between the two tissues (Figure 2F), reinforcing the notion of transcriptional reprogramming in response to a food challenge of specific peripheral clocks (Eckel-Mahan et al., 2013).

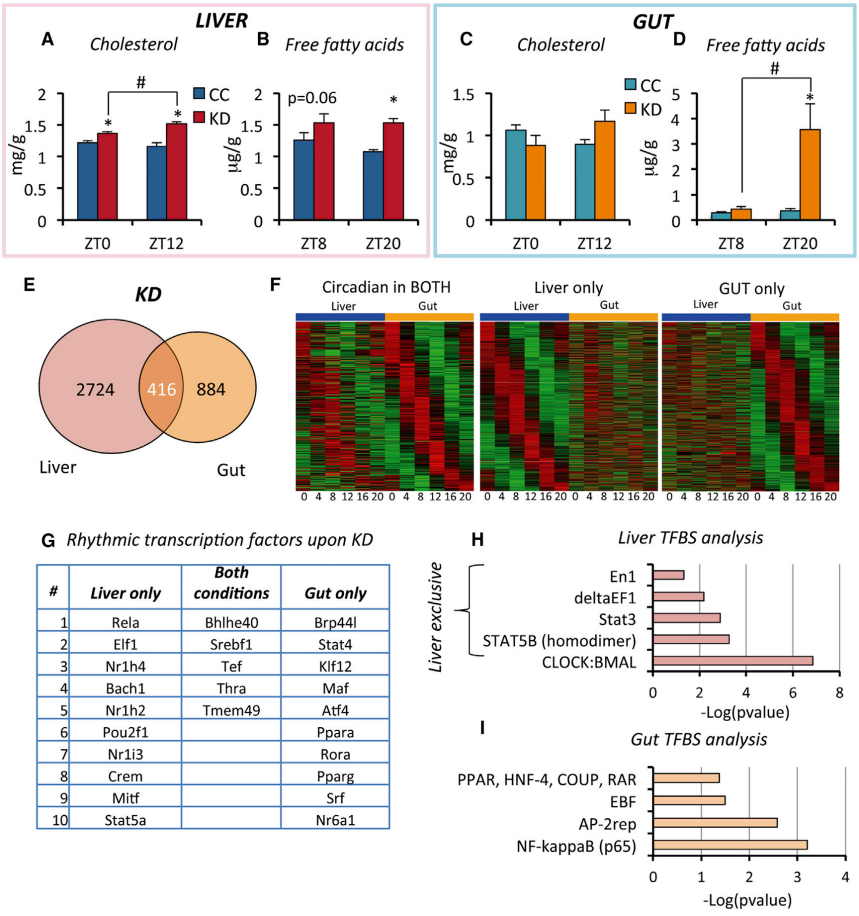
Furthermore, to gain insight into the differential influence of KD on the liver and gut clock, transcription factor binding site (TFBS) enrichment analysis was performed based on MotifMap (Daily et al., 2011; Xie et al., 2009) binding site results (Table S6, sheet 4), and a further meta-analysis was used to identify tissue-specific rhythmic transcription factors (Figure 2G). Interestingly, *Ppar*α, a nuclear receptor involved in ketogenic responses, was a transcription factor rhythmic in KD-only IECs (Figure 2G) and its binding motif was highly significant in genes rhythmic exclusively under this condition (Figure 2I). This suggested that the PPARα pathway could represent a specific diurnal signature in the KD-dependent rhythmic transcriptional rewiring of the gut clock. Other transcription factors displayed an oscillatory profile in the liver (Figure 2G) and some of the motifs bound by the STAT protein family were enriched in KD-only liver (Figure 2H). Importantly, "CLOCK:BMAL1" binding site was enriched in virtually all the conditions (Table S6, sheet 4) except for KD-only IECs when compared to KD-only liver. Moreover, "CLOCK:BMAL1" binding site was strongly significant in KD-only liver (Table S6, sheet 4; Figure 2H), implying a possible direct involvement of the core clock in interpreting a ketogenic nutritional challenge on hepatic tissue physiology.

Limited Effect of KD on Core Clock Gene Oscillation

The significant influence of KD on cyclic gene expression prompted us to determine whether it would directly affect the core clock machinery. To do so, we analyzed the expression of *Bmal1*, *Cry1*, *Per2*, and *Rev-erb*α genes and found no significant alteration in phase and amplitude when comparing CC- and KD-fed mice in both liver (Figure S3C) and IECs (Figure S3D). Moreover, BMAL1 protein and phosphorylation levels were virtually unaltered by KD as compared to CC in both liver (Figure S3E) and gut (Figure S3F) along the entire circadian cycle. Thus, core clock gene expression appears to be resistant to potential perturbations caused by a food challenge in two distinct peripheral clock systems.

Tissue-Specific Effect of KD on BMAL1 Chromatin Recruitment

Given the changes in gene cycling but subtle difference in core clock gene expression, we investigated other possible molecular mechanisms by which KD influences the diurnal transcriptional landscape. Therefore, we analyzed the expression profiles of



#	Liver only	Both conditions	Gut only
1	Rela	Bhlhe40	Brp44l
2	Elf1	Srebf1	Stat4
3	Nr1h4	Tef	Klf12
4	Bach1	Thra	Maf
5	Nr1h2	Tmem49	Atf4
6	Pou2f1		Ppara
7	Nr1i3		Rora
8	Crem		Pparg
9	Mitf		Srf
10	Stat5a		Nr6a1

Figure 2. Metabolite Levels and Circadian Transcriptional Signature in Liver versus IECs upon KD

(A) Hepatic total cholesterol levels at ZT0 and ZT12 in CC- and KD-fed mice (n = 4 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05; ZT within KD, #p < 0.05).

(B) Free fatty acid levels in the liver of CC- and KD-fed mice at ZT8 and ZT20 (n = 4 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05).

(C) Intestinal total cholesterol levels at ZT0 and ZT12 in CC- and KD-fed mice (n = 4 per time point, per group; two-way ANOVA, no significant difference).

(D) Free fatty acid levels in the ileum of CC- and KD-fed mice at ZT8 and ZT20 (n = 4 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05; ZT within KD, #p < 0.05). Error bars represent SEM.

(E) Venn diagram representing the transcripts oscillating in "liver only," "both tissues," and "IECs only" upon KD feeding (n = 3 per time point, per group, p < 0.01, FDR < 0.1).

(F) Heatmaps representing the genes diurnal in both tissues, liver only, and intestine only upon KD (n = 3 per time point, per group, p < 0.01, FDR < 0.1).

(G) Analysis of the rhythmic transcription factors (p < 0.01) in animals fed KD, exclusively in the liver, in both tissues, or exclusively in the gut.

(H) Transcription factor binding site (TFBS) analysis in the liver of KD-fed mice. In the graph is reported the –log (pValue, 10). The first four sites are enriched exclusively in the liver upon KD. CLOCK:BMAL1 binding site is not exclusive, although it is highly enriched in KD liver, as shown by the value of the –log (pValue).

(I) Transcription factor binding site (TFBS) analysis in the IECs of KD-fed mice. In the graph is reported the –log (pValue, 10).

CCGs other than the core clock that, as previously mentioned, did not display any major alterations both in the liver and the gut (Figures S3C and S3D).

In the liver, KD induced a significant increase in the amplitude of CCG oscillation, such as *Dbp* (albumin D-box binding protein) and *Nampt* (nicotinamide phosphoribosyl transferase) (Figure 3A). The

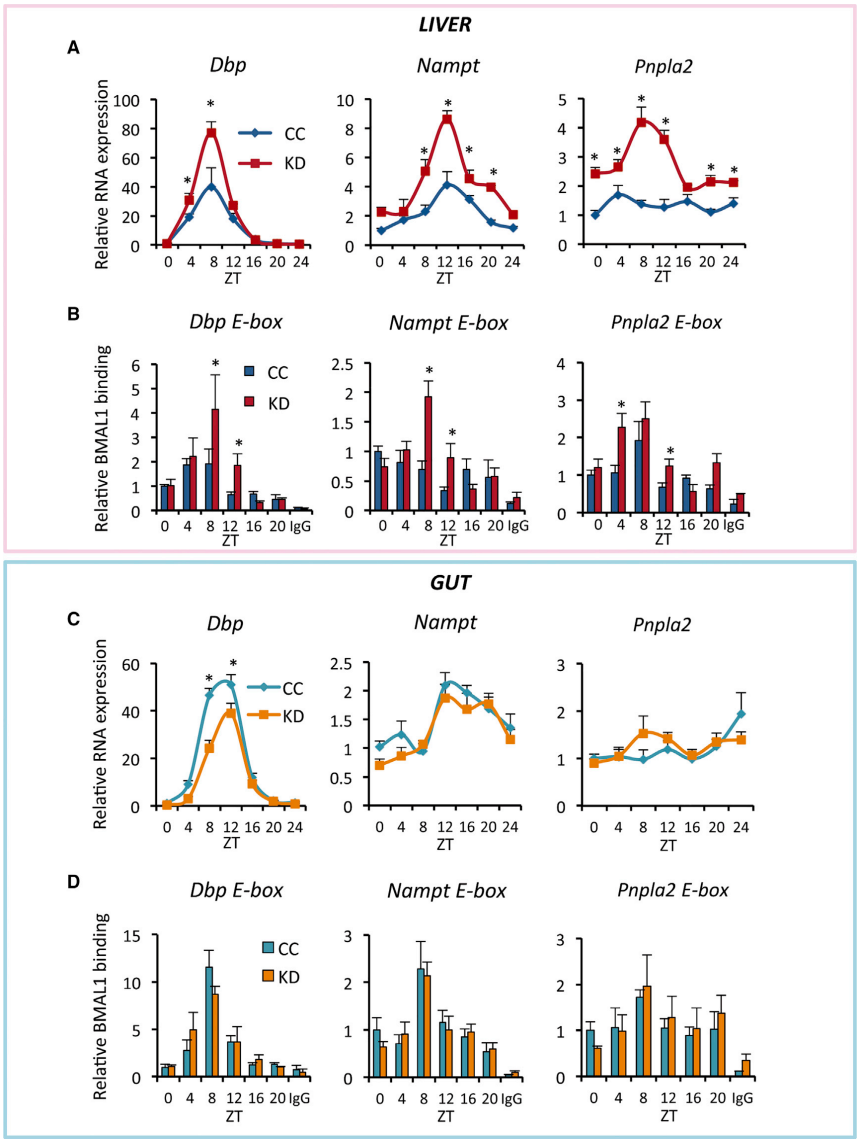


Figure 3. Clock-Controlled Gene Expression and BMAL1 Chromatin Recruitment on Specific Target Promoters

(A) Expression of clock output genes *Dbp*, *Nampt*, and *Pnpla2* measured by qPCR in the liver (n = 5 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05).

(B) qPCR- ChIP results showing BMAL1 chromatin recruitment on the E-boxes of the *Dbp*, *Nampt*, and *Pnpla2* gene promoters in the liver (n = 5 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05).

(C) Expression of clock output genes *Dbp*, *Nampt*, and *Pnpla2* measured by qPCR in IECs (n = 5 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05).

(legend continued on next page)

example of *Nampt* is notable since, in contrast to KD, its circadian expression is completely abolished by high-fat diet in the liver (Eckel-Mahan et al., 2013). The profile of *Pnpla2* (patatin-like phospholipase domain containing 2, also known as *Atgl*), a rhythmic gene involved in lipid metabolism, was also altered (Figure 3A), displaying a significant amplitude enhancement in the liver (Figure 3A). In the gut, the effect of KD on these genes was drastically different. In fact, the expression of *Dbp*, *Nampt*, and *Pnpla2* was virtually identical to the profile under CC (Figure 3C). Importantly, there was no increase of *Nampt* and *Pnpla2* gene expression at ZT12 upon KD in *Clock*-deficient mice, both in liver and intestine (Figures S3G and S3H). This result indicates that a functional clock is required to mediate the effect of KD on hepatic rhythmic transcriptional changes.

To decipher the molecular mechanism responsible for the enhancement in CCG amplitude, we performed chromatin immunoprecipitation (ChIP) experiments using BMAL1 antibodies. Strikingly, BMAL1 recruitment on the E-boxes within *Dbp* and *Nampt* promoter regions was significantly increased at ZT8 and ZT12, and at ZT4 and ZT12 on *Pnpla2* E-box in the liver of KD-fed animals (Figure 3B). In contrast, BMAL1 binding to the E-boxes of the same genes showed no differences in the IECs (Figure 3D). In line with these results, cross-analysis of our diurnal transcriptome with BMAL1 ChIP sequencing (ChIP-seq) data (Koike et al., 2012; Rey et al., 2011) indicated that the majority of cyclic BMAL1 targets in the liver of KD-fed mice had a robust peak at ZT8 (Figure S4A) and an increase in their expression at the same ZT with respect to CC-fed mice (Figure S4E). Similarly, other clock output genes, such as *Nmnat3* and *Bhlhe41*, displayed a significant enhancement in their oscillation with a zenith at ZT8 (Figures S4G and S4H), accompanied by a significant increase in BMAL1 chromatin recruitment on their promoters upon KD feeding (Figures S4G and S4H). In contrast, no changes in BMAL1 binding were observed at the E-boxes on *Per2* and *Rev-erbα* promoters (Figure S4I), perfectly correlating with no major alterations in core clock gene transcription in KD liver (Figure S3C). On the other hand, BMAL1 target genes displayed no increase in their level under KD in the gut (Figures S4B and S4F). Moreover, the phase-lag analysis of BMAL1 target cyclic genes showed a different pattern in the peak and nadir between liver and gut both for CC- or KD-fed animals (Figures S4C and S4D). Thus, while the global levels of BMAL1 protein seem to be unaffected by KD feeding, BMAL1 chromatin recruitment to specific targets appears to be a critical control step in the liver of KD-fed mice. It is worth noting that these results are in keeping with the TFBS analysis, in which the "CLOCK:BMAL1" motif was highly enriched in the rhythmic liver transcriptome of KD-fed mice (Figure 2H; Table S6, sheet 4).

Taken together, these data show that KD modulates specific CCG oscillation by influencing the chromatin recruitment of the endogenous clock complex in the liver, but not in the gut. Thus, our findings underlie a tissue-dependent effect on clock protein functions upon a nutritional challenge.

Cyclic Activation of PPARα by KD

A central pathway highly enriched in our high-throughput genomic analysis is "PPAR signaling," clustered in the class of genes oscillating in "KD-only" in both liver and intestine (Figures 1C and 1G). It is worth noting that this GO category was present also in "CC-only" cycling genes in the gut, although the significance of the cluster in "CC-only" is much lower than in "KD-only" (Figure S2B). Moreover, the PPAR binding motif was significantly enriched in KD-only IECs (Figure 2I). Importantly, PPARα-driven transcription is implicated in the metabolic response to energy deprivation (Kersten et al., 1999). Furthermore, PPARα is a key regulator of lipid metabolic homeostasis and a prime mediator of ketogenesis (Badman et al., 2007; Leone et al., 1999). Thus, we examined the possible role of PPARα in the KD-induced cyclic transcriptional reprogramming in liver and gut. Importantly, PPARα nuclear protein levels were robustly upregulated upon KD in both liver and gut (Figures 4A and 4D). Daily oscillations of nuclear PPARα protein expression in KD-fed animals were anti-phasic in the two tissues, with a peak at around ZT8 in the liver and a trough at the same ZT in the gut (Figures 4A and 4D). Remarkably, PPARα target genes in the liver (Figure 4B) had a distinct profile with respect to the cyclic profile of the corresponding genes in the gut (Figure 4E), following the nuclear accumulation of PPARα in the two tissues (Figures 4A and 4D). Notably, PPARα targets *Hmgcs2* (3-hydroxy-3-methylglutaryl-CoA synthase 2), the rate limiting enzyme of the ketogenic pathway; *Acot2* (acyl-CoA thioesterase 2); *Angptl4* (angiopoietin-like 4); and *Cpt1a* (carnitine palmitoyltransferase 1A) robustly oscillate in IECs (Figure 4E) in phase with PPARα nuclear protein abundance (Figure 4D) and following the mouse feeding rhythm. It is worth noting that *Hmgcs2* exhibited a robust gain in oscillation in IECs, showing an increase of more than 25-fold compared to CC at the peak time point (Figure 4E). On the other hand, not all the liver PPARα targets displayed a clear rhythmic oscillation (Figure 4B). For example, hepatic *Hmgcs2* was only moderately increased during the daily ZTs, despite the significant upregulation in βOHB in KD-fed mice (Figure S1B). It is conceivable that HMGCS2 activity may be mainly regulated by post-translational modifications, in keeping with previously reported findings (Rardin et al., 2013; Shimazu et al., 2010). *Acot2* was significantly upregulated in KD-fed mice at all times of the circadian cycle, although it did not present a clear rhythmic behavior. Finally, hepatic *Cpt1a* and *Angptl4* exhibited a significant expression increase and gain in the amplitude of oscillation (Figure 4B).

To gain further insight into the cyclic activation of the PPARα pathway upon KD, we analyzed the daily expression of known PPARα target genes (Rakhshandehroo et al., 2010). The major biological pathways in which PPARα targets are involved are presented as a pie chart (Figure S5A). Strikingly, in the gut ~20% of these genes oscillated in phase with PPARα nuclear accumulation, as shown by the heatmap (Figure S5B, see the arrow). Thereby, it appears that the KD is imposing a stricter phase of oscillation for a number of genes in the gut. The majority of

(D) ChIP results showing BMAL1 chromatin recruitment on the E-boxes of *Dbp*, *Nampt*, and *Pnpla2* gene promoters in IECs (n = 5 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, no significant difference).
IgG represents ChIP experiment performed with an isotype-matched control immunoglobulin (normal rabbit IgG) to BMAL1.
Error bars represent SEM.

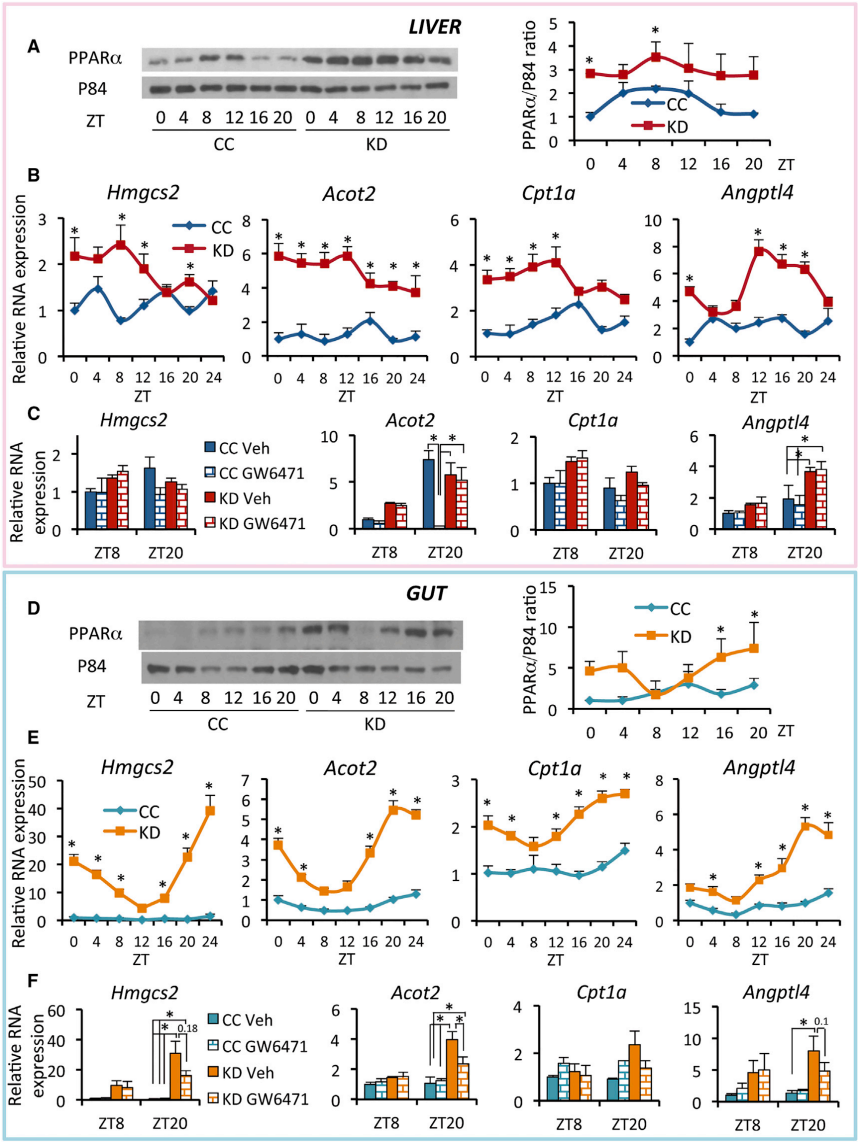


Figure 4. Tissue-Specific Circadian PPARα Signature in Liver and Intestine upon KD

(A) Western blot analysis of PPARα in liver nuclear extract. On the left, representative blot; on the right, graph reporting the average of the blot band density (n = 3 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05).

(B) Expression of the PPARα target genes *Hmgcs2*, *Acot2*, *Cpt1a*, and *Angptl4* in liver extract (n = 5 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05).

(C) Liver PPARα target gene expression in mice fed CC or KD for 4 weeks and treated with the specific PPARα inhibitor GW6471 (n = 3–4 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05).

(legend continued on next page)

these targets participated in specific metabolic pathways such as "mitochondrial β-oxidation/oxidative phosphorylation" and "peroxisomal β-oxidation." In contrast, the same genes in the liver were not rhythmic and their profile was completely different with respect to the one in IECs (Figure S5C), as already indicated by the qPCR data (Figure 4B). Importantly, treatment of CC- or KD-fed mice with the specific PPARα inhibitor GW6471 influenced PPARα target gene levels at ZT20, the peak time point in KD-fed mice, in the ileum (Figure 4F), but not in the liver (Figure 4C).

Therefore, our results show that KD activates a tissue-specific rhythmic PPARα-dependent transcriptional reprogramming characterized by a well-defined phase of oscillation exclusively in the intestine, which differs from the liver.

KD Induces Circadian βOHB Levels and Rhythmic Histone Acetylation in the Gut

To determine the molecular mechanisms contributing to the different diurnal expression of PPARα targets in liver and intestine, we monitored the abundance of acetyl-(Lys9/Lys14) on histone H3 at PPRE (PPAR response element) sites on target promoters. The levels of acetylated H3 followed the daily transcriptional changes in IECs of KD-fed animals, with a higher increase at ZT20 than at ZT8 for all tested promoters (*Hmgcs2-PPRE*, *Acot2-PPRE*, *Cpt1a-PPRE*, and *Angptl4-PPRE*; Figure 5B). In keeping with PPARα targets profile, no significant changes in the rhythmicity of acetyl-(Lys9/Lys14) could be observed in the liver (Figure 5A). Strikingly, there was a robust oscillation of βOHB serum levels along the diurnal cycle in KD-fed mice, peaking at ZT0–ZT20 and reaching nadir at ZT8 (Figure 5C), with a diurnal profile mirroring the one of PPARα nuclear accumulation and PPARα target expression in the intestine (Figures 4D and 4E). Indeed, cyclic expression of ileal *Hmgcs2* may partially be responsible for the daily profile of βOHB levels. As βOHB is an endogenous histone deacetylase inhibitor (Shimazu et al., 2013), we reasoned that cyclic βOHB might generate rhythmicity in HDAC activity contributing to de novo oscillation of PPARα target genes. To explore this possibility, we analyzed HDAC activity in nuclear extract from intestine during daytime (ZT8) and nighttime (ZT20). We found that HDAC activity displayed an opposite profile with respect to serum βOHB oscillation in KD gut, being higher at ZT8 and lower at ZT20 (Figure 5G). Moreover, HDAC activity was significantly decreased in the gut of KD-fed mice as compared to CC-fed mice at ZT20 (Figure 5G), in line with the increase in histone H3 acetylation at specific promoter regions (Figure 5B). HDAC activity in nuclear extracts from CC-fed mice was virtually identical at ZT8 and ZT20, in keeping with the lack of βOHB rhythmicity (Figure 5C) and H3 acetyl-(Lys9/Lys14) in CC-fed mice (Figure 5B). Notably, the local concentration of βOHB in KD intestine remarkably mirrored the diurnal profile of the same metabolite in the serum (Figure 5E), strongly supporting the inhibitory effect on HDAC activity and the time-of-the-day-dependent changes in histone acetylation in this specific tissue. On the other hand, βOHB levels in the liver of KD-fed animals were still increased, although displaying a different oscillatory pattern with a peak at ZT12–ZT16 (Figure 5D). Furthermore, hepatic HDAC activity did not show any significant difference between CC and KD and only a negligible day-night difference (Figure 5F). To rule out the possibility that HDAC inhibition may occur at a diverse ZT in the liver, we monitored the abundance of acetyl-(Lys9/Lys14) on histone H3 on the same promoter regions (*Hmgcs2-PPRE*, *Acot2-PPRE*, *Cpt1a-PPRE*, and *Angptl4-PPRE*) at ZT0 and ZT12, the nadir and zenith, respectively, of the local hepatic βOHB. No changes were observed on H3 acetylation between CC and KD conditions or between the two ZTs analyzed (Figure S4J).

To demonstrate the direct link between βOHB, gene expression, and changes in histone acetylation in the gut, mice were fed CC or a diet corresponding to CC containing 10% (w/w) 1,3-butanediol (1,3 butanediol diet, BD) for 4 weeks. This is a strategy to increase the endogenous βOHB levels without inducing the metabolic state characteristic of a KD (Hashim and VanItallie, 2014). 1,3-butanediol is an alcohol precursor of βOHB and it is quickly converted to βOHB in the liver by the alcohol dehydrogenase system (Veech, 2014). As expected, βOHB serum concentration was significantly increased both at ZT8 and ZT20 upon BD (Figure S6A). PPARα target gene expression in the liver did not show any substantial difference after 4 weeks of BD (Figure S6B), as predicted because 1,3-butanediol does not involve any aspect of fatty acid metabolism. In keeping with this observation, only marginal changes in H3 acetyl-(Lys9/Lys14) on the promoter of the same genes were detected (Figure S6C). Notably, a significant upregulation in PPARα target genes was present in the intestine of BD-fed mice (Figure S6D). Interestingly, H3 acetyl-(Lys9/Lys14) abundance on the promoter region of *Acot2*, *Cpt1a*, and *Angptl4* was higher upon BD (Figure S6E), paralleling gene expression and reinforcing our hypothesis that βOHB specifically impacts this type of histone post-translational modification in the gut. Taken together, our results unveil a novel and tissue-specific epigenetic role of βOHB in rewiring cyclic gene expression upon a nutritional challenge.

Sucrose and Fructose Complementation of KD

As KD is devoid of carbohydrates, we questioned whether complementation with either fructose or sucrose would influence the metabolic effect of KD on the diurnal program. Carbohydrates have been shown to inhibit the induction of ketogenesis (Fukao et al., 2004). To do so, KD- or CC-fed mice were subjected to oral gavage with fructose (4 g/kg of body weight) for 7 days once a day, during the last week of their dietary regimen. Tissues were collected at ZT0 and ZT12, respectively the end

(D) Nuclear PPARα protein accumulation in IECs. On the left, representative blot; on the right, graph reporting the average of the blot band density (n = 3 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05).
(E) PPARα target gene expression of *Hmgcs2*, *Acot2*, *Cpt1a*, and *Angptl4* in IECs (n = 5 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05).
(F) Ileum PPARα target gene expression in mice fed CC or KD for 4 weeks and treated with the specific PPARα inhibitor GW6471 (n = 3–4 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05).
Error bars represent SEM.

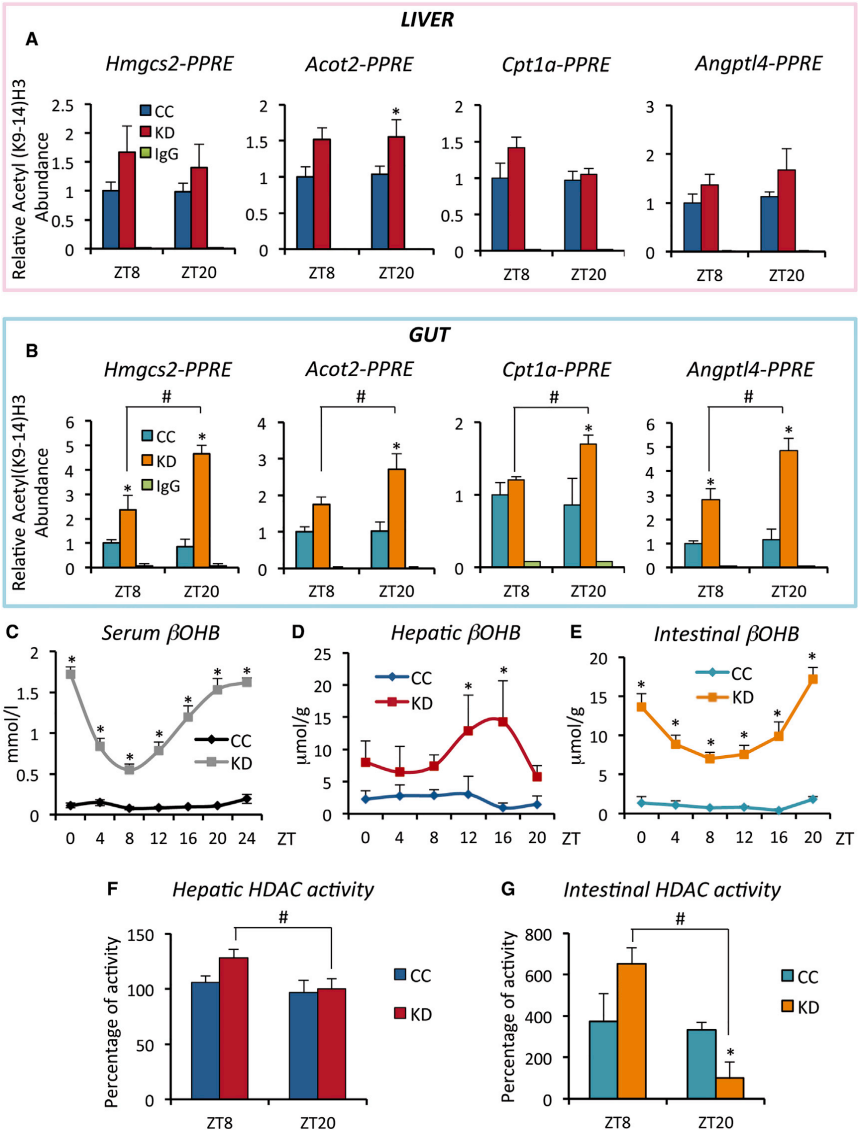


Figure 5. KD-Driven Oscillation in Serum βOHB and Daily Changes in Epigenetic Marks

(A) Liver ChIP of acetylated (Acetyl) lysine (K) 9-14 on histone H3. qPCR showing the relative abundance of H3 Acetyl(K9-K14) on the PPAR-responsive element (PPRE) of specific promoter: *Hmgcs2*, *Acot2*, *Cpt1a*, and *Angptl4* (n = 4–6 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05).

(B) H3 Acetyl(K9-K14) abundance on the PPRE of *Hmgcs2*, *Acot2*, *Cpt1a*, and *Angptl4* promoter in the intestine (n = 4 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05; ZT within KD, #p < 0.05).

(C) Serum βOHB levels along the diurnal cycle in CC- and KD-fed mice (n = 5 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05).

(legend continued on next page)

and the beginning of the active phase. We focused our attention on a set of genes involved in KD metabolic response both in hepatic and ileal tissues. No significant effect of fructose was noted in the liver of KD-fed mice on *Hmgcs2*, *Acot2*, and *Angptl4* gene expression (Figure 6A). In sharp contrast, fructose induced a significant reduction in the expression of all these transcripts in the intestine (Figure 6C). Moreover, core clock genes and CCGs were only marginally influenced by fructose supplementation both in the liver (Figure S7A) and in the gut (Figure S7C). Thus, clock effects in the liver may be due to altered upstream signaling pathways that are selectively recruited over time to adapt to the change in fuel availability and drive downstream circadian events in a tissue-specific manner.

To test the effect of sucrose administration on KD-induced clock response, CC- or KD-fed mice were treated with 30% sucrose in drinking water for all 4 weeks of diet. Concomitant sucrose administration was insufficient to alter KD-specific clock response in the liver as only the expression of *Acot2* was moderately affected (Figure 6B). In contrast, the intestinal clock was robustly responsive to sucrose administration with an almost complete reversal of KD-induced metabolic changes. Indeed, gene expression of all the tested genes (*Hmgcs2*, *Acot2*, and *Angptl4*) returned to the levels distinctive of CC-fed mice (Figure 6D). Interestingly, sucrose did not interfere with core clock genes and CCGs in the liver of KD-fed mice, although it significantly influenced gene expression in CC-fed mice (Figure S7B), suggesting that KD could reinforce the core clock robustness in the liver. Similarly, sucrose affected core clock gene expression in the gut of CC-fed mice, although it also impacted CCG levels upon KD (Figure S7D).

Thus, in contrast to the liver, the intestinal clock exhibits a rapid response to carbohydrates that leads to increased plasticity in transcriptional reprogramming, further underlying the distinct and tissue-specific response of peripheral clocks upon dietary perturbation.

DISCUSSION

Metabolic homeostasis is connected to circadian function and both central and peripheral clocks contribute to its maintenance. The liver clock has been extensively studied because of its role in metabolic regulation. The intestinal clock, although obviously of importance, has been less investigated despite its central involvement in sustaining organismal metabolic responses (Mukherji et al., 2013). For the first time, we have questioned how two different tissues would operate their intrinsic, specialized plasticity to respond to a distinct nutritional challenge. We reveal a tissue-specific cyclic signature in response to a KD regimen that leads to a dramatic increase in β-oxidation and consequent ketosis (Paoli et al., 2015). While the impact of KD on the circadian clock has been noted (Genzer et al., 2015; Oishi et al., 2009), the molecular mechanisms underlying its action have not been satisfactorily explored. Our study represents the first exhaustive analysis of how KD influences rhythmic genomic reprogramming. In addition, as KD has been extensively used for the treatment of refractory epilepsy in children and for many other applications in adults such as obesity, diabetes, neurodegenerative disorders, and cancer (Paoli et al., 2013), our findings may provide a framework for future explorations of how circadian control could contribute to these pathophysiological conditions.

One of the most interesting properties of KD is its ability to drive metabolic pathways generally induced by fasting or caloric restriction. In fact, gluconeogenesis, fatty acid oxidation, and ketogenesis are upregulated while glycolysis and de novo lipogenesis are shut down (Kennedy et al., 2007; Paoli et al., 2015). Interestingly, enzymes and transcription factors involved in these metabolic cascades display circadian rhythmicity or start to oscillate upon KD, underscoring the tight link with the circadian clock. Our results show that KD is able to modulate the clock machinery recruitment to chromatin, which led to gain in oscillation of CCGs in the liver. Despite a very high fat content in KD, this effect was the opposite of a high-fat diet, which hindered BMAL1 chromatin recruitment to target genes (Eckel-Mahan et al., 2013), opening new questions about how nutrient composition can modulate the core clock-chromatin interaction. Here we report a significant increase in BMAL1 binding on *Nampt* and *Dbp* promoters both at ZT8 and ZT12, and at ZT4 and ZT12 on *Pnpla2* promoter, that results in a robust increase in their amplitude. Intriguingly, crossing our diurnal microarray data with BMAL1 ChIP-seq unveiled that BMAL1 target genes cycling in KD liver displayed a robust peak at ZT8. Taken together, these data suggest that KD-driven BMAL1 chromatin recruitment participates in the regulation of systemic adaptive responses to KD. On the other hand, the gut core clock was not affected by KD, indicating that the core clock machinery contributes to the physiological adaptation to a KD challenge in a tissue-specific manner, crosstalking directly with the metabolic clock exclusively in the liver.

The robust PPARα signaling activation induced in the gut of KD-fed mice paralleled the mouse food intake, suggesting that daily changes in PPARα signaling were primarily mediated by local physiological responses to feeding behavior rather than alterations in the core clock machinery. Indeed, the intestinal local levels of free fatty acids were remarkably increased at ZT20 with respect to ZT8 upon KD, respectively the peak and trough of PPARα nuclear concentration and its target gene expression (Figures 4D and 4E). Intriguingly, although PPARα pathway was induced both in liver and gut during ketogenesis, the two tissues displayed different phase of oscillation in both PPARα

(D) Hepatic βOHB levels along the diurnal cycle in CC- and KD-fed mice (n = 3 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05).
(E) Intestinal βOHB levels along the diurnal cycle in CC- and KD-fed mice (n = 3 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05).
(F) HDAC activity from liver nuclear extract at ZT8 and ZT20 (n = 5 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, no significant difference; ZT within KD, #p < 0.05).
(G) HDAC activity in intestinal nuclear extract at ZT8 and ZT20 (n = 3–5 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05; ZT within KD, #p < 0.05).
Error bars represent SEM.

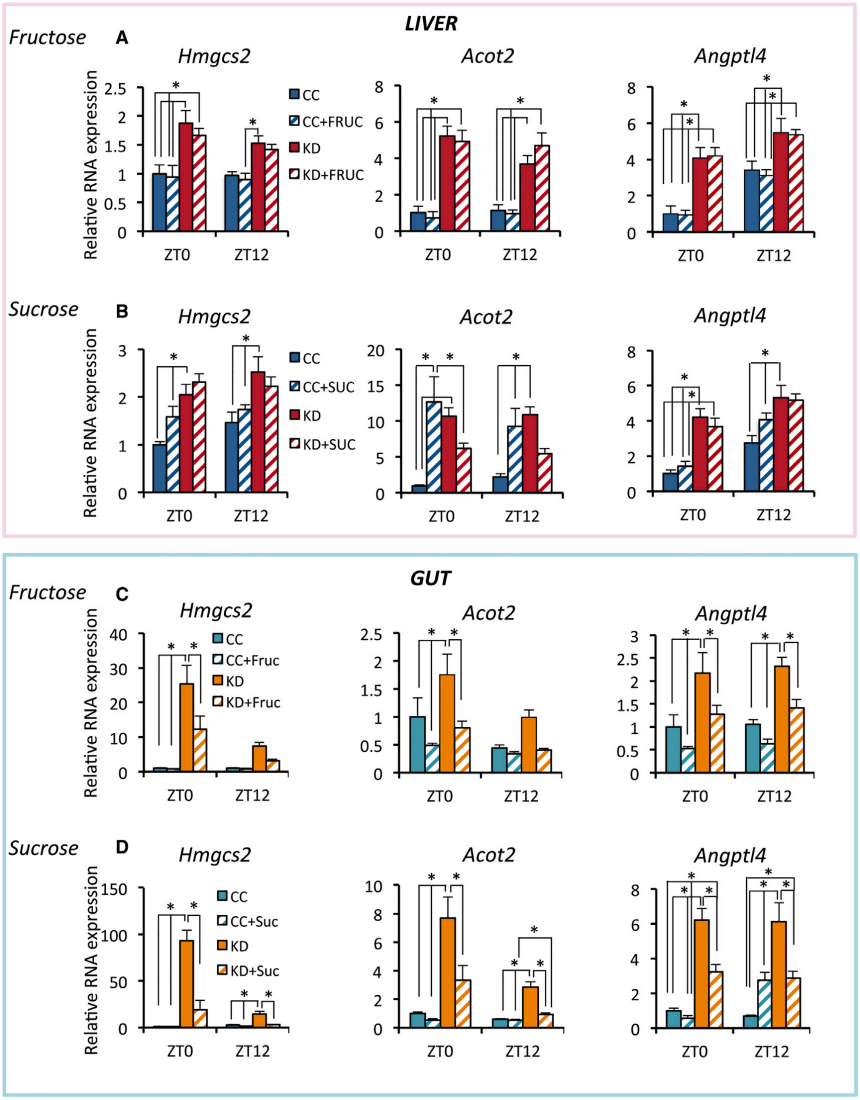


Figure 6. Liver and Intestine Response to Perturbation in Diet Composition

(A) qPCR of genes involved in the ketogenic response, *Hmgcs2*, *Acot2*, and *Angptl4*, in the liver of mice fed 4-week KD or CC and administered with fructose oral gavage (4 g/Kg) once a day, per 7 days during the last week of dietary regimen (n = 5–6 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05).

(B) *Hmgcs2*, *Acot2*, and *Angptl4* gene expression in the liver of animals fed 4-week KD or CC and simultaneously treated with 30% sucrose in the drinking water (n = 5 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05).

(C) Intestinal gene expression of *Hmgcs2*, *Acot2*, and *Angptl4* in mice subjected to fructose oral gavage (n = 5–6 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05).

(D) qPCR analysis of *Hmgcs2*, *Acot2*, and *Angptl4* genes in the intestine of animals treated with 30% sucrose in the drinking water (n = 5 per time point, per group; two-way ANOVA, Holm-Sidak post hoc; comparisons for factor, diet within ZT, *p < 0.05).

Error bars represent SEM.

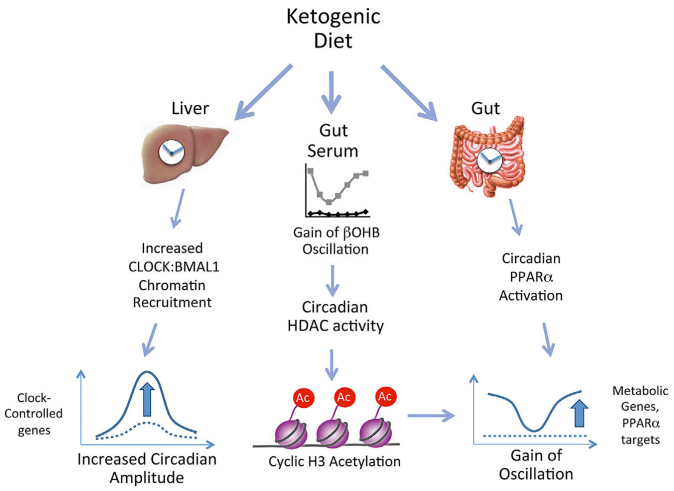


Figure 7. Schematic Representation of KD-Dependent Diurnal Signature in Liver and Intestine

A very-low-carbohydrate KD is able to reprogram peripheral clocks in liver and intestine through the activation of distinct transcription factors in a diurnal manner. KD enhances BMAL1 chromatin recruitment in the liver, resulting in an increase in the amplitude of oscillation of CCGs. KD activates the PPARα pathway both in liver and intestine but is highly diurnal only in the gut. The activation of ketogenesis increases serum βOHB concentration that oscillates in a diurnal fashion. Through circulation, βOHB is distributed to all the tissues as an energy source. The effect of βOHB is primarily evident in the intestine, where the metabolite oscillates with the same phase of serum βOHB, and it leads to regulated HDAC activity in a time-of-the-day-dependent fashion. This results in the increase of histone H3 acetylation on metabolic gene promoters at specific ZTs, contributing to rhythmic gene expression. Indeed, PPARα targets display a unique diurnal profile that nicely parallels PPARα nuclear accumulation and serum/intestinal βOHB daily changes.

nuclear accumulation and target gene expression. Finally, the pharmacological inhibition of PPARα during KD significantly influenced its target gene expression in the gut, but not in the liver. This evidence points toward unique circadian properties of metabolic clocks upon a food stress, even when the same molecular player, such as PPARα, is implicated. Furthermore the liver and intestine clock exhibited different sensitivity to perturbations in the diet macronutrient. Indeed, the hepatic clock was refractory to carbohydrate supplementation under KD, maintaining almost unvaried gene expression. This is in sharp contrast to the intestinal clock that is highly plastic to either sucrose or fructose complementation, as visualized by the changes in gene expression as compared to KD (Figures 6C and 6D). Thus, the sole KD complementation by either sucrose or fructose appears to restore the metabolic clock in the gut to the CC scenario, stressing the specificity in the response of distinct peripheral clocks to food challenges.

An interesting facet of this study relates to the cyclic profile of βOHB upon KD feeding. From a clinical point of view, this notion could be relevant since KD has been broadly used for the treatment of epilepsy (van der Louw et al., 2016). Indeed, circadian control might contribute to the intricate and still poorly explored relationship between seizure control and serum levels of ketone bodies (Gilbert et al., 2000; McNally and Hartman, 2012). Moreover, KD seems to be beneficial as an adjuvant for cancer therapy (Allen et al., 2014; Wright and Simone, 2016), and in obesity, diabetes, and some inherited metabolic disorders (Mobbs et al., 2013; Scholl-Bürgi et al., 2015). Remarkably, βOHB has been demonstrated to be protective in a variety of conditions, such as neurodegenerative diseases (Kashiwaya et al., 2000; Lim et al., 2011), in support of a view in which ketone bodies may have multiple functions in addition to energy source (Puchalska and Crawford, 2017). Our findings support a role of βOHB as an epigenetic regulator in keeping with previous studies (Shimazu et al., 2013; Sleiman et al., 2016). Intriguingly, the rhythmic nature of βOHB in the serum and locally in the gut upon KD is translated into a time-of-the-day-dependent modulation of HDAC activity specifically in the intestine, resulting in differences in the level of histone H3 acetylation and paralleled circadian changes in gene expression. The oscillatory levels of βOHB appear to drive effects on histone acetylation in a tissue-specific manner, being robust in the gut and almost absent in the liver. While the reasons for this intriguing difference are yet to be explored, our data reveal new avenues of future directions to better understand the impact of βOHB, and possibly other metabolites, on the epigenetic landscape and gene expression control of different peripheral clocks.

Finally, our study brings further evidence to the link between chromatin remodeling, diurnal rhythms, and cell energy metabolism. Additional experiments will determine whether KD and the subsequent βOHB oscillation could influence the levels and/or rhythmicity of histone β-hydroxyl-butyrylation, a recently described epigenetic mark involved in metabolic regulation of transcription (Xie et al., 2016).

Our work contributes to the concept that metabolic cues can influence chromatin landscape and gene expression (Katada et al., 2012; Martinez-Pastor et al., 2013) and points to the ability of the clock system to interpret changes in the nutritional regimen in a tissue-specific fashion, leading to distinct rhythmic signatures that designate the physiology and function of the tissue analyzed (Figure 7). It is noteworthy that while the genes and enzymes studied here are expressed in both the ileum and the liver, the clock has the unique ability to recruit them to carry out precise temporal roles tailored for tissue-specific response.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.cmet.2017.08.015.

AUTHOR CONTRIBUTIONS

P.T. and M.M. performed all the experiment. K.L.E.-M. helped in collecting samples and in interpreting the data. Y.L. and P.B. performed the bioinformatic analyses. J.C.N. and E.V. helped in data discussion and project design. P.T., M.M., and P.S.-C. designed the study, prepared the figures, and wrote the manuscript.

ACKNOWLEDGMENTS

We thank all the members of Sassone-Corsi laboratory for constructive comments and Selma Masri for critical discussions. We thank Melanie Oakes, Seung-Ah Chung, and Valentina Ciobanu of the UCI Genomics High-Throughput Facility. We are grateful to Haig Derderian for help and Qin Yang for access to his laboratory's instrumentation. This work was supported by NIH grant DA036408, the Institute National de la Sante et de la Recherche Medicale (INSERM, France) INSERM-44790 and the Novo Nordisk Foundation Challenge Grant NNF-202585 (P.S.-C.), and the Japan Foundation for Pediatric Research 13-003 (M.M.). M.M. was supported by the Sumitomo Life Welfare and Culture Foundation and P.T. was supported by EMBO ALTF 1510-2012 and Human Frontiers Science Program LT 000576/2013.

Received: October 27, 2016
Revised: May 4, 2017
Accepted: August 15, 2017
Published: September 5, 2017

REFERENCES

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*, 3051.

Allen, B.G., Bhatia, S.K., Anderson, C.M., Eichenberger-Gilmore, J.M., Sibenaller, Z.A., Mapuskar, K.A., Schoenfeld, J.D., Buatti, J.M., Spitz, D.R., and Fath, M.A. (2014). Ketogenic diets as an adjuvant cancer therapy: history and potential mechanism. Redox Biol. *2*, 963–970.

Asher, G., and Sassone-Corsi, P. (2015). Time for food: the intimate interplay between nutrition, metabolism, and the circadian clock. Cell *161*, 84–92.

Asher, G., Gatfield, D., Stratmann, M., Reinke, H., Dibner, C., Kreppel, F., Mostoslavsky, R., Alt, F.W., and Schibler, U. (2008). SIRT1 regulates circadian clock gene expression through PER2 deacetylation. Cell *134*, 317–328.

Badman, M.K., Pissios, P., Kennedy, A.R., Koukos, G., Flier, J.S., and Maratos-Flier, E. (2007). Hepatic fibroblast growth factor 21 is regulated by PPARalpha and is a key mediator of hepatic lipid metabolism in ketotic states. Cell Metab. *5*, 426–437.

Baldi, P., and Long, A.D. (2001). A Bayesian framework for the analysis of microarray expression data: regularized t -test and statistical inferences of gene changes. Bioinformatics *17*, 509–519.

Bass, J., and Takahashi, J.S. (2010). Circadian integration of metabolism and energetics. Science *330*, 1349–1354.

Bellet, M.M., Deriu, E., Liu, J.Z., Grimaldi, B., Blaschitz, C., Zeller, M., Edwards, R.A., Sahar, S., Dandekar, S., Baldi, P., et al. (2013). Circadian clock regulates the host response to Salmonella. Proc. Natl. Acad. Sci. USA *110*, 9897–9902.

Chavan, R., Feillet, C., Costa, S.S., Delorme, J.E., Okabe, T., Ripperger, J.A., and Albrecht, U. (2016). Liver-derived ketone bodies are necessary for food anticipation. Nat. Commun. *7*, 10580.

Daily, K., Patel, V.R., Rigor, P., Xie, X., and Baldi, P. (2011). MotifMap: integrative genome-wide maps of regulatory motif sites for model species. BMC Bioinformatics *12*, 495.

Debruyne, J.P., Noton, E., Lambert, C.M., Maywood, E.S., Weaver, D.R., and Reppert, S.M. (2006). A clock shock: mouse CLOCK is not required for circadian oscillator function. Neuron *50*, 465–477.

Douris, N., Melman, T., Pecherer, J.M., Pissios, P., Flier, J.S., Cantley, L.C., Locasale, J.W., and Maratos-Flier, E. (2015). Adaptive changes in amino acid metabolism permit normal longevity in mice consuming a low-carbohydrate ketogenic diet. Biochim. Biophys. Acta *1852* (10 Pt A), 2056–2065.

Eckel-Mahan, K., and Sassone-Corsi, P. (2013). Metabolism and the circadian clock converge. Physiol. Rev. *93*, 107–135.

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. USA *109*, 5541–5546.

Eckel-Mahan, K.L., Patel, V.R., de Mateo, S., Orozco-Solis, R., Ceglia, N.J., Sahar, S., Dilag-Penilla, S.A., Dyar, K.A., Baldi, P., and Sassone-Corsi, P. (2013). Reprogramming of the circadian clock by nutritional challenge. Cell *155*, 1464–1478.

Everett, L.J., and Lazar, M.A. (2014). Nuclear receptor Rev-erbα: up, down, and all around. Trends Endocrinol. Metab. *25*, 586–592.

Fukao, T., Lopaschuk, G.D., and Mitchell, G.A. (2004). Pathways and control of ketone body metabolism: on the fringe of lipid biochemistry. Prostaglandins Leukot. Essent. Fatty Acids *70*, 243–251.

Genzer, Y., Dadon, M., Burg, C., Chapnik, N., and Froy, O. (2015). Ketogenic diet delays the phase of circadian rhythms and does not affect AMP-activated protein kinase (AMPK) in mouse liver. Mol. Cell. Endocrinol. *417*, 124–130.

Gilbert, D.L., Pyzik, P.L., and Freeman, J.M. (2000). The ketogenic diet: seizure control correlates better with serum beta-hydroxybutyrate than with urine ketones. J. Child Neurol. *15*, 787–790.

Grygiel-Górniak, B. (2014). Peroxisome proliferator-activated receptors and their ligands: nutritional and clinical implications–a review. Nutr. J. *13*, 17.

Hashim, S.A., and VanItallie, T.B. (2014). Ketone body therapy: from the ketogenic diet to the oral administration of ketone ester. J. Lipid Res. *55*, 1818–1826.

Hatori, M., Vollmers, C., Zarrinpar, A., DiTacchio, L., Bushong, E.A., Gill, S., Leblanc, M., Chaix, A., Joens, M., Fitzpatrick, J.A., et al. (2012). Time-restricted feeding without reducing caloric intake prevents metabolic diseases in mice fed a high-fat diet. Cell Metab. *15*, 848–860.

Hawley, J.A., Hargreaves, M., Joyner, M.J., and Zierath, J.R. (2014). Integrative biology of exercise. Cell *159*, 738–749.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

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.

Kashiwaya, Y., Takeshima, T., Mori, N., Nakashima, K., Clarke, K., and Veech, R.L. (2000). D-beta-hydroxybutyrate protects neurons in models of Alzheimer's and Parkinson's disease. Proc. Natl. Acad. Sci. USA *97*, 5440–5444.

Katada, S., Imhof, A., and Sassone-Corsi, P. (2012). Connecting threads: epigenetics and metabolism. Cell *148*, 24–28.

Kayala, M.A., and Baldi, P. (2012). Cyber-T web server: differential analysis of high-throughput data. Nucleic Acids Res. *40*, W553–W559.

Kennedy, A.R., Pissios, P., Otu, H., Roberson, R., Xue, B., Asakura, K., Furukawa, N., Marino, F.E., Liu, F.F., Kahn, B.B., et al. (2007). A high-fat, ketogenic diet induces a unique metabolic state in mice. Am. J. Physiol. Endocrinol. Metab. *292*, E1724–E1739.

Kersten, S., Seydoux, J., Peters, J.M., Gonzalez, F.J., Desvergne, B., and Wahli, W. (1999). Peroxisome proliferator-activated receptor alpha mediates the adaptive response to fasting. J. Clin. Invest. *103*, 1489–1498.

Kohsaka, A., Laposky, A.D., Ramsey, K.M., Estrada, C., Joshu, C., Kobayashi, Y., Turek, F.W., and Bass, J. (2007). High-fat diet disrupts behavioral and molecular circadian rhythms in mice. Cell Metab. *6*, 414–421.

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.

Leone, T.C., Weinheimer, C.J., and Kelly, D.P. (1999). A critical role for the peroxisome proliferator-activated receptor alpha (PPARalpha) in the cellular fasting response: the PPARalpha-null mouse as a model of fatty acid oxidation disorders. Proc. Natl. Acad. Sci. USA *96*, 7473–7478.

Leone, V., Gibbons, S.M., Martinez, K., Hutchison, A.L., Huang, E.Y., Cham, C.M., Pierre, J.F., Heneghan, A.F., Nadimpalli, A., Hubert, N., et al. (2015). Effects of diurnal variation of gut microbes and high-fat feeding on host circadian clock function and metabolism. Cell Host Microbe *17*, 681–689.

Lim, S., Chesser, A.S., Grima, J.C., Rappold, P.M., Blum, D., Przedborski, S., and Tieu, K. (2011). D-β-hydroxybutyrate is protective in mouse models of Huntington's disease. PLoS One *6*, e24620.

Martinez-Pastor, B., Cosentino, C., and Mostoslavsky, R. (2013). A tale of metabolites: the cross-talk between chromatin and energy metabolism. Cancer Discov. *3*, 497–501.

Masri, S., Rigor, P., Cervantes, M., Ceglia, N., Sebastian, C., Xiao, C., Roqueta-Rivera, M., Deng, C., Osborne, T.F., Mostoslavsky, R., et al. (2014). Partitioning circadian transcription by SIRT6 leads to segregated control of cellular metabolism. Cell *158*, 659–672.

McNally, M.A., and Hartman, A.L. (2012). Ketone bodies in epilepsy. J. Neurochem. *121*, 28–35.

Mobbs, C.V., Mastaitis, J., Isoda, F., and Poplawski, M. (2013). Treatment of diabetes and diabetic complications with a ketogenic diet. J. Child Neurol. *28*, 1009–1014.

Mohawk, J.A., Green, C.B., and Takahashi, J.S. (2012). Central and peripheral circadian clocks in mammals. Annu. Rev. Neurosci. *35*, 445–462.

Mukherji, A., Kobiita, A., Ye, T., and Chambon, P. (2013). Homeostasis in intestinal epithelium is orchestrated by the circadian clock and microbiota cues transduced by TLRs. Cell *153*, 812–827.

Murakami, M., Tognini, P., Liu, Y., Eckel-Mahan, K.L., Baldi, P., and Sassone-Corsi, P. (2016). Gut microbiota directs PPARγ-driven reprogramming of the liver circadian clock by nutritional challenge. EMBO Rep. *17*, 1292–1303.

Nakahata, Y., Kaluzova, M., Grimaldi, B., Sahar, S., Hirayama, J., Chen, D., Guarente, L.P., and Sassone-Corsi, P. (2008). The NAD+-dependent deacetylase SIRT1 modulates CLOCK-mediated chromatin remodeling and circadian control. Cell *134*, 329–340.

Nakahata, Y., Sahar, S., Astarita, G., Kaluzova, M., and Sassone-Corsi, P. (2009). Circadian control of the NAD+ salvage pathway by CLOCK-SIRT1. Science *324*, 654–657.

Newman, J.C., and Verdin, E. (2014). Ketone bodies as signaling metabolites. Trends Endocrinol. Metab. *25*, 42–52.

Oishi, K., Uchida, D., Ohkura, N., Doi, R., Ishida, N., Kadota, K., and Horie, S. (2009). Ketogenic diet disrupts the circadian clock and increases hypofibrinolytic risk by inducing expression of plasminogen activator inhibitor-1. Arterioscler. Thromb. Vasc. Biol. *29*, 1571–1577.

Okamura, H. (2004). Clock genes in cell clocks: roles, actions, and mysteries. J. Biol. Rhythms *19*, 388–399.

Paoli, A., Rubini, A., Volek, J.S., and Grimaldi, K.A. (2013). Beyond weight loss: a review of the therapeutic uses of very-low-carbohydrate (ketogenic) diets. Eur. J. Clin. Nutr. *67*, 789–796.

Paoli, A., Bosco, G., Camporesi, E.M., and Mangar, D. (2015). Ketosis, ketogenic diet and food intake control: a complex relationship. Front. Psychol. *6*, 27.

Partch, C.L., Green, C.B., and Takahashi, J.S. (2014). Molecular architecture of the mammalian circadian clock. Trends Cell Biol. *24*, 90–99.

Patel, V.R., Eckel-Mahan, K., Sassone-Corsi, P., and Baldi, P. (2012). CircadiOmics: integrating circadian genomics, transcriptomics, proteomics and metabolomics. Nat. Methods *9*, 772–773.

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.

Puchalska, P., and Crawford, P.A. (2017). Multi-dimensional roles of ketone bodies in fuel metabolism, signaling, and therapeutics. Cell Metab. *25*, 262–284.

Rakhshandehroo, M., Knoch, B., Müller, M., and Kersten, S. (2010). Peroxisome proliferator-activated receptor alpha target genes. PPAR Res. *2010*, http://dx.doi.org/10.1155/2010/612089.

Ramsey, K.M., Yoshino, J., Brace, C.S., Abrassart, D., Kobayashi, Y., Marcheva, B., Hong, H.K., Chong, J.L., Buhr, E.D., Lee, C., et al. (2009). Circadian clock feedback cycle through NAMPT-mediated NAD+ biosynthesis. Science *324*, 651–654.

Rardin, M.J., He, W., Nishida, Y., Newman, J.C., Carrico, C., Danielson, S.R., Guo, A., Gut, P., Sahu, A.K., Li, B., et al. (2013). SIRT5 regulates the mitochondrial lysine succinylome and metabolic networks. Cell Metab. *18*, 920–933.

Rey, G., Cesbron, F., Rougemont, J., Reinke, H., Brunner, M., and Naef, F. (2011). Genome-wide and phase-specific DNA-binding rhythms of BMAL1 control circadian output functions in mouse liver. PLoS Biol. *9*, e1000595.

Ripperger, J.A., and Schibler, U. (2006). Rhythmic CLOCK-BMAL1 binding to multiple E-box motifs drives circadian Dbp transcription and chromatin transitions. Nat. Genet. *38*, 369–374.

Schibler, U., and Sassone-Corsi, P. (2002). A web of circadian pacemakers. Cell *111*, 919–922.

Scholl-Bürgi, S., Höller, A., Pichler, K., Michel, M., Haberlandt, E., and Karall, D. (2015). Ketogenic diets in patients with inherited metabolic disorders. J. Inherit. Metab. Dis. *38*, 765–773.

Shimazu, T., Hirschey, M.D., Hua, L., Dittenhafer-Reed, K.E., Schwer, B., Lombard, D.B., Li, Y., Bunkenborg, J., Alt, F.W., Denu, J.M., et al. (2010). SIRT3 deacetylates mitochondrial 3-hydroxy-3-methylglutaryl CoA synthase 2 and regulates ketone body production. Cell Metab. *12*, 654–661.

Shimazu, T., Hirschey, M.D., Newman, J., He, W., Shirakawa, K., Le Moan, N., Grueter, C.A., Lim, H., Saunders, L.R., Stevens, R.D., et al. (2013). Suppression of oxidative stress by β-hydroxybutyrate, an endogenous histone deacetylase inhibitor. Science *339*, 211–214.

Sleiman, S.F., Henry, J., Al-Haddad, R., El Hayek, L., Abou Haidar, E., Stringer, T., Ulja, D., Karuppagounder, S.S., Holson, E.B., Ratan, R.R., et al. (2016). Exercise promotes the expression of brain derived neurotrophic factor (BDNF) through the action of the ketone body β-hydroxybutyrate. Elife *5*, http://dx.doi.org/10.7554/eLife.15092.

Stokkan, K.A., Yamazaki, S., Tei, H., Sakaki, Y., and Menaker, M. (2001). Entrainment of the circadian clock in the liver by feeding. Science *291*, 490–493.

Summa, K.C., Voigt, R.M., Forsyth, C.B., Shaikh, M., Cavanaugh, K., Tang, Y., Vitaterna, M.H., Song, S., Turek, F.W., and Keshavarzian, A. (2013). Disruption of the circadian clock in mice increases intestinal permeability and promotes alcohol-induced hepatic pathology and inflammation. PLoS One *8*, e67102.

Tamanini, F., Chaves, I., Bajek, M.I., and van der Horst, G.T. (2007). Structure function analysis of mammalian cryptochromes. Cold Spring Harb. Symp. Quant. Biol. *72*, 133–139.

Tognini, P., Thaiss, C.A., Elinav, E., and Sassone-Corsi, P. (2017). Circadian coordination of antimicrobial responses. Cell Host Microbe *22*, 185–192.

van der Louw, E., van den Hurk, D., Neal, E., Leiendecker, B., Fitzsimmon, G., Dority, L., Thompson, L., Marchió, M., Dudzińska, M., Dressler, A., et al. (2016). Ketogenic diet guidelines for infants with refractory epilepsy. Eur. J. Paediatr. Neurol. *20*, 798–809.

Veech, R.L. (2014). Ketone ester effects on metabolism and transcription. J. Lipid Res. *55*, 2004–2006.

Vollmers, C., Gill, S., DiTacchio, L., Pulivarthy, S.R., Le, H.D., and Panda, S. (2009). Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression. Proc. Natl. Acad. Sci. USA *106*, 21453–21458.

Wright, C., and Simone, N.L. (2016). Obesity and tumor growth: inflammation, immunity, and the role of a ketogenic diet. Curr. Opin. Clin. Nutr. Metab. Care *19*, 294–299.

Xie, X., Rigor, P., and Baldi, P. (2009). MotifMap: a human genome-wide map of candidate regulatory motif sites. Bioinformatics *25*, 167–174.

Xie, Z., Zhang, D., Chung, D., Tang, Z., Huang, H., Dai, L., Qi, S., Li, J., Colak, G., Chen, Y., et al. (2016). Metabolic regulation of gene expression by histone lysine β-hydroxybutyrylation. Mol. Cell *62*, 194–206.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-BMAL1 antibody	Abcam	ab93806; RRID: AB_10675117
Anti-PPARα antibody	Santa Cruz Biotechnology	Sc-9000; RRID: AB_2165737
Anti-P84 antibody	GeneTex	GTX70220; RRID: AB_372637
Anti-H3K9/14Ac antibody	Diagenode	C15410200; RRID: AB_2637059
Normal Rabbit IgG	Santa Cruz Biotechnology	SC-2027; RRID: AB_737197
Biological Samples		
Mouse Liver	This paper	See Experimental Model and Subject Details
Mouse Intestine	This paper	See Experimental Model and Subject Details
Mouse Serum	This paper	See Experimental Model and Subject Details
Chemicals, Peptides, and Recombinant Proteins		
GW6471 PPARα inhibitor	Cayman Chemical	11697
Disuccinimidyl Glutarate (DSG)	Thermo Scientific	20593
Trizol Lysis Reagent	Ambion	155696-018
Protein G Sepharose, Fast Flow	Sigma-Aldrich	P3296
Critical Commercial Assays		
β-Hydroxybutyrate LiquiColor Test (Endpoint)	Stanbio	2440-058
BHB Assay Kit	Abcam	ab83390
Free Fatty Acid Quantitation Kit	Sigma-Aldrich	MAK044
Cholesterol Quantitation Kit	Sigma-Aldrich	MAK043
HDAC Assay Kit	Active Motif	56200
Deposited Data		
Liver and IECs circadian microarray dataset	This paper	GEO: GSE87425
Experimental Models: Organisms/Strains		
Male C57BL/6J mice	Jackson Laboratory	000664
Male *Clock*-deficient mice	Laboratory of Dr. S. Reppert (Worcester, MA)	Debruyne et al., 2006
Oligonucleotides		
Primers for RT-PCR, see Table S7	This paper	N/A
Software and Algorithms		
JTK_CYCLE	Hughes et al., 2010	N/A
BIO_CYCLE	Agostinelli et al., 2016	N/A
DAVID v6.7	N/A	N/A
GeneChip Scanner 3000 7G	N/A	N/A
Command Console Software v. 3.2.3	N/A	N/A
Other		
Control Chow Diet (CC)	Envigo, Teklad Custom diet	TD.150345
Ketogenic Diet (KD)	Envigo, Teklad Custom diet	TD.160153
1,3-Butanediol diet (BD)	Envigo, Teklad Custom diet	TD. 160257

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Paolo Sassone-Corsi (psc@uci.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals and Feeding

Eight week-old male C57BL/6J mice (JAX, 00064), maintained at 24°C on a 12 hr light/ 12 hr dark cycle, were fed ad libitum to a ketogenic diet (KD, Envigo, Teklad Custom diet TD.160153, 90.5% kcal from fat, 9.5% kcal from protein), to a control chow (CC, Envigo Teklad Custom diet TD.150345, 12.6% kcal from fat, 9.8% kcal from protein, 77.7% from carbohydrate) or to 1,3-Butanediol diet (10% 1,3-butanediol) (BD, Envigo, Teklad Custom diet TD. 160257) for 4 weeks.

At the end of the 4 weeks 5-7 mice per time-point, per group were sacrificed via CO_2 and cervical dislocation every four hours along the diurnal cycle (from ZT0 to ZT24). *Clock*-deficient animals were a gift from S. Reppert (Worcester, MA). Study animals were backcrossed to the C57BL/6 background for four generations (Debruyne et al., 2006). Eight week-old male *Clock*-deficient mice and their WT littermates were maintained in the same condition described for C57BL/6J. 4-5 *Clock*-deficient animals or WT littermates per time-point were sacrificed via CO_2 and cervical dislocation at ZT12 (light off, starting of the active phase). Liver and intestine were harvested from both C57BL/6J mice and *Clock*-deficient mice and immediately frozen in liquid nitrogen. A portion of the ileum was processed for intestinal epithelial cells isolation as reported below.

A second cohort of animals was sacrificed at ZT0-8-12-20 (n = 5 per time-point) during an independent experiment. Animal care and use was in accordance with guidelines of the institutional Animal Care and Use Committee at the University of California at Irvine.

METHOD DETAILS

Mouse Diets

All diets are matched on a per-calorie basis for micronutrient content, fiber, and preservatives.

The composition of the three experimental diets (CC = control chow, KD = Ketogenic diet, BD = 1,3-Butanediol diet) is as follows (g/Kg):

	CC TD.150345	KD TD.160153	BD TD.160257
Casein	100	180	106.44
DL-methionine	1.6	2.88	1.7
Corn starch	512.46	0	436.973
Sucrose	100	0	85
Maltodextrin	155	0	130.5
Crisco	25	440	0
Cocoa butter	0	150	0
Corn oil	25	85	26.6
Cellulose	35	59.19	37.17
1,3 Butanediol	0	0	100
Calories per gram	3.6	6.7	3.9

Vitamin supplements included AIN-93-VX vitamin mix (Envigo 110068), thiamin, phylloquinone, choline bitartate, mineral mix (Envigo 98057), calcium phosphate, and calcium carbonate. Crisco is a proprietary blend of partially hydrogenated vegetable oil, with minimal trans-fat content. Fatty acids in KD are, by weight, approximately 24% saturated, 39% monounsaturated, and 37% polyunsaturated.

GW6471 Treatment

Eight week-old male C57BL/6J mice (JAX, 00064), maintained at 24°C on a 12 hr light/ 12 hr dark cycle, were fed ad libitum to a ketogenic diet (KD, Envigo diet TD. 160153, 90.5% kcal from fat, 9.5% kcal from protein) or to a control chow (CC, Envigo diet TD.150345, 12.6% kcal from fat, 9.8% kcal from protein, 77.7% from carbohydrate) for 4 weeks. Twice per week the animals were intraperitoneally injected once a day (ZT11) with the specific PPARα inhibitor GW6471 (Cayman chemical, Cat. N. 11697), at a dose of 10mg per Kg of body weight. GW6471 was dissolved in DMSO at a concentration of 10mg/ml and diluted in PBS before the injection. At the end of the 4 weeks 3-4 mice per time-point, per group were sacrificed via CO2 and cervical dislocation at ZT8 and at ZT20. Animal care and use was in accordance with guidelines of the institutional Animal Care and Use Committee at the University of California at Irvine.

Fructose and Sucrose Treatment

Eight week-old male C57BL/6J mice, maintained on a 12 hr light/ 12 hr dark cycle, were fed ad libitum to KD or CC for four weeks. The last week of feeding regimen, Fructose (4g/Kg) dissolved in PBS was administered to the mice through oral gavage (Volume = 200ul), once a day for seven days.

For sucrose experiments, mice were supplemented with 30% sucrose in drinking water ad libitum during the four weeks of dietary regimen. At the end of the 4th week animals were sacrificed via CO2 and cervical dislocation at ZT0 (light on) and ZT12 (light off). Liver and intestine were harvested and immediately frozen in liquid nitrogen.

Indirect Calorimetry

Calorimetry was performed as described in Eckel-Mahan et al. (2012), using negative-flow CLAMS hardware system cages (Columbus Instruments, Columbus, Ohio). Briefly, animals were housed at a temperature of 24°C and subjected to individual indirect calorimetry measurements for a period of 5 consecutive days under a 12 hr light/ 12 hr dark cycle. The first 48 hr of the experiment served to allow acclimation of the mouse to the metabolic cage, and were not included in the analysis. Food and water were available ad libitum during the whole experiment. VO2, VCO2, and food intake were measured every 10 min. RER (VCO2/VO2) was calculated with Oxymax software (Columbus Instruments).

Intestinal Epithelial Cells (IECs) Isolation

IECs was isolated as previously described (Mukherji et al., 2013) with slight modifications. Specifically, 12 cm of ileum (0-12 cm upstream of cecum) was taken, opened longitudinally and washed vigorously in Hanks Balanced Salt Solution (HBSS). The samples were put in 10 mM EDTA in HBSS with 5% fetal bovine serum and shaken at 200 rpm for 20 min at 37°C. The supernatant containing IECs was centrifuged (720 g, 5 min, 4°C) and the pellet was washed in PBS. Following centrifugation (720 g, 5 min, 4°C), the pellet was frozen at −80°C for future analysis.

RNA Extraction

Liver samples were homogenized in Trizol Lysis Reagent (Ambion Cat. N. 155696-018). Chloroform was added and the samples were shaken for 15 s. The samples were left at RT for 3 min and then centrifuged (12000 g, 15 min, 4°C). The aqueous solution, containing RNA, was collected in a fresh tube and the RNA was precipitated by the addition of isopropanol. Samples were mixed, left at RT for 10 min and then centrifuged (12000 g, 10 min, 4°C). Supernatant was discarded and the RNA pellet was washed in 75% ethanol by centrifugation (7500 g, 5 min, 4°C). Supernatant was discarded and the pellet was re-suspended in DEPC-treated water. To increase RNA purity, RNA was purified using the RNeasy mini kit (Qiagen Cat. N. 74106). Total RNA concentrations were determined by Nanodrop Spectrophotometer (Thermo Scientific). RNA quality was analyzed via agarose gel electrophoresis (1% agarose). Total RNA was reverse transcribed using Iscript reverse transcription Super mix (Biorad Cat. N. 1708840).

Gene expression was analyzed by Real time PCR (Biorad CFX96 Real-Time System) using SsoAdvanced Universal SYBR Green Supermix (Biorad Cat. N. 172-5270).

Microarray Analysis

All starting total RNA samples were quality assessed prior target preparation/processing steps by running out a small amount of each sample (typically 25-250 ng/well) onto a RNA 6000 Nano LabChip that was evaluated on an Agilent Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA). Microarray analyses were performed at the Genomics High-Throughput Facility, University of California Irvine.

Gene Ontology Analysis

Gene ontology analysis was performed using Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7, using genomic background; and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway was chosen for gene clustering.

Protein Extraction and Western Blot

For whole cell extracts, liver samples were homogenized in modified RIPA buffer (50 mM Tris pH8, 150mM NaCl, 5mm EDTA, 15mM MgCl2, 1% NP40) plus protease inhibitors and centrifuge for 15 min at 14000 g 4oC. The supernatant was recovered and the protein concentration was determined by Bradford assay (Biorad Cat. N. 500-0006).

For liver nuclear fraction, approximately 250 mg of liver was homogenized in 4 mL buffer A (10 mM HEPES, pH 7.8, 25 mM KCl, 0.5 mM spermidine, 1 mM EGTA, 1 mM EDTA, 0.32 M sucrose, 0.3% triton) with protease inhibitors. Samples were centrifuged (1,000 g, 10 min, 4°C) and the pellets were resuspended in 4 mL buffer A. Following centrifugation (1,000 g, 10 min, 4°C), 4 mL low salt buffer (10 mM HEPES, pH 7.8, 25 mM KCl, 0.5 mM spermidine, 1 mM EGTA, 1 mM EDTA, and 20% glycerol) was added to the pellets and then centrifuged again (1,000 g, 10 min, 4°C). The pellets were resuspended in 1 mL low salt buffer, centrifuged, and resuspended in 1x volume low salt buffer and 2x high salt buffer (10 mM HEPES, pH 7.8, 25 mM KCl, 0.5 mM spermidine, 1 mM EGTA, 1 mM EDTA, 20% glycerol, and 0.5 M KCl). Suspensions were nutated for 1 hr at 4°C and then centrifuged (12,000 g, 20 min, 4°C). The resulting supernatant was used as the nuclear fraction. For nuclear fraction of IECs, buffer B (10 mM HEPES-KOH, pH 7.9, 1.5 mM MgCl2, 10 mM KCl) was added to wash the frozen aliquots of IECs and centrifuged (1,000 g, 5min, 4°C). 0.2% NP40 in buffer B was added to the pellets and left on ice for 10min before centrifugation (10,000 g, 5min, 4°C). The pellets were washed with buffer B and centrifuged again (10,000 g, 5min, 4°C). The pellets were re-suspended in modified RIPA (500 mM Tris-HCl, pH 7.4, 1% NP-40, 0.25% deoxycholic acid-Na, 150 mM NaCl, 1mM EDTA), sonicated and centrifuged (12,000 g, 20min, 4°C). The supernatant was used as the nuclear fraction.

8% SDS-PAGE was performed to check BMAL1 and PPARα expression. The samples were blotted onto nitrocellulose membranes (Biorad) and blocked in 5% non-fat dry milk in Tris-buffered saline (TBS) for 2 hr at room temperature RT. The nitrocellulose membrane was incubated at 4°C overnight with the following antibodies: anti-BMAL1 (Abcam ab93806) 1:2000, anti-PPARα (Santa Cruz Biotechnology Sc 9000) 1:1000 and anti-P84 1:3000 (GeneTex GTX70220). Blots were then washed 3 times in TTBS for thirty minutes, incubated in HRP conjugated anti-mouse or anti-rabbit diluted (1:8000) in 2.5% milk in TTBS for one hour at RT. The membranes were than rinsed three times in TTBS and incubated in enhanced chemiluminescent substrate (Millipore Cat. N. WBKLS0500) and exposed to film. The films were scanned and densitometry was analyzed through ImageJ software.

Chromatin Immunoprecipitation (ChIP)

Minced frozen liver or ileum tissue was double crosslinked with DSG for 40 min and 1% formaldehyde for 10 min followed by Glycine (0.125M final concentration) at room temperature for 10 min. After homogenizing tissue pellet in PBS, 1 mL of lysis buffer was added. Samples were sonicated by Bioruptor (20 cycles, every cycle: 30 s ON / 30 s OFF, power high) to generate 200-500 base pairs fragments and centrifuged at 14000 g at 4°C. Supernatants were diluted in dilution buffer (1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl, 167 mM NaCl), precleared with Protein-G beads, blocked with salmon sperm DNA and BSA for 2 hr. Precleared supernatant was incubated with the following primary antibodies: 2 μg BMAL1 (Abcam Cat. N. ab93806) or 1 μg H3K9/14ac (Diagenode Cat. N. C15410200) overnight at 4°C. To monitor the specificity of ChIP assays, samples were also immunoprecipitated with a specific-antibody isotype matched control immunoglobulin (IgG). Protein-G beads Sepharose, Fast Flow (Sigma-Aldrich Cat. N. P3296) were added to the supernatant and incubated for 2 hr at 4°C and centrifuged. Beads were recovered, washed in low salt buffer, high salt buffer, LiCl buffer, followed by washing in TE for three times. Elution buffer (300 mM NaCl, 0.5% SDS, 10 mM Tris-HCl, 5mM EDTA) was added to the washed beads, treated with RNase at 37°C for 2 hr and Proteinase K at 65°C overnight. Equal amount of Phenol-Chloroform-Isoamyl alcohol was added to the samples and the aqueous phase was recovered. DNA was precipitated by adding 100% Ethanol, NaOAc and glycogen and kept at −20°C overnight. Samples were centrifuged at 14000 g for 30 min at 4°C and washed with 70% ethanol followed by centrifugation at 14000 g for 30 min at 4°C. Quantitative PCRs were performed using SsoAdvanced Universal SYBR Green Supermix (Biorad), according to the manufacturer's protocol.

qPCR Primers

The primers for gene expression and ChIP were designed by Primer 3 software (v. 0.4.0) or obtained from previous publications: Dbp E-box ChIP primers are based on Ripperger and Schibler (2006), and Nampt E-box chip primers are based on Nakahata et al. (2009). All the primer sequences are reported in Table S7.

β-Hydroxyl-Butyrate Quantitation

Fresh blood was centrifuged at 1500 g for 15 min and serum isolated. β-Hydroxyl-butyrate level were quantified using β-Hydroxybutyrate LiquiColor Test (Endpoint) (StanBio Cat. N. 2440-058) according to the manufacturer's instructions. For β-Hydroxybutyrate levels in intestine and liver, we used Abcam BHB Assay kit (Cat. N. ab83390). The samples were prepared and deproteinated with PCA according to the manufacturer's instruction.

Fatty Acid and Cholesterol Quantitation

10-15 mg of liver or intestinal tissue were used to check free fatty acid and total cholesterol levels. Free fatty acid were quantified using the "Free fatty acid quantitation kit" (Sigma-Aldrich Cat. N. MAK044) according to manufacturer's instructions. Total cholesterol was quantified using the "Cholesterol quantitation kit" (Sigma-Aldrich Cat. N. MAK043) according to manufacturer's instructions.

HDAC Activity Assay

15 μg of intestinal or liver nuclear extract was tested using HDAC assay kit (Active Motif Cat. N. 56200) according to manufacturer's instructions.

QUANTIFICATION AND STATISTICAL ANALYSIS

Microarray Data Analysis and Statistics

3 animals per time-point, per group were analyzed. Arrays were scanned using GeneChip Scanner 3000 7G and Command Console Software v. 3.2.3 to produce .CEL intensity files. The intensity files were then analyzed in Affymetrix Expression Console software v1.1.1 using the PLIER algorithm to generate probe level summarization files (.CHP), where signal levels were normalized between KD and CC conditions within the same tissue. Expression levels time series were then used to determine circadian behaviors of transcripts using JTK-CYCLE (Hughes et al., 2010) and the results were corroborated using BIO_CYCLE (Agostinelli et al., 2016), an improved circadian rhythm analysis software. Circadian behaviors include whether or not a transcript is circadian as well as its phase and amplitude. Both methods produced similar results and importantly, key transcripts such as the core clock genes and PPARα targets showed the same behavior. In the rest of the analysis, we report the results obtained with JTK-CYCLE. For JTK-CYCLE, a gene was considered circadian, if at least one of its transcripts passed a p value cutoff of 0.01. Data visualization was done in Python using 'matplotlib' and R using 'gplots'. The Database for Annotation, Visualization and Integrated Discovery (DAVID) pathway

analysis tool was used to identify GO terms related to biological process and potentially enriched pathways (Huang et al., 2009). Pathways were ranked by the number of circadian genes they contained with a JTK-CYCLE p value < 0.01, and results were compared to the genomic background for enrichment. For the background, we selected all the possible genes in the whole genome that can be detected by the specific microarray (extracted from the microarray annotation). The total number of background genes is around 25,000. The lists of PPARα and BMAL1 target genes used in focused analysis were generated using a combination of literature evidence (Rakhshandehroo et al., 2010), MotifMap (Daily et al., 2011; Xie et al., 2009) and publically available ChIP Seq datasets from GEO (GEO: GSE35262 and GSE39977). Differential analysis of transcript expression levels between KD and CC at certain ZTs was done using CyberT (Baldi and Long, 2001; Kayala and Baldi, 2012), a differential analysis program using a Bayesian-regularized t test. Variance Stabilizing Normalization was used process the data for input to CyberT using the R package 'vsn'. Further statistical analysis including multiple hypothesis testing corrections was done in R using 'fdrtool'.

Analysis was done using pipelines implemented for the Circadiomics (Patel et al., 2012, 2015) database and web portal (http://circadiomics.ics.uci.edu/) where all the transcriptomic data associated with this work is publicly available.

Phase lag analysis was performed based on the "LAG" (predicted phase) obtained by JTK-CYCLE for every single gene (See also Tables S3 and S4). The genes with the same "LAG" were grouped together and the data reported in the radar plot as percentage. Anderson-Darling tests were performed using the R package 'kSamples' to determine whether or not the phase distributions of KD and CC circadian transcripts were statistically different ($p < 0.0001$). Student's t tests were conducted to determine if changes, in terms of JTK amplitudes or specific expression at ZT8, were significant between KD and CC conditions for circadian transcripts.

Transcription Factor Enrichment Analysis

Transcription factors were analyzed in terms of enrichment for binding sites in the promoter regions (−10000∼+2000 bps of TSS) of rhythmic genes in each circadian group, where binding sites were extracted from high quality MotifMap results for the mouse genome build mm9 (BBLS > 1, FDR < 0.25). A Fisher's exact test was performed between the circadian genes and the whole genome to establish enrichment (odds > 1, $p < 0.05$). TFs with motifs that are too short or degenerate (more than 50000 binding sites under the filtering criteria) were removed, as they tend to be unreliable.

Enrichment results for different conditions across circadian groups were then compared in a meta-analysis to identify tissue-specific TFs. As an example, a PPARα related motif was found to be uniquely enriched in IECs under ketogenic diet, when compared to both IECs under normal chow diet or liver under ketogenic diet, which suggests that it may explain the tissue-specific response to ketogenic diet treatment. In contrast, Clock:Bmal motif was found to be enriched in virtually all circadian groups except for ketogenic IECs when compared to ketogenic liver, which suggests that there is a stronger tissue-specific effect of Clock:Bmal controlled genes in ketogenic liver when compared to IECs, despite the fact that Clock:Bmal is virtually ubiquitously enriched as a core circadian TF. Full details of the analysis are provided as Table S6 (sheet 4).

Other Statistical Analysis

Data are expressed as mean ± SEM. The significance of differences was analyzed by Student's t test or two-way ANOVA and post hoc analysis for multiple group comparison. The test was considered significant with a P value < 0.05.

DATA AND SOFTWARE AVAILABILITY

The GEO accession number for the microarray dataset reported in this paper is GEO: GSE87425. All the transcriptomic data associated with this work is publicly available on the resource http://circadiomics.ics.uci.edu/.

4.4 pyfuncgraph

`pyfuncgraph` utilizes the same data sources for the enrichment pipelines (e.g., MotifMap, MotifMap-RNA, KEGG and GO), but in a much more sophisticated way. Instead of performing enrichment analyses based on categorical data derived from functional information, it uses functional data directly to construct complex functional graphs.

For MotifMap, MotifMap-RNA and ChIPSeq data, sources such as transcription factors, RNA binding proteins or histones are directly added as nodes in a directed, weighted graph. Their targets are also added as nodes in the graph (note that sources can be targets as well). Edges are constructed by binding sites or peaks, which may exist at the promoter, intron or UTRs depending on the source and target. This graph is also a multigraph, permitting multiple edges of the same direction between a set of source and target. Furthermore, functional graphs derived from each of the databases are also merged to generate a meta-functional graph embodying the totality our knowledge of transcriptional and post-transcriptional regulation in the cell (for mouse). Different filtering parameters are used for different edges in order to generate comparable and reliable results.

For pathway and GO term databases, a bipartite pathway graph is created where different functional entities (e.g. pathway, GO term) inhabit one half and their targets (transcripts, metabolites) inhabit the other. Targets are linked if they associate with the same functional entity. Again, metagraphs are created by combining different data sources.

Due to the gigantic sizes of these functional graphs (the MotifMap-MotifMapRNA-ChIPSeq meta graph has 30,334 nodes with 1,749,132 edges), `pyfuncgraph` does not provide any client-side service by its own. Instead graphical resources are stored as network objects on a webserver, in netowrkx (https://networkx.github.io/). Client side software, such as igb-pipelines (Chapter 5) and PyCircadiOmics (Chapter 7), query this server to obtain subgraphs, target nodes, outgoing edges or other information derived from the functional graph.

By preserving the complex regulatory relationships between transcripts and their regulators, `pyfuncgraph` provides the basis for systematic graphical analysis of the circadian transcriptome. It contributed to the discovery of novel circadian regulators and helped established a hierarchical model for the organization of circadian transcripts centering around the core clock (Chapter 7).

Chapter 5

Circadian Web Development–viz and goworg

5.1 Introduction

The software described in Chapter 2 and 3 (with data derived partially from results described in Chapter 1) constitute the most important back ends of the `igb-pipelines` codebase. Through these software, we can rapidly generate massive amounts of bioinformtics results, such as circadian statistics, figures generated from visualization and machine learning analyses, functional enrichment results, etc. Although these results qualify as biological interpretations of some kind, by themselves they cannot produce meaningful biological discoveries. Indeed, they must be further summarized and experimentally validated by bioinformaticians and biologists. To facilitate this collaborative process, it is necessary to construct a management system that can host and organize large amounts of bioinformatics data. Moreover, such a system should be readily accessible to many biologists, some of whom may not be interested or well-equipped to directly process data and understand it at a low level. There-

fore, it is necessary to provide extensive tools for documentation and interactive visualization so that users can "play with" the data and obtain intuitive understanding which may inspire further experiments.

The `viz` and `goworg` softwares are designed specifically to enable this kind of collaboration. They serve as the front end of the `igb-pipelines` software system. Importantly, they can also function as reliable archives of past research projects that can make reproduction of bioinformatics analysis very convenient.

5.2 Viz

`viz` serves as a HTML templating system that takes various outputs from circadian softwares such as `mpfe` and renders them into static HTML with interactive and scalable elements are realized by javascript. `viz` contains a data layer in `tabular.py` which works similarly to a dataframe in pandas, but is very lightweight and implemented in pure python. Datatables are loaded by this data layer and converted to JQuery datatables (https://datatables.net/), Google Charts (https://developers.google.com/chart/) or worg-javascript tables (a project by Christophe Magnan). These elements are injected into static HTML with additional interactive visualization elements provided by Google Charts. Various HTML templates with dropdown selection and flexible client side search are available for rendering. A brief summary of this codebase in the form of class dependency graph generated by doxygen (https://www.stack.nl/ dimitri/doxygen/) is shown in the following figure. An example of an interactive project page output showcasing flexible and searchable datatables with visualization elements is shown further below.

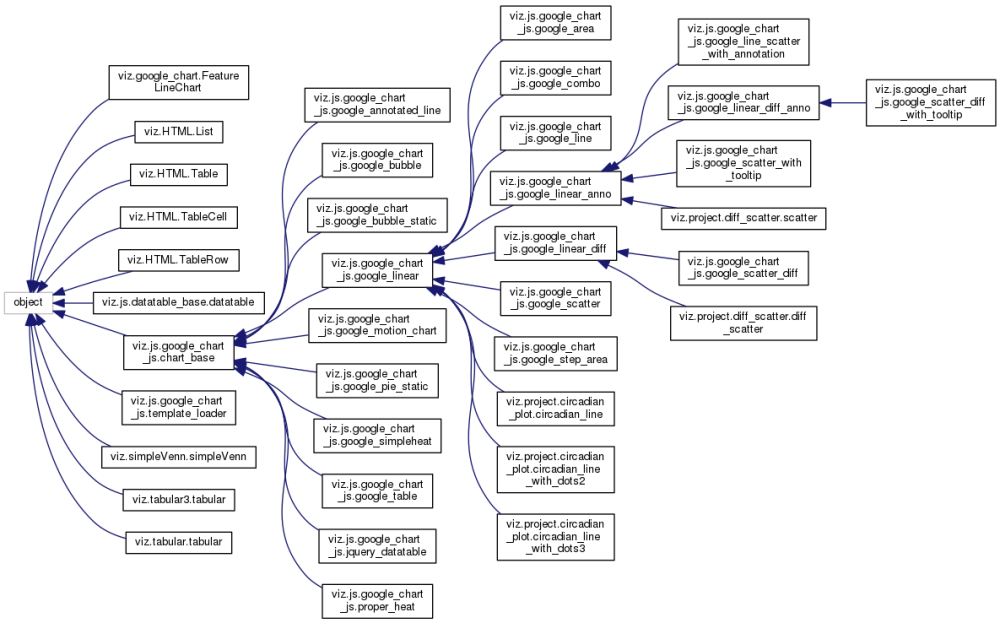


Figure 5.1: Class inheritance graph for the viz codebase. Tabular is used as a low level data layer while more than a dozen interactive visualization template elements are available.

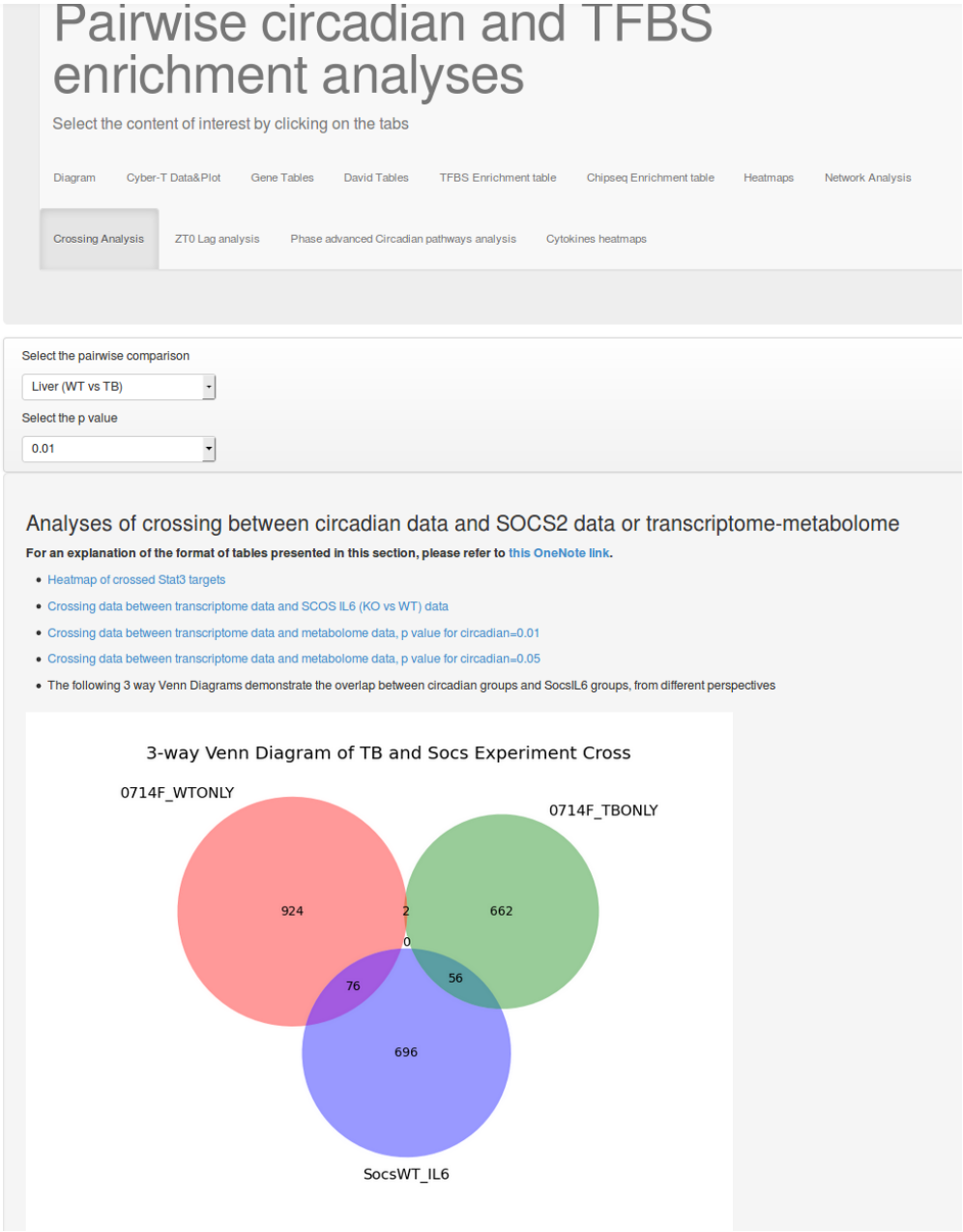


Figure 5.2: Example output of the templated HTML generated by Viz. It contains interactive elements such as tabs, dropdowns, and static/vectorized figures. Data taken from a tumor bearing mouse study [37].

5.3 goworg

The `viz` templating system is accompanied by `goworg`, a webserver inspired by the R project Worg (https://orgmode.org/worg/). `worg` is written in GO and supported by the Echo framework (https://echo.labstack.com/). Instead of being programmable like Worg, `goworg` interacts only with static HTMLs. This is not a drawback because the automation functionalities provided by the viz-goworg system greatly reduce the workload necessary for developing a project website. Briefly, to construct a project folder in `goworg`, one only needs to create a folder of static files under monitored paths. The webserver scans the path and automatically creates templates for raw data files such as tables and static figures. Any static HTML generated by `viz` works properly under its original directory structure. Moreover, markdown files are rendered to HTML by the server, which makes writing richly styled documentation trivial for the user.

The goworg system is available at http://worg.ics.uci.edu/. Here we note that individual project pages are not available for public browsing.

5.4 Application: Tumor Bearing Mouse Liver Study

The viz-goworg system and the project websites generated using these software have been utilized in many projects [41, 49]. Here I append a highlighted paper by Masri et al. on the circadian reprogramming in mouse liver where the host bears a cancerous lung, for which I am a contributing author [37].

This research utilized results hosted on the aforementioned systems, as well as other results generated by the whole `igb-pipelines`. In particular, Figure 2 A C D and Figure 3 A B utilized results extracted from the viz-worg system. Extensive bioinformatics analysis,

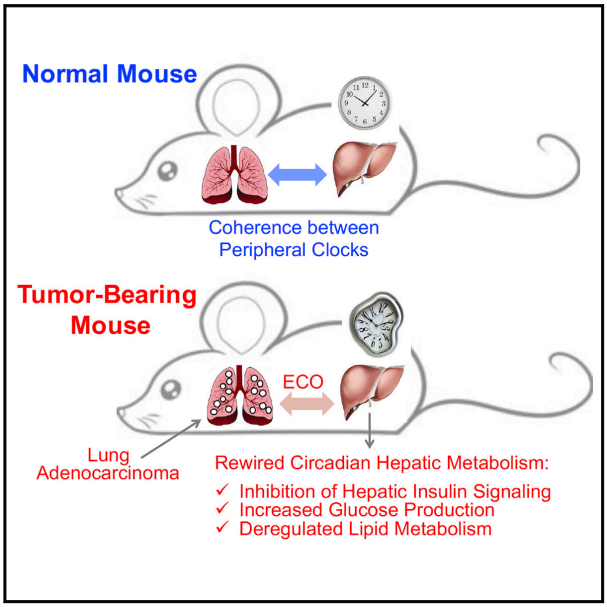
rendered through the viz-worg system, revealed that there is a distinct set of oscillating genes in mouse liver when the host also has a tumor bearing lung. These distinct oscillating transcripts exhibit dramatic differences in circadian behavior, including phase shifts. Functional enrichment analysis revealed that a significant number of these transcripts belong to metabolism related pathways including insulin response and cell proliferation. Further analysis of the metabolome identified metabolites linked to the the same pathways, thus establishing a circadian reprogramming event driven by metabolism and inflammatory response related pathways. The full scope and details of the study are shown in the paper below.

Article

Cell

Lung Adenocarcinoma Distally Rewires Hepatic Circadian Homeostasis

Graphical Abstract



Authors

Selma Masri,
Thales Papagiannakopoulos,
Kenichiro Kinouchi, ..., Pierre Baldi,
Tyler Jacks, Paolo Sassone-Corsi

Correspondence

psc@uci.edu

In Brief

Lung adenocarcinoma rewires circadian metabolic cycles to inhibit hepatic insulin signaling, decrease glucose sensitivity, and reorganize lipid metabolism.

Highlights

- Lung adenocarcinoma rewires circadian transcription and metabolism in the liver
- This rewiring involves the STAT3-Socs3 inflammatory signaling axis
- Inhibition of hepatic insulin signaling and glucose intolerance are tumor driven
- Lung adenocarcinoma drives deregulation of lipid metabolism in the liver

Accession Numbers

GSE73222

Masri et al., 2016, Cell *165*, 896–909
May 5, 2016 ©2016 Elsevier Inc.
http://dx.doi.org/10.1016/j.cell.2016.04.039

CellPress

Cell

Article

Lung Adenocarcinoma Distally Rewires Hepatic Circadian Homeostasis

Selma Masri,[1] Thales Papagiannakopoulos,[2,5] Kenichiro Kinouchi,[1] Yu Liu,[3] Marlene Cervantes,[1] Pierre Baldi,[3] Tyler Jacks,[2,4] and Paolo Sassone-Corsi[1,*]

[1]Center for Epigenetics and Metabolism, INSERM Unit 904, Department of Biological Chemistry, University of California, Irvine (UCI), Irvine, CA 92697, USA

[2]Division of Comparative Medicine, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[3]Institute for Genomics and Bioinformatics, Department of Computer Science, UCI, Irvine, CA 92697, USA

[4]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[5]Present address: Department of Pathology/Perlmutter Cancer Center, New York University (NYU) School of Medicine, New York, NY 10016, USA

*Correspondence: psc@uci.edu

http://dx.doi.org/10.1016/j.cell.2016.04.039

SUMMARY

The circadian clock controls metabolic and physiological processes through finely tuned molecular mechanisms. The clock is remarkably plastic and adapts to exogenous "zeitgebers," such as light and nutrition. How a pathological condition in a given tissue influences systemic circadian homeostasis in other tissues remains an unanswered question of conceptual and biomedical importance. Here, we show that lung adenocarcinoma operates as an endogenous reorganizer of circadian metabolism. High-throughput transcriptomics and metabolomics revealed unique signatures of transcripts and metabolites cycling exclusively in livers of tumor-bearing mice. Remarkably, lung cancer has no effect on the core clock but rather reprograms hepatic metabolism through altered pro-inflammatory response via the STAT3-Socs3 pathway. This results in disruption of AKT, AMPK, and SREBP signaling, leading to altered insulin, glucose, and lipid metabolism. Thus, lung adenocarcinoma functions as a potent endogenous circadian organizer (ECO), which rewires the pathophysiological dimension of a distal tissue such as the liver.

INTRODUCTION

Metabolic, endocrine, and behavioral functions are largely circadian and their disruption is associated with a number of disorders and pathologies, including cancer (Asher and Sassone-Corsi, 2015; Bass, 2012; Fu and Lee, 2003; Gamble et al., 2014; Masri et al., 2015; Partch et al., 2014). Circadian rhythms are governed by molecular machinery whose function is to maintain rhythmic precision within cells and synchrony between central and peripheral clocks. Importantly, circadian transcriptional circuits function in a defined tissue-specific manner by interplaying with specialized nuclear factors through poorly understood mechanisms (Masri and Sassone-Corsi, 2010; Panda et al., 2002). Under standard physiological states, the core clock machinery is coupled to the metabolic cycles with which it operates in a coherent, concerted manner. However, the clock is also able to adapt to changing metabolic fluctuations as a compensatory mechanism and it does so by utilizing alternative transcriptional strategies. For instance, restricted feeding temporally phase shifts circadian gene expression in the liver (Damiola et al., 2000; Stokkan et al., 2001; Vollmers et al., 2009) and nutritional challenge is able to reprogram circadian transcription and subsequently alter cyclic metabolism (Eckel-Mahan et al., 2013; Hatori et al., 2012; Kohsaka et al., 2007). Therefore, timing of food intake and nutritional challenge are able to uncouple the timekeeping of hepatic metabolic oscillations from the core clock machinery. Yet, aside from the consequences of nutritional challenge, the effects of other non-dietary factors that could uncouple and disrupt the hepatic clock remain largely unexplored.

Cancer cells thrive based on a heightened metabolic rate that circumvents typical physiological means for energy production through the so-called Warburg effect (Hsu and Sabatini, 2008; Vander Heiden et al., 2009). In addition, cancer cells excrete a number of factors systemically, including metabolic "waste" by-products and/or inflammatory signals (Hanahan and Weinberg, 2011; Lin and Karin, 2007). For example, tumor-secreted lactate, a product of increased aerobic glycolysis of cancer cells, is associated with heightened metastatic incidence and increased angiogenesis, is responsible for metabolic reprogramming in adjacent tissues, and can induce a pro-inflammatory state (Colegio et al., 2014; Doherty and Cleveland, 2013). Similarly, the cooperative effects of the inflammatory response during tumorigenesis are well-documented (Gao et al., 2007; Sansone et al., 2007). Tumor-secreted cytokines, such as Interleukin-6 (IL-6), can regulate metabolism in multiple tissues (Mauer et al., 2015), suggesting a possible role in mediating tumor-induced metabolic changes systemically. Collectively, these tumor-derived metabolites and cytokines constitute the so-called tumor "macroenvironment" (Al-Zoughbi et al., 2014), the systemic metabolic consequences of which remain elusive.

Importantly, the effects of a tumor on organismal homeostasis are poorly understood, and given the unique ability of the clock in sensing metabolic discrepancies, a potential role of cancer in rewiring clock-controlled metabolism is intriguing (Sahar and Sassone-Corsi, 2009). Indeed, our results demonstrate that lung adenocarcinoma rewires the circadian hepatic transcriptome and corresponding metabolome, yet the core clock machinery remains virtually unperturbed. The tumor imposes a profound metabolic reprogramming that implicates a number of signaling pathways, which operate within the framework of the tumor macroenvironment. As a paradigm, we reveal that the inflammatory STAT3-Socs3 signaling axis is induced in the liver of lung-tumor-bearing mice, resulting in inhibition of hepatic insulin signaling, glucose intolerance, and deregulated lipid metabolism. In conclusion, we illustrate a previously unappreciated role played by a distally located lung adenocarcinoma as an endogenous circadian organizer (ECO) in the rewiring of circadian homeostasis of the liver.

RESULTS

Lung Cancer as an Endogenous Reorganizer of Circadian Rhythms

The *Kras*$^{LSL-G12D}$;*p53*$^{fl/fl}$ mice are a genetic model of lung adenocarcinoma that mimics human non-small cell lung cancer (NSCLC) (Jackson et al., 2005; Jackson et al., 2001). Upon intratracheal delivery of equivalent adenoviral titer of Cre recombinase, which induces the genetic rearrangement of the *Lox-stop-Lox* cassette to activate oncogenic Kirsten rat sarcoma viral oncogene homolog (*Kras*) and to knock out the tumor suppressor *p53*, mice developed defined lung adenocarcinoma (Figure S1). This mouse model generates lung adenocarcinoma with 100% penetrance and uniform tumor burden among all mice (Jackson et al., 2001). Equivalent adenoviral titer of FlpO recombinase was administered to *p53*$^{fl/fl}$ littermates of the same pure C57BL/6J background as a control that does not induce recombination. Upon sacrifice, lung-tumor-bearing (TB) mice exhibited wild-type (WT) expression of *Kras* in the liver, white adipose tissue (WAT), and muscle, and no metastatic lesions were observed in the liver (Figure S2).

To investigate the distal effects of lung adenocarcinoma on circadian hepatic function, WT and TB mice were sacrificed every 4 hr over the circadian cycle (zeitgeber time [ZT] 0, 4, 8, 12, 16 and 20) and livers were subjected to transcriptomics and metabolomics analyses. Heat maps for oscillating genes based on transcriptomics, as determined by JTK_cycle, display striking differences in unique sets of oscillating genes from WT (left) and TB (right) mice (Figures 1A and 1B). Gene ontology (GO) biological function was determined using DAVID pathway analysis for WT or TB oscillating genes. Pathway analysis revealed that WT-specific genes were enriched for a number of metabolic processes, including insulin response and regulation of cell cycle and proliferation, while TB-only oscillating genes were selectively enriched for endoplasmic reticulum (ER) signaling, unfolded protein response, cholesterol biosynthesis, and redox state (Figures 1C and S3). Phase analysis was performed for uniquely oscillating WT- and TB-specific genes to determine the relative phase of circadian gene expression. The peak in phase of expression was around ZT 8 in the WT category, whereas TB oscillating genes exhibited a bi-phasic profile that peaked around ZT 0 and again at ZT 12 (Figure 1D). Using the set of 505 genes that retain oscillation in both WT and TB mice, phase analysis was performed to determine if rhythmic genes retained their peak in expression. Strikingly, 46% of circadian genes exhibited a phase change, with 68% of these genes being phase advanced and 32% were phase delayed by at least 1 hr (Figure 1E). These results demonstrate that lung adenocarcinoma significantly reprograms the circadian hepatic transcriptome.

Similar to the circadian transcriptome, metabolomics analysis revealed unique sets of oscillating metabolites in the livers of WT (left heat maps) or TB (right heat maps) mice (Figure 2A). Of ~600 identified metabolites, two-way ANOVA analysis identified that 235 metabolites were differentially altered by the lung tumor and 328 metabolites were differentially expressed by time point (Figure 2B). Oscillating metabolites were further determined using JTK_cycle, and though the oscillation of 159 metabolites persisted, 90 were rhythmic exclusively in WT and 84 exclusively in TB mice (Figure 2B). Of these 159 metabolites that oscillate in WT and TB mice, 53% exhibited a change in phase, with 62% and 38% being phase advanced and delayed, respectively (Figure 2C). Classification of these metabolites into pathways demonstrated a clear reduction in oscillating lipids in TB versus WT mice (Figure 2D). In addition, a reduction in the levels of energetic metabolites NAD^+, ATP and acetyl-CoA was seen (Figure 2E). This indicates altered usage or production of these molecules resulting in disruption of liver homeostasis in TB animals. Thus, the presence of lung tumors acts to distally rewire both transcriptional and metabolic programs in the liver. As further depicted below, this circadian reorganization appears to coordinately contribute to a TB-specific hepatic metabolic profile.

Lung Adenocarcinoma Does Not Affect Hepatic Core Clock Components

A detailed analysis of the genes that were not altered between WT and TB mice was carried out, as shown in the heat map in Figure 3A. GO pathway analysis revealed this category is enriched in not only select metabolic genes, but also in rhythmic genes pertaining to the circadian clock (Figure 3B). The phosphorylation of the aryl hydrocarbon receptor nuclear-translocator-like (ARNTL or BMAL1) protein and expression of all core clock genes, including circadian locomotor output cycles kaput (*Clock), Bmal1, Period (Per1-3)*, Cryptochrome (*Cry1/2)* and nuclear receptor subfamily 1, group D (*Nr1d1* or *Rev-Erbα)*, as well as the clock-controlled D site of albumin promoter binding protein (*Dbp)* gene, were unchanged in the livers of TB animals (Figures 3C and S4). In order to better characterize the effects of lung adenocarcinoma on the clock, locomotor behavior was analyzed, and no change in the free-running period was observed between WT and TB mice (Figure 3D). Similarly, behavioral actograms show that the circadian activity profile was equal during the light/dark cycles in WT and TB mice (Figure S4). Also, the feeding behavior remained rhythmic in TB mice while a non-significant decrease in food intake was observed (Figure 3E). The respiratory exchange ratio (RER) remained rhythmic, but TB mice displayed an elevated RER during the light phase and a dampened RER during the dark phase (Figure 3F), in keeping with a reduction in VO2, VCO2, and heat production

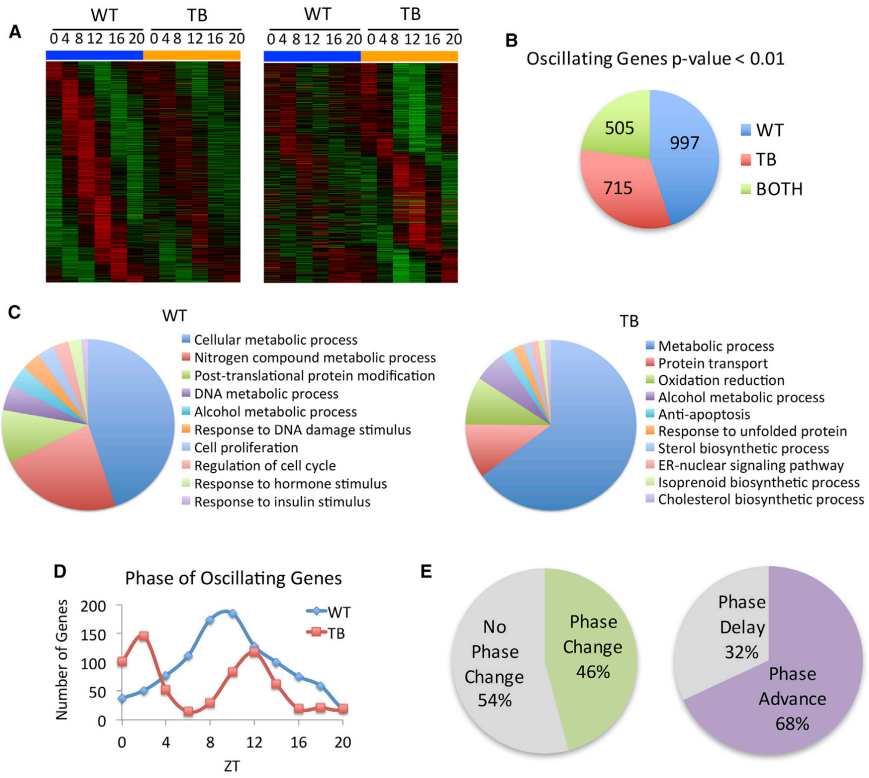


Figure 1. Lung Adenocarcinoma Rewires the Circadian Hepatic Transcriptome

A) DNA microarray analysis was performed using mouse liver total RNA from ZT 0, 4, 8, 12, 16, and 20. Using JTK_cycle, genes selected to be circadian at a p value < 0.01 are displayed as heat maps for WT and lung-tumor-bearing (TB) livers. Left displays circadian genes exclusively in WT mice and right shows genes with more robust oscillation in TB mice.

B) Pie charts indicate actual numbers of circadian genes that oscillate exclusively in WT, TB, or BOTH conditions.

C) Top ten gene ontology (GO) terms for biological process were identified by DAVID pathway analysis tool, based on a 0.01 p value cutoff.

D) Phase analysis of WT- and TB-specific oscillating gene expression profiles.

E) Phase analysis of "BOTH" genes that remain circadian in WT and TB mice.

(Figure S5). The altered circadian metabolites (Figures 2D and 2E) in conjunction with dampened RER levels (Figures 3F and S5) revealed a significant shift in the metabolic state of TB mice. Indeed, repressed energy expenditure might be a contributing factor to the uncoupling of the core clock and metabolic rhythms. Timing of food intake, which functions as a powerful zeitgeber (Damiola et al., 2000; Eckel-Mahan et al., 2013; Vollmers et al., 2009), also remains virtually unaltered in TB mice (Figure 3E).

Lung Adenocarcinoma Rewires Hepatic Metabolism, but Not the Core Clock

Given the changes in energy expenditure as measured by RER (Figure 3F) and the dampened lipid profiles identified by metabolomics in TB mice (Figure 2D), the effect of lung adenocarcinoma on fatty acid synthesis, breakdown by beta-oxidation, and utilization for cholesterol production were further investigated. The sterol regulatory element binding protein (SREBP) pathway is known to control lipid metabolism in the liver in a circadian manner (Gilardi et al., 2014; Le Martelot et al., 2009), and its deregulation is in accordance with the observed alteration of lipid levels (Figure 2D). The SREBP pathway is known to be inhibited by the energy sensor AMP-activated protein kinase (AMPK) (Li et al., 2011; Vavvas et al., 1997). Indeed, activation of AMPKα by phosphorylation of threonine (Thr) 172 was markedly elevated in TB mice and peaked at ZT 16 (Figure 4A). Given the dampened ATP levels in TB mice (Figure 2E), these effects are aligned with the increased intracellular AMP/ATP ratios over the circadian cycle (Figure 4A). Accordingly, the SREBP1 pathway was suppressed, as both gene expression profiles and the levels of the mature form of nuclear SREBP1c protein were repressed at ZT 16 in TB mice

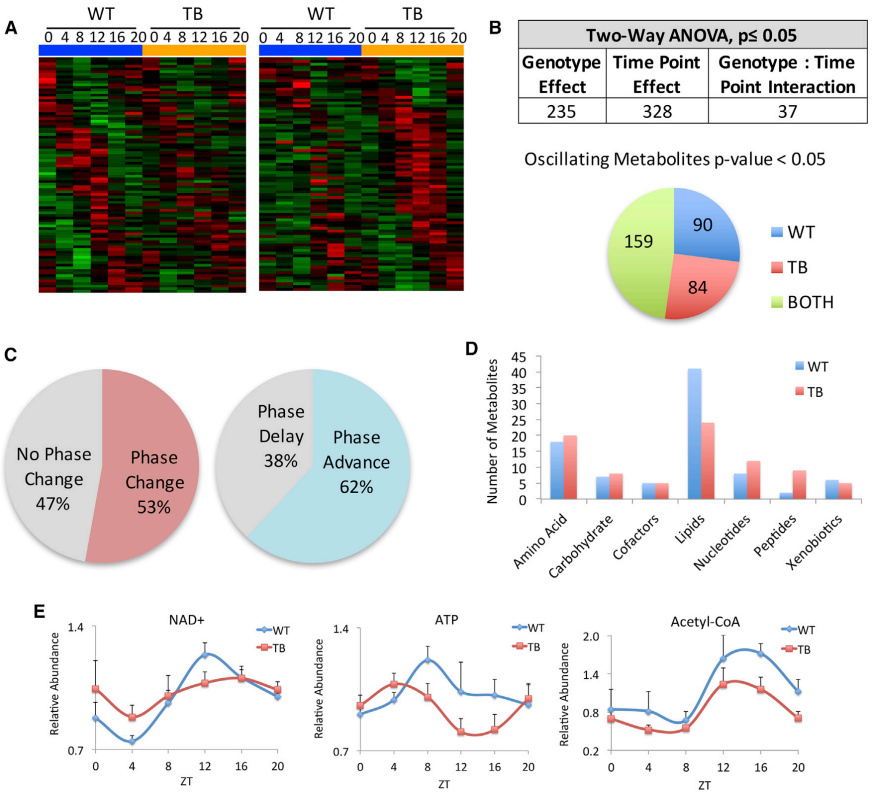


Figure 2. The Circadian Metabolome is Reorganized by Lung Cancer

A) Heat maps displaying oscillating metabolites as determined by JTK_cycle (p value < 0.05) in WT and TB mice. Left displays circadian metabolites exclusively in WT liver and right shows metabolites with more robust oscillation in TB mice.

B) Two-way ANOVA analysis using a p value cutoff of 0.05 reveals metabolites that are responsive to lung tumors, circadian time point, or both. Numbers of oscillating metabolites using JTK_cycle are indicated from WT, TB, or BOTH categories.

C) Phase analysis was performed using JTK_cycle to identify the phase of peak metabolite expression.

D) Oscillating metabolites are displayed based on biological sub-pathway, including amino acid, carbohydrate, cofactors, lipids, nucleotides, peptides, and xenobiotics.

E) Examples of energetic metabolites that are dampened in TB mice. NAD^+, nicotinamide adenine dinucleotide; ATP, adenosine 5′-triphosphate.

Error bars indicate standard error of mean (SEM).

(Figure 4B). Similarly, significant inhibition of SREBP1 target genes was observed, as seen with *Fasn, Acaca*, and *Elovl6* expression (Figure 4C). The repression of SREBP1-dependent signaling in the livers of TB mice was further substantiated by the decreased levels of long-chain fatty acids and esterified fatty acids (Figure 4D), including myristate, linolenate, palmitoleate, and eicosapentaenoate (EPA). This suggests either a decrease in fatty acid biosynthesis or an increase in breakdown by beta-oxidation, the former case being most likely given the suppression of SREBP1 signaling and the unaltered peroxisome proliferator-activated receptor alpha (PPARα) and beta-oxidation gene expression profiles in livers of TB mice (Figure S6).

In contrast to the suppression in the SREBP1 pathway, SREBP2 gene expression is not repressed in TB mice and its target genes lanosterol synthase (*Lss)*, 3-hydroxy-3-methylglutaryl-CoA synthase (*Hmgcs1*), and phosphomevalonate kinase (*Pmvk)* showed a significant and coordinated increased peak in expression at ZT 16 (Figure 4E). As SREBP1 is primarily involved in fatty acid biosynthesis and SREBP2 is critical for cholesterol production (Brown and Goldstein, 1997; Horton et al., 2003), an increase in total cholesterol levels was observed in TB mice (Figure 4F), which paralleled SREBP2-dependent gene expression profiles (Figure 4E). Overall, TB mice displayed a deregulation of SREBP signaling, with a suppression of fatty acid synthesis

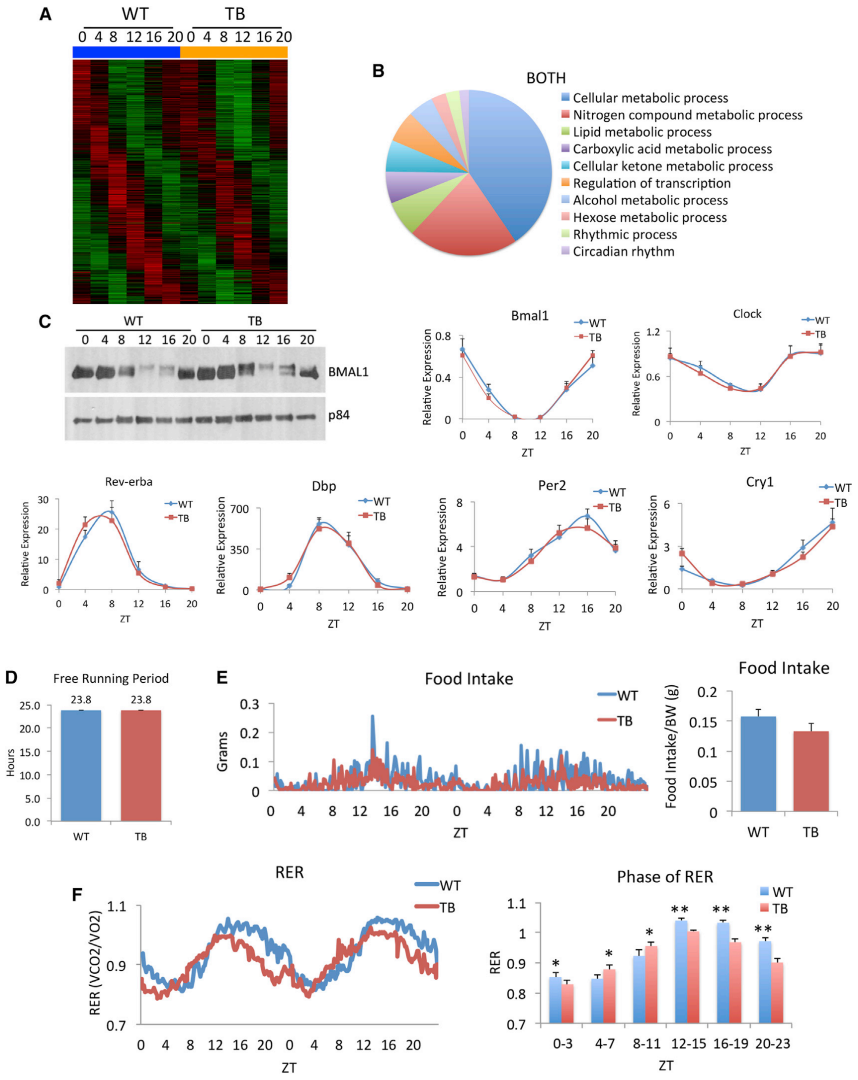


Figure 3. The Circadian Clock Is Unaffected by Lung Adenocarcinoma

A) Heatmap for BOTH category genes that are unaltered in expression between WT and TB mice.

B) GO pathway analysis using biological process for BOTH oscillating genes.

C) BMAL1 protein phosphorylation by western and circadian expression of the clock genes, *Bmal1, Clock, Rev-erbα, Dbp, Per2*, and *Cry1*, as determined by quantitative real-time PCR.

D) Locomotor activity analysis for WT and TB mice, as calculated by the free-running period (Tau) in dark/dark (D/D) conditions. An n = 12 WT and n = 12 TB mice were used for behavioral analysis.

E) Food intake of WT and TB mice shown over a 48 hr period (left). Total food intake was normalized to body weight of each animal. An n = 7 WT and n = 8 TB mice were used for indirect calorimetry analysis.

F) VCO_2/VO_2 is shown as the respiratory exchange ratio (RER) for WT and TB mice over a 48 hr period. Average RER is quantified during the light and dark phases (right panel). Error bars indicate SEM. Significance was calculated using Student's t test and * and ** indicate p value cutoffs of 0.001 and 0.0001, respectively.

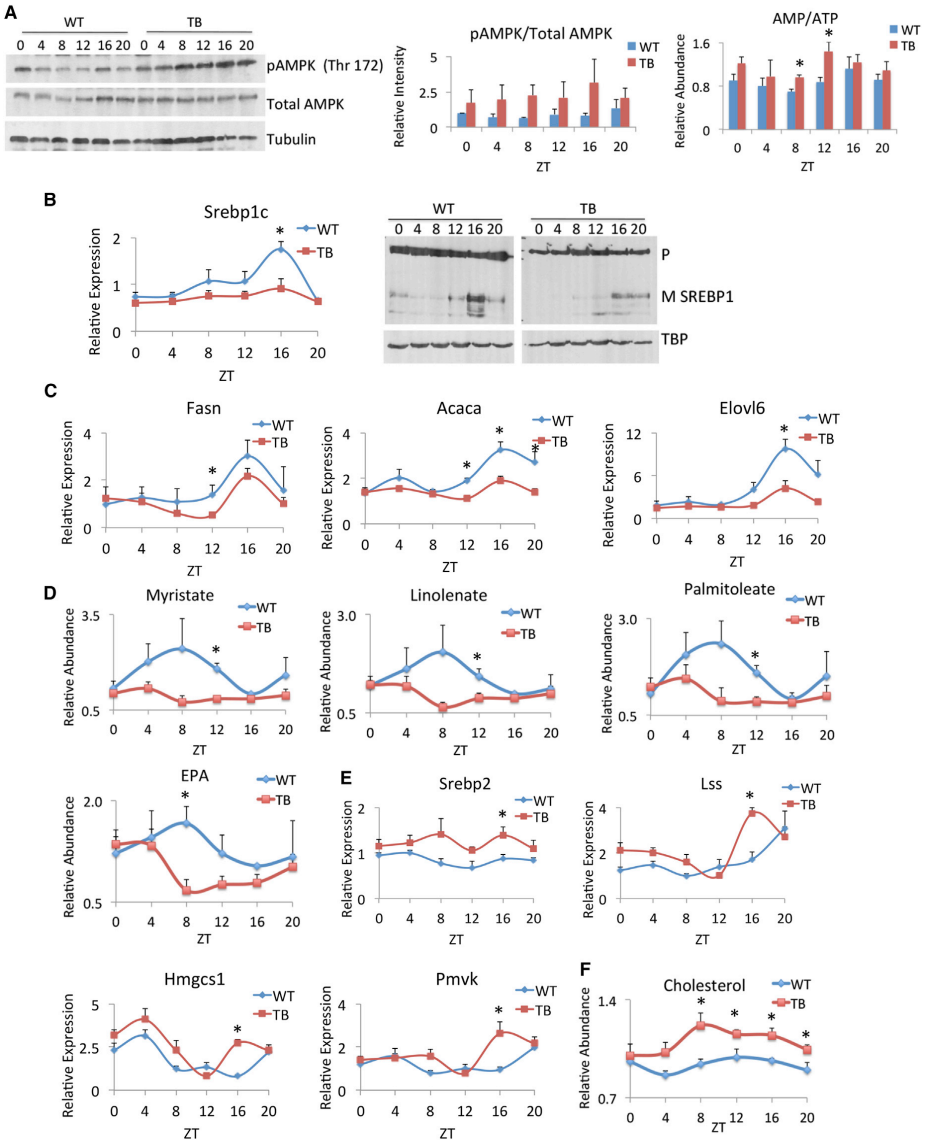


Figure 4. Lipid Metabolism Is Altered in TB Mice

A) Western blot analysis for phospho-AMPK (Thr 172) and total AMPK in WT and TB mice, at the indicated circadian times. Quantification of pAMPK/total AMPK signaling is shown as a histogram. The ratio of AMP/ATP is shown over the circadian cycle and is elevated at all ZTs.

B) Gene expression as determined by quantitative real-time PCR and protein expression of SREBP1c in WT and TB mice. Precursor (P) indicates uncleaved protein and mature (M) shows cleaved SREBP1 protein.

C) Gene expression by quantitative real-time PCR was performed for *Fasn, Acaca,* and *Elovl6* in WT and TB mice over the indicated ZTs.

(legend continued on next page)

and an induction of cholesterol biosynthesis. Repressed fatty acid levels and elevated cholesterol biosynthesis suggests a preferential shunting of lipids to produce cholesterol. Importantly, increased cholesterol levels are associated with a heightened inflammatory response (Ma et al., 2008; Zhao et al., 2011), and recent evidence links sustained inflammation with hepatic glucose production through the mevalonate pathway (Okin and Medzhitov, 2016). These findings reveal a coordinated disruption in metabolic homeostasis in TB mice, which converge to activate AMPK in the liver and thereby differentially modulate SREBP-dependent lipid signaling. Thus, these results demonstrate that while the hepatic core clock is resilient to the distal effects of lung adenocarcinoma (Figure 3), the liver metabolic clock is altered in response to tumors.

Tumor-Driven Targeting of Circadian Inflammatory Response

Pro-inflammatory responses mediated by tumor-secreted cytokines and chemokines are critical in cancer initiation and progression (Grivennikov et al., 2010). Specifically, the janus kinase (JAK)/signal transducer and activator of transcription (STAT) pathway has been demonstrated to play a role in multiple types of cancer (Lesina et al., 2011; Michalaki et al., 2004; Yu et al., 2014) and is known to be activated by IL-6, tumor necrosis factor alpha (TNFα), interferon gamma (IFNγ), and leukemia inhibitory factor (LIF) (Darnell et al., 1994; Fitzgerald et al., 2005; Grivennikov et al., 2009; Guo et al., 1998). Also, inflammation plays an important role in progression of lung adenocarcinoma (Gao et al., 2007; Yeh et al., 2006). Therefore, to validate the inflammatory response in TB mice, serum cytokine levels were assessed in an unbiased, multiplexed platform at ZT 12, which is the reported peak of circadian inflammatory response (Gibbs et al., 2012). Of the 31 cytokines assayed, a number were increased, decreased, or unchanged in TB mouse serum versus WT and shown in a heat map (Figure 5A). Specifically, significant elevation of IL-6 was observed in TB mouse serum, along with a clear, though statistically non-significant, increase in IL-1α, TNFα, LIF, and IFNγ (Figures 5A and S7). In contrast, serum levels of the anti-inflammatory cytokine IL-10 and its receptor expression in the liver did not change between WT and TB mice (Figures 5A and S7). Importantly, concomitant gene expression profiles of the cytokine receptors, interleukin 6 receptor (*Il6r*α), interleukin 1 receptor (*Il1r1*), tumor necrosis factor receptor subfamily (*Tnfrsf1b*), and interleukin 17 receptor (*Il17ra*) were significantly elevated in the liver and displayed circadian profiles that peaked at ZT12 in TB mice (Figures 5B and S7). We focused on the effects of IL6-dependent signaling as a paradigm of how signals in TB mice lead to the expression and phosphorylation of STAT3. Gene expression of *Stat3* was significantly elevated at ZT 8, ZT 12, and ZT 16, along with a corresponding increase in total protein levels in TB mice (Figure 5C), in keeping with reported IL-6-dependent STAT3 auto-regulation (Narimatsu et al., 2001). JAK-dependent phosphorylation of STAT3 at tyrosine (Tyr) 705 is known to activate STAT3 and induce its transcriptional activity by nuclear translocation (Darnell, 1997). We observed marked elevation of p-STAT3 Tyr705 in the livers of TB mice that peaked at ZT8 and ZT12 (Figure 5C). The transcriptional activation of STAT3 resulted in a significant increase in the expression of its downstream targets. Specifically, gene expression of suppressor of cytokine signaling 3 (*Socs3*) was drastically elevated and peaked at ZT 8 in TB mice, while the expression of *Socs1* and *Socs7* was unchanged (Figure 5D). Our transcriptomics data were compared to known STAT3 target genes (Bonetto et al., 2011) to determine the extent of STAT transcriptional activation in the liver. Heat maps display genes that were differentially regulated in TB normalized to WT, and of these genes, an enrichment was observed in the TB-specific gene set versus WT (Figure 5E; Table S1). Finally, there is a time-specific, significant increase of transcriptionally active p-STAT3 in TB mice at the STAT binding element (SBE), as demonstrated by chromatin immunoprecipitation (ChIP), on the *Socs3* promoter (Figure 5F).

These results demonstrate that the pro-inflammatory response can induce transcriptional activation of STAT3 signaling in the liver, which may play a role in the hepatic metabolic rewiring observed in TB mice. Yet, this transcriptional rewiring observed in the liver is representative of a localized response, as the WAT and muscle gene expression profiles differ (Figure S6), in keeping with a tissue-specific inflammatory response. Moreover, the pro-inflammatory response is most likely not the only cause of metabolic rewiring: our preliminary results of the circadian serum metabolome from WT and TB mice show that a number of factors could be involved in tumor-dependent crosstalk with peripheral tissues (Figure S7). These data suggest that the extent of the tumor macroenvironment remains inadequately defined and that complex tissue-specific responses to these tumor-derived signaling molecules exist.

Lung Cancer Alters Hepatic Insulin Signaling and Glucose Production

SOCS3 has been shown to play a role in modulating insulin sensitivity in adipose tissue and liver, and these effects have been linked with IL-6- or TNFα-mediated inflammation (Emanuelli et al., 2001; Sachithanandan et al., 2010; Senn et al., 2003; Torisu et al., 2007). Given the induction of the STAT3 inflammatory axis and the increase in *Socs3* gene expression in TB mice, hepatic insulin signaling was further investigated. Insulin-dependent phosphorylation of V-Akt murine thymoma viral oncogene homolog (AKT) at serine (Ser) 473 was dramatically inhibited in TB mice versus WT, while total levels of AKT remained unchanged (Figure 6A). Strikingly, the total levels of insulin receptor substrate 1 (IRS-1) protein were noticeably decreased in TB mice versus WT, especially from ZT8 to ZT20 (Figure 6A). Notably, these changes in IRS-1 protein levels coincided with the peak in STAT3 activation and *Socs3* expression (Figures 5C and 5D), as SOCS3 is reported to target and degrade IRS-1

D) Levels of fatty acids and fatty acid esters as determined by metabolomics analysis for myristate, linolenate, palmitoleate, and eicosapentaenoate (EPA).
E) Gene expression of *Srebp2* and its target genes *Lss, Hmgcs1*, and *Pmvk* as determined by quantitative real-time PCR.
F) Total cholesterol levels in WT and TB mice over the circadian cycle were determined by metabolomics analysis. Error bars indicate SEM. Significance was calculated using Student's t test, and * indicates a p value cutoff of 0.05.

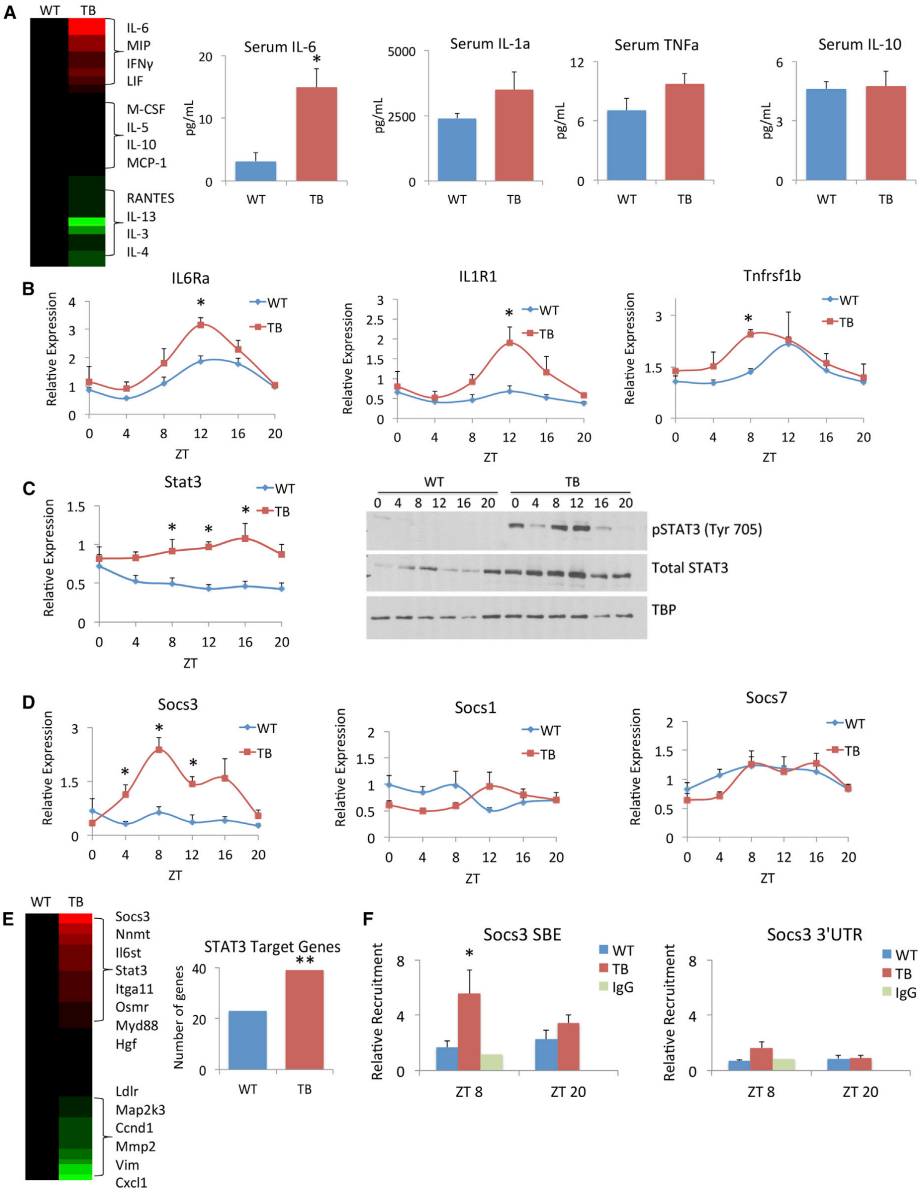


Figure 5. Lung-Tumor-Induced Inflammation in the Liver

A) Serum samples from WT and TB mice were assayed in an unbiased, multiplexed cytokine array platform and displayed as a heat map. Red indicates increased levels of cytokines, and green indicates decreased cytokine levels in TB normalized to WT mouse serum. Specific profiles of pro-inflammatory cytokines (IL-6, IL-1α, and TNFα) and anti-inflammatory IL-10 are shown.

B) Gene expression as profiled by quantitative real-time PCR is shown for *Il6rα*, *Il1r1*, and *Tnfrsf1b*.

(legend continued on next page)

protein and thereby further repress hepatic insulin signaling (Rui et al., 2002). Serum insulin levels were investigated, and in line with decreased insulin signaling in the liver, systemic insulin levels were significantly low and lose their circadian oscillation in TB mice (Figure 6B). Yet, TB animals retained insulin sensitivity, as determined by insulin tolerance tests (ITT) (Figure 6C). Given these changes in insulin signaling, it would be expected that TB mice exhibit elevated levels of serum glucose. Indeed, TB mice displayed elevated fasting serum glucose levels (Figure 6D), and these mice were significantly less sensitive to exogenous glucose challenge using a glucose tolerance test (GTT) (Figure 6E). In order to elucidate the mechanism by which glucose levels were increased in TB mice, hepatic glucose production through gluconeogenesis was investigated, as this pathway is known to be clock-controlled (Zhang et al., 2010). Gene expression of phosphoenolpyruvate carboxykinase 1 (*Pck1* or Pepck) remained circadian but was significantly induced at ZT 8, ZT12, and ZT16 in TB mice (Figure 6F). The level of phosphoenolpyruvate (PEP), the product of PEPCK, was elevated at ZT 8 and ZT 12 (Figure 6G). In contrast, expression of key glycolytic enzymes, such as rate-limiting glucokinase (*Gck*) and liver pyruvate kinase (*Pklr*), was significantly inhibited (Figure 6H). Though their expression was not circadian, lactate dehydrogenases (*Ldha* and *Ldhc*) that interconvert lactate and pyruvate were elevated in TB mice (Figure 6I), in keeping with the increased pyruvate levels that could be shunted into gluconeogenesis (Figure 6J). These results demonstrate that a lung tumor is responsible for the drastic change in insulin-dependent AKT signaling in the liver, leading to significant alterations in clock-controlled hepatic glucose production.

DISCUSSION

Lung Tumor as an Endogenous Zeitgeber?

Circadian homeostasis is essential for organismal physiology and its intrinsic plasticity constitutes a highly efficient adaptation system to the changing environment. Specifically, zeitgebers such as light and nutrition are referred to as external stimuli that operate to entrain central and peripheral clocks, respectively. Here we have reported on findings that identify an endogenous circadian reorganizer that has the unique feature of rewiring circadian metabolism under unaltered light and feeding conditions. Indeed, lung adenocarcinoma contributes to the distal reprogramming of circadian hepatic gene expression and metabolic function (Figure 7). These results collectively demonstrate that lung tumors, independently of any nutritional challenge paradigm, alter circadian physiology leading to changes in cyclic energy expenditure, lipid metabolism, and hepatic insulin and glucose signaling. Importantly, since the core components of the liver molecular clock do not seem to be influenced in TB mice (Figure 3), the tumor does not appear to function as a classical zeitgeber. Yet, lung tumors act prominently on the liver by profoundly rewiring circadian metabolic control. Thus, these findings illustrate that circadian metabolism can be reprogrammed independently of exogenous inputs, such as the classical zeitgebers, light and nutrition. Indeed, lung adenocarcinoma operates as a distinctive ECO that dictates the changing pathophysiological dimension of a distal tissue such as the liver. We speculate that an ECO may function in differential manners depending on the tissue type, range of action, and metabolic state.

Endogenous Circadian Organizer

Given that these metabolic effects are systemic, tumor-dependent rewiring is most likely taking place in multiple organs to disrupt homeostasis. In this context it is notable that the effects of IL-6 are pleiotropic and function in a context- and tissue-specific manner to alter multiple signaling pathways (Mauer et al., 2015) and that the inflammatory response is the likely mediator of a complex web of physiological adjustments. Indeed, it does not escape our attention that the effect of a lung tumor will not be restricted to the liver alone (Figure S6), although this tissue is critical to decipher the effects of lung cancer on organismal metabolism. In further support of our results, a number of intriguing links can be made to connect tumor-derived inflammation with deregulated metabolism. Our extensive analysis demonstrates that a number of cytokines could be implicated in the rewiring observed in TB mice (Figures 5A and S7). Also, our preliminary analysis of the serum metabolome indicates that the cyclic profiles of a variety of potentially critical metabolites change significantly in the TB mice (Figure S7). We have focused on IL-6 as a paradigm since the IL-6 inflammatory response has been investigated in fatty liver disease, and these effects link inflammation to altered lipid accumulation through SREBP signaling (Miller et al., 2011; Yamaguchi et al., 2010). Also, another layer of complexity exists in that IL-6 is known to activate AMPK especially in the context of exercise (Carey et al., 2006; Ruderman et al., 2006) and systemic IL-6 knockout mice are unable to stimulate AMPK signaling (Adser et al., 2011; Kelly et al., 2004). Moreover, AMPK suppresses SREBP target gene expression and attenuates hepatic steatosis (Li et al., 2011). Collectively, these notions suggest an interconnected network between inflammatory cytokines, AMPK, and SREBP that could contribute to the tumor-induced liver reprogramming we observe. Moreover, the influence of cachexia on muscle and WAT is another factor that can feedback and alter liver homeostasis (Bonetto et al., 2011; Narsale et al., 2015; Tsoli et al., 2014). Our findings

C) *Stat3* gene expression as shown by quantitative real-time PCR. Phospho-STAT3 (Tyr 705) and total STAT3 protein levels over the circadian cycle in WT and TB mice.

D) Gene expression profiles of *Socs3, Socs1*, and *Socs7* by quantitative real-time PCR.

E) Known STAT3 target genes were compared to our transcriptomics data. Heat map displays gene expression profiles in the TB-specific group normalized to WT, with red and green representing up- and down-regulated genes, respectively. Additionally, to determine enrichment of STAT3 target genes in WT and TB, Fisher's exact test was used. The ** indicates the odds that the overlap of 39 genes in TB over random is 1.716 and the p value is 0.00358, which satisfies a $p < 0.005$ threshold. For WT, the odds ratio is 1.345 with a p value of 0.209.

F) Recruitment of p-STAT3 to the STAT binding element (SBE) in the *Socs3* promoter or to the 3′ UTR as determined by chromatin immunoprecipitation (ChIP). Error bars indicate SEM. Significance was calculated using Student's t test, and * indicates a p value cutoff of 0.05.

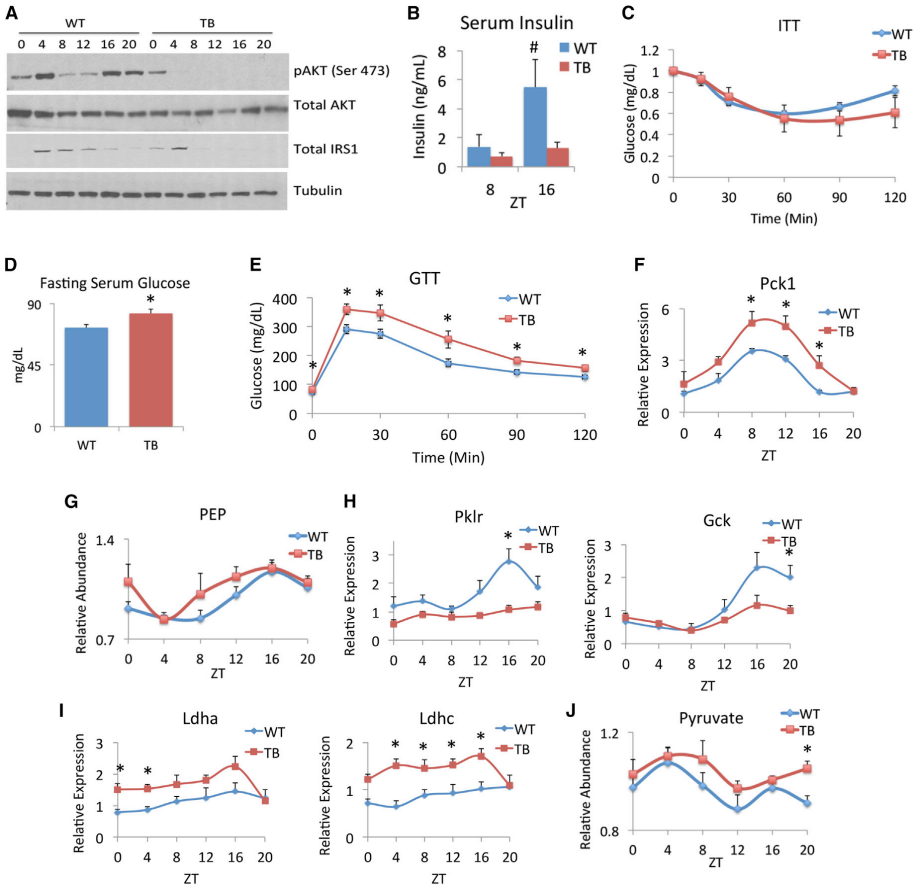


Figure 6. Lung Adenocarcinoma Alters Hepatic Insulin Signaling and Glucose Production

A) Western analysis of phospho-AKT (Ser 473), total AKT, and total IRS1 in WT and TB mice over the circadian cycle.

B) Serum insulin levels were measured by ELISA at ZT 8 and 16 in WT and TB mouse serum. Insulin levels at ZT 16 are statistically significant as indicated by # (p value = 0.053, using Student's t test).

C) Insulin tolerance test (ITT) in WT and TB mice.

D) Overnight fasting glucose levels in WT and TB mice.

E) Glucose tolerance test (GTT) in overnight-fasted WT and TB mice.

F) Gluconeogenic gene expression profile of Pepck (*Pck1*) by quantitative real-time PCR was done in livers of WT and TB mice.

G) Levels of phosphoenolpyruvate (PEP) were determined by metabolomics analysis from livers of WT and TB mice.

H) Quantitative real-time PCR of glycolytic gene expression of L-PK (*Pklr*) and GK (*Gck*) over the circadian cycle.

I) Gene expression of lactate dehydrogenases *Ldha* and *Ldhc* in WT and TB mice by quantitative real-time PCR.

J) Levels of pyruvate were determined by metabolomics in livers of WT and TB mice. Error bars indicate SEM. Significance was calculated using Student's t test and * indicates a p value cutoff of 0.05.

identify two converging pathways that might work in a coordinated manner to modulate circadian hepatic homeostasis. Altered energy expenditure can mediate a circadian metabolic rewiring that is likely compounded by the pro-inflammatory effects on the liver.

The Tumor Macroenvironment Reorganizes Homeostasis

Our results illustrate that lung adenocarcinoma has a profound effect on a variety of metabolic and signaling pathways in the liver. This tumor-derived macroenvironment is constituted by

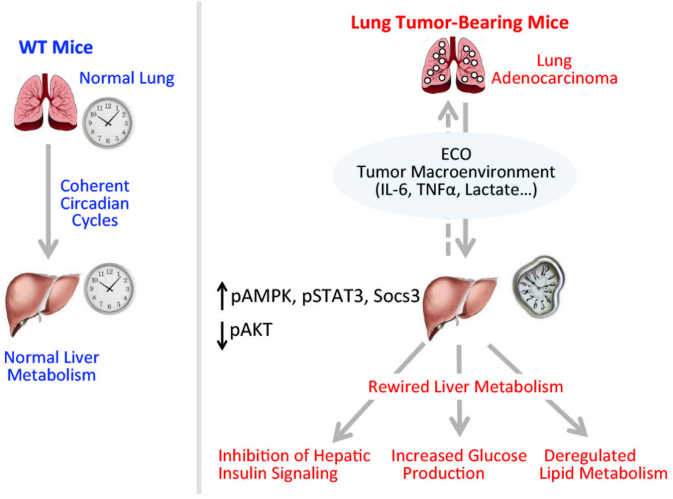


Figure 7. Lung Adenocarcinoma Distally Rewires Circadian Hepatic Metabolism

Schematic overview depicting the effects of the tumor macroenvironment on circadian hepatic metabolism. Our results show that lung tumors, acting through the inflammatory STAT3-Socs3 axis, operate to distally rewire circadian transcription and metabolism by acting as an endogenous circadian organizer (ECO). This manifests in loss of hepatic insulin signaling, glucose intolerance, and deregulated lipid metabolism through the AMPK/SREBP pathway.

glycolytic metabolic byproducts, inflammatory cytokines, and other poorly defined circulating components (Al-Zoughbi et al., 2014). In this context, we have explored the liver pro-inflammatory response, a paradigm of the metabolic rewiring induced by the lung tumor. An additional twist to this scenario is the likely contribution of other metabolic tissues that respond to the tumor-derived macroenvironment. In turn, these tissues could provide an additional layer of signaling that would further adjust circadian homeostasis.

The pro-inflammatory response alters hepatic insulin signaling and subsequent glucose production through the STAT3-dependent activation of *Socs3* (Figure 5). These findings bring to light another critical pathway that is deregulated by the distant actions of the lung tumor. The suppression of serum insulin levels suggests an important role of inflammation in the pancreas. Yet, these effects on hepatic insulin sensitivity are most likely regulated coordinately with the inflammatory response. For instance, the expression of the insulin-responsive *Insig2* gene, which is known to be involved in clock-controlled SREBP function (Le Martelot et al., 2009), is unaltered in TB mice and suggests a complex signaling mechanism beyond a simple model of hypoinsulinemia. Moreover, the circadian clock controls hepatic glucose production through gluconeogenesis (Zhang et al., 2010) and our results suggest a potential crosstalk between the tumor and the liver clock. We speculate that tumor-secreted "waste" such as lactate is converted to pyruvate and shunted through gluconeogenesis to produce glucose, which can further satisfy the heightened energetic demand of cancer cells. Interestingly, TB mice exhibit increased expression of lactate dehydrogenases (*Ldha, Ldhc*), Pepck (*Pck1*), and PEP, which could result in enhanced glucose intolerance (Figure 6). In conclusion, the circadian clock is highly responsive to its environment and is able to adapt to changes in energetic demand. In this context, the lung tumor macroenvironment operates as an ECO on hepatic circadian metabolism—a process that could potentially further drive tumorigenesis.

EXPERIMENTAL PROCEDURES

Animal Housing and Experimental Procedures

Kras$^{LSL-G12D}$;*p53*$^{fl/fl}$ mice have been previously described (Johnson et al., 2001). Detailed housing and infection procedures are provided as Supplemental Experimental Procedures.

DNA Microarray Analysis

Microarray analysis was performed as previously described (Masri et al., 2014), and further information is provided as Supplemental Experimental Procedures.

Metabolomics Analysis

Metabolomics analysis was carried out by Metabolon (Durham) as previously described (Evans et al., 2009; Masri et al., 2014). See the Supplemental Experimental Proceduresfor further details.

Bioinformatics and Pathway Analysis

Bioinformatics analysis was performed using JTK_cycle, and metabolomics and transcriptomics data are accessible at circadiomics.ics.uci.edu. Detailed methodology is available in the Supplemental Experimental Procedures.

Metabolic Cage Analysis

Indirect calorimetry was performed using negative-flow CLAMS hardware system cages (Columbus Instruments). VO_2, VCO_2, RER, and food intake were measured and calculated with Oxymax software (Columbus Instruments).

Locomotor Activity Analysis

Animals were individually housed, using an n = 12 mice/genotype for behavioral analysis. Mice were housed for 2 weeks in 12 hr standard L/D conditions and subsequently released into D/D conditions for 2 weeks. Activity was measured using optical beam motion detection (Philips Respironics) and data analyzed using Minimitter VitalView data acquisition software.

Gene Expression Analysis

Detailed methodology and primer sequences can be found in the Supplemental Experimental Procedures section.

Western Blot Analysis

Livers were homogenized in RIPA lysis buffer containing protease inhibitor cocktail, NaF and PMSF, sonicated briefly, and rocked to lyse cells at 4°C. 10–30 μg of protein lysate was resolved on SDS-PAGE gels. Antibodies used for western blots include: TBP, BMAL1, SREBP1 (Abcam), pAMPK, total

AMPK, pSTAT3, total STAT3, pAKT, total AKT, and IRS1 (Cell Signaling Technology).

Cytokine Profiling

A total of 31 mouse cytokines were profiled using a multiplex platform, and data were extracted based on cytokine-specific standards by Eve Technologies (Calgary). Five independent serum samples were used from WT and TB mice. Relative change in cytokine expression between TB and WT was determined using absolute deviation values from the median and then used for heat map generation.

Chromatin Immunoprecipitation

Chromatin immunoprecipitation (ChIP) methodology was previously described (Masri et al., 2014). The pSTAT3 antibody used for ChIP was obtained from Cell Signaling technology.

Tolerance Tests: GTT and ITT

8 WT and 8 TB mice were fasted overnight, and fasting glucose levels were measured using an ACCU-CHEK Aviva Plus glucometer (Roche). Body weight measurements were taken, insulin (0.75U/kg) or glucose (2g/kg) was intraperitoneally injected, and blood glucose measurements were taken at 15, 30, 60, 90, and 120 min post injection.

ACCESSION NUMBERS

The accession number for the microarray data reported in this paper is GEO: GSE73222.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2016.04.039.

AUTHOR CONTRIBUTION

S.M., T.P., and P.S.-C. conceptually designed the study. T.P. and T.J. provided the mouse model. S.M., T.P., K.K., and M.C. performed experiments. Y.L. and P.B. performed bioinformatics analysis. S.M. and P.S.-C. wrote the paper with input from all authors.

ACKNOWLEDGMENTS

We thank all members of the Sassone-Corsi laboratory for scientific discussion or technical assistance and T. Leff, G. Servillo, and M. Oakes for discussions. Funding for S.M. was provided by NIH post-doctoral fellowship GM097899. K.K. was supported by a post-doctoral fellowship from the Japan Society for the Promotion of Science (JSPS). T.P. was supported by the Hope Funds for Cancer Research. Financial support for T.J. was provided by the Ludwig Center for Molecular Oncology at MIT and by the Koch Institute Frontier Fund. Grants from the NSF (IIS-1321053) and the NIH (LM010235) to P.B. supported the work of Y.L. and P.B. Financial support for P.S.-C. was provided by NIH (AG043745), Merieux Research Grant (53923), and the University of California Irvine Chao Family Comprehensive Cancer Center.

Received: October 21, 2015
Revised: January 2, 2016
Accepted: April 12, 2016
Published: May 5, 2016

REFERENCES

Adser, H., Wojtaszewski, J.F., Jakobsen, A.H., Kiilerich, K., Hidalgo, J., and Pilegaard, H. (2011). Interleukin-6 modifies mRNA expression in mouse skeletal muscle. Acta Physiol. (Oxf.) *202*, 165–173.

Al-Zoughbi, W., Huang, J., Paramasivan, G.S., Till, H., Pichler, M., Guertl-Lackner, B., and Hoefler, G. (2014). Tumor macroenvironment and metabolism. Semin. Oncol. *41*, 281–295.

Asher, G., and Sassone-Corsi, P. (2015). Time for food: the intimate interplay between nutrition, metabolism, and the circadian clock. Cell *161*, 84–92.

Bass, J. (2012). Circadian topology of metabolism. Nature *491*, 348–356.

Bonetto, A., Aydogdu, T., Kunzevitzky, N., Guttridge, D.C., Khuri, S., Koniaris, L.G., and Zimmers, T.A. (2011). STAT3 activation in skeletal muscle links muscle wasting and the acute phase response in cancer cachexia. PLoS ONE *6*, e22538.

Brown, M.S., and Goldstein, J.L. (1997). The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. Cell *89*, 331–340.

Carey, A.L., Steinberg, G.R., Macaulay, S.L., Thomas, W.G., Holmes, A.G., Ramm, G., Prelovsek, O., Hohnen-Behrens, C., Watt, M.J., James, D.E., et al. (2006). Interleukin-6 increases insulin-stimulated glucose disposal in humans and glucose uptake and fatty acid oxidation in vitro via AMP-activated protein kinase. Diabetes *55*, 2688–2697.

Colegio, O.R., Chu, N.Q., Szabo, A.L., Chu, T., Rhebergen, A.M., Jairam, V., Cyrus, N., Brokowski, C.E., Eisenbarth, S.C., Phillips, G.M., et al. (2014). Functional polarization of tumour-associated macrophages by tumour-derived lactic acid. Nature *513*, 559–563.

Damiola, F., Le Minh, N., Preitner, N., Kornmann, B., Fleury-Olela, F., and Schibler, U. (2000). Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. Genes Dev. *14*, 2950–2961.

Darnell, J.E., Jr. (1997). STATs and gene regulation. Science *277*, 1630–1635.

Darnell, J.E., Jr., Kerr, I.M., and Stark, G.R. (1994). Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. Science *264*, 1415–1421.

Doherty, J.R., and Cleveland, J.L. (2013). Targeting lactate metabolism for cancer therapeutics. J. Clin. Invest. *123*, 3685–3692.

Eckel-Mahan, K.L., Patel, V.R., de Mateo, S., Orozco-Solis, R., Ceglia, N.J., Sahar, S., Dilag-Penilla, S.A., Dyar, K.A., Baldi, P., and Sassone-Corsi, P. (2013). Reprogramming of the circadian clock by nutritional challenge. Cell *155*, 1464–1478.

Emanuelli, B., Peraldi, P., Filloux, C., Chavey, C., Freidinger, K., Hilton, D.J., Hotamisligil, G.S., and Van Obberghen, E. (2001). SOCS-3 inhibits insulin signaling and is up-regulated in response to tumor necrosis factor-alpha in the adipose tissue of obese mice. J. Biol. Chem. *276*, 47944–47949.

Evans, A.M., DeHaven, C.D., Barrett, T., Mitchell, M., and Milgram, E. (2009). Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. Anal. Chem. *81*, 6656–6667.

Fitzgerald, J.S., Tsareva, S.A., Poehlmann, T.G., Berod, L., Meissner, A., Corvinus, F.M., Wiederanders, B., Pfitzner, E., Markert, U.R., and Friedrich, K. (2005). Leukemia inhibitory factor triggers activation of signal transducer and activator of transcription 3, proliferation, invasiveness, and altered protease expression in choriocarcinoma cells. Int. J. Biochem. Cell Biol. *37*, 2284–2296.

Fu, L., and Lee, C.C. (2003). The circadian clock: pacemaker and tumour suppressor. Nat. Rev. Cancer *3*, 350–361.

Gamble, K.L., Berry, R., Frank, S.J., and Young, M.E. (2014). Circadian clock control of endocrine factors. Nat. Rev. Endocrinol. *10*, 466–475.

Gao, S.P., Mark, K.G., Leslie, K., Pao, W., Motoi, N., Gerald, W.L., Travis, W.D., Bornmann, W., Veach, D., Clarkson, B., and Bromberg, J.F. (2007). Mutations in the EGFR kinase domain mediate STAT3 activation via IL-6 production in human lung adenocarcinomas. J. Clin. Invest. *117*, 3846–3856.

Gibbs, J.E., Blaikley, J., Beesley, S., Matthews, L., Simpson, K.D., Boyce, S.H., Farrow, S.N., Else, K.J., Singh, D., Ray, D.W., and Loudon, A.S. (2012). The nuclear receptor REV-ERBα mediates circadian regulation of innate immunity through selective regulation of inflammatory cytokines. Proc. Natl. Acad. Sci. USA *109*, 582–587.

Gilardi, F., Migliavacca, E., Naldi, A., Baruchet, M., Canella, D., Le Martelot, G., Guex, N., and Desvergne, B.; CycliX Consortium (2014). Genome-wide analysis of SREBP1 activity around the clock reveals its combined dependency on nutrient and circadian signals. PLoS Genet. *10*, e1004155.

Grivennikov, S., Karin, E., Terzic, J., Mucida, D., Yu, G.Y., Vallabhapurapu, S., Scheller, J., Rose-John, S., Cheroutre, H., Eckmann, L., and Karin, M. (2009). IL-6 and Stat3 are required for survival of intestinal epithelial cells and development of colitis-associated cancer. Cancer Cell *15*, 103–113.

Grivennikov, S.I., Greten, F.R., and Karin, M. (2010). Immunity, inflammation, and cancer. Cell *140*, 883–899.

Guo, D., Dunbar, J.D., Yang, C.H., Pfeffer, L.M., and Donner, D.B. (1998). Induction of Jak/STAT signaling by activation of the type 1 TNF receptor. J. Immunol. *160*, 2742–2750.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674.

Hatori, M., Vollmers, C., Zarrinpar, A., DiTacchio, L., Bushong, E.A., Gill, S., Leblanc, M., Chaix, A., Joens, M., Fitzpatrick, J.A., et al. (2012). Time-restricted feeding without reducing caloric intake prevents metabolic diseases in mice fed a high-fat diet. Cell Metab. *15*, 848–860.

Horton, J.D., Shah, N.A., Warrington, J.A., Anderson, N.N., Park, S.W., Brown, M.S., and Goldstein, J.L. (2003). Combined analysis of oligonucleotide microarray data from transgenic and knockout mice identifies direct SREBP target genes. Proc. Natl. Acad. Sci. USA *100*, 12027–12032.

Hsu, P.P., and Sabatini, D.M. (2008). Cancer cell metabolism: Warburg and beyond. Cell *134*, 703–707.

Jackson, E.L., Willis, N., Mercer, K., Bronson, R.T., Crowley, D., Montoya, R., Jacks, T., and Tuveson, D.A. (2001). Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. Genes Dev. *15*, 3243–3248.

Jackson, E.L., Olive, K.P., Tuveson, D.A., Bronson, R., Crowley, D., Brown, M., and Jacks, T. (2005). The differential effects of mutant p53 alleles on advanced murine lung cancer. Cancer Res. *65*, 10280–10288.

Johnson, L., Mercer, K., Greenbaum, D., Bronson, R.T., Crowley, D., Tuveson, D.A., and Jacks, T. (2001). Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. Nature *410*, 1111–1116.

Kelly, M., Keller, C., Avilucea, P.R., Keller, P., Luo, Z., Xiang, X., Giralt, M., Hidalgo, J., Saha, A.K., Pedersen, B.K., and Ruderman, N.B. (2004). AMPK activity is diminished in tissues of IL-6 knockout mice: the effect of exercise. Biochem. Biophys. Res. Commun. *320*, 449–454.

Kohsaka, A., Laposky, A.D., Ramsey, K.M., Estrada, C., Joshu, C., Kobayashi, Y., Turek, F.W., and Bass, J. (2007). High-fat diet disrupts behavioral and molecular circadian rhythms in mice. Cell Metab. *6*, 414–421.

Le Martelot, G., Claudel, T., Gatfield, D., Schaad, O., Kornmann, B., Lo Sasso, G., Moschetta, A., and Schibler, U. (2009). REV-ERBalpha participates in circadian SREBP signaling and bile acid homeostasis. PLoS Biol. *7*, e1000181.

Lesina, M., Kurkowski, M.U., Ludes, K., Rose-John, S., Treiber, M., Klöppel, G., Yoshimura, A., Reindl, W., Sipos, B., Akira, S., et al. (2011). Stat3/Socs3 activation by IL-6 transsignaling promotes progression of pancreatic intraepithelial neoplasia and development of pancreatic cancer. Cancer Cell *19*, 456–469.

Li, Y., Xu, S., Mihaylova, M.M., Zheng, B., Hou, X., Jiang, B., Park, O., Luo, Z., Lefai, E., Shyy, J.Y., et al. (2011). AMPK phosphorylates and inhibits SREBP activity to attenuate hepatic steatosis and atherosclerosis in diet-induced insulin-resistant mice. Cell Metab. *13*, 376–388.

Lin, W.W., and Karin, M. (2007). A cytokine-mediated link between innate immunity, inflammation, and cancer. J. Clin. Invest. *117*, 1175–1183.

Ma, K.L., Ruan, X.Z., Powis, S.H., Chen, Y., Moorhead, J.F., and Varghese, Z. (2008). Inflammatory stress exacerbates lipid accumulation in hepatic cells and fatty livers of apolipoprotein E knockout mice. Hepatology *48*, 770–781.

Masri, S., and Sassone-Corsi, P. (2010). Plasticity and specificity of the circadian epigenome. Nat. Neurosci. *13*, 1324–1329.

Masri, S., Rigor, P., Cervantes, M., Ceglia, N., Sebastian, C., Xiao, C., Roqueta-Rivera, M., Deng, C., Osborne, T.F., Mostoslavsky, R., et al. (2014). Partitioning circadian transcription by SIRT6 leads to segregated control of cellular metabolism. Cell *158*, 659–672.

Masri, S., Kinouchi, K., and Sassone-Corsi, P. (2015). Circadian clocks, epigenetics, and cancer. Curr. Opin. Oncol. *27*, 50–56.

Mauer, J., Denson, J.L., and Brüning, J.C. (2015). Versatile functions for IL-6 in metabolism and cancer. Trends Immunol. *36*, 92–101.

Michalaki, V., Syrigos, K., Charles, P., and Waxman, J. (2004). Serum levels of IL-6 and TNF-alpha correlate with clinicopathological features and patient survival in patients with prostate cancer. Br. J. Cancer *90*, 2312–2316.

Miller, A.M., Wang, H., Bertola, A., Park, O., Horiguchi, N., Ki, S.H., Yin, S., Lafdil, F., and Gao, B. (2011). Inflammation-associated interleukin-6/signal transducer and activator of transcription 3 activation ameliorates alcoholic and nonalcoholic fatty liver diseases in interleukin-10-deficient mice. Hepatology *54*, 846–856.

Narimatsu, M., Maeda, H., Itoh, S., Atsumi, T., Ohtani, T., Nishida, K., Itoh, M., Kamimura, D., Park, S.J., Mizuno, K., et al. (2001). Tissue-specific autoregulation of the stat3 gene and its role in interleukin-6-induced survival signals in T cells. Mol. Cell. Biol. *21*, 6615–6625.

Narsale, A.A., Enos, R.T., Puppa, M.J., Chatterjee, S., Murphy, E.A., Fayad, R., Pena, M.O., Durstine, J.L., and Carson, J.A. (2015). Liver inflammation and metabolic signaling in ApcMin/+ mice: the role of cachexia progression. PLoS ONE *10*, e0119888.

Okin, D., and Medzhitov, R. (2016). The Effect of Sustained Inflammation on Hepatic Mevalonate Pathway Results in Hyperglycemia. Cell *165*, 343–356.

Panda, S., Antoch, M.P., Miller, B.H., Su, A.I., Schook, A.B., Straume, M., Schultz, P.G., Kay, S.A., Takahashi, J.S., and Hogenesch, J.B. (2002). Coordinated transcription of key pathways in the mouse by the circadian clock. Cell *109*, 307–320.

Partch, C.L., Green, C.B., and Takahashi, J.S. (2014). Molecular architecture of the mammalian circadian clock. Trends Cell Biol. *24*, 90–99.

Ruderman, N.B., Keller, C., Richard, A.M., Saha, A.K., Luo, Z., Xiang, X., Giralt, M., Ritov, V.B., Menshikova, E.V., Kelley, D.E., et al. (2006). Interleukin-6 regulation of AMP-activated protein kinase. Potential role in the systemic response to exercise and prevention of the metabolic syndrome. Diabetes *55* (*Suppl 2*), S48–S54.

Rui, L., Yuan, M., Frantz, D., Shoelson, S., and White, M.F. (2002). SOCS-1 and SOCS-3 block insulin signaling by ubiquitin-mediated degradation of IRS1 and IRS2. J. Biol. Chem. *277*, 42394–42398.

Sachithanandan, N., Fam, B.C., Fynch, S., Dzamko, N., Watt, M.J., Wormald, S., Honeyman, J., Galic, S., Proietto, J., Andrikopoulos, S., et al. (2010). Liver-specific suppressor of cytokine signaling-3 deletion in mice enhances hepatic insulin sensitivity and lipogenesis resulting in fatty liver and obesity. Hepatology *52*, 1632–1642.

Sahar, S., and Sassone-Corsi, P. (2009). Metabolism and cancer: the circadian clock connection. Nat. Rev. Cancer *9*, 886–896.

Sansone, P., Storci, G., Tavolari, S., Guarnieri, T., Giovannini, C., Taffurelli, M., Ceccarelli, C., Santini, D., Paterini, P., Marcu, K.B., et al. (2007). IL-6 triggers malignant features in mammospheres from human ductal breast carcinoma and normal mammary gland. J. Clin. Invest. *117*, 3988–4002.

Senn, J.J., Klover, P.J., Nowak, I.A., Zimmers, T.A., Koniaris, L.G., Furlanetto, R.W., and Mooney, R.A. (2003). Suppressor of cytokine signaling-3 (SOCS-3), a potential mediator of interleukin-6-dependent insulin resistance in hepatocytes. J. Biol. Chem. *278*, 13740–13746.

Stokkan, K.A., Yamazaki, S., Tei, H., Sakaki, Y., and Menaker, M. (2001). Entrainment of the circadian clock in the liver by feeding. Science *291*, 490–493.

Torisu, T., Sato, N., Yoshiga, D., Kobayashi, T., Yoshioka, T., Mori, H., Iida, M., and Yoshimura, A. (2007). The dual function of hepatic SOCS3 in insulin resistance in vivo. Genes Cells *12*, 143–154.

Tsoli, M., Schweiger, M., Vanniasinghe, A.S., Painter, A., Zechner, R., Clarke, S., and Robertson, G. (2014). Depletion of white adipose tissue in cancer cachexia syndrome is associated with inflammatory signaling and disrupted circadian regulation. PLoS ONE *9*, e92966.

Vander Heiden, M.G., Cantley, L.C., and Thompson, C.B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science *324*, 1029–1033.

Vavvas, D., Apazidis, A., Saha, A.K., Gamble, J., Patel, A., Kemp, B.E., Witters, L.A., and Ruderman, N.B. (1997). Contraction-induced changes in acetyl-CoA carboxylase and 5′-AMP-activated kinase in skeletal muscle. J. Biol. Chem. *272*, 13255–13261.

Vollmers, C., Gill, S., DiTacchio, L., Pulivarthy, S.R., Le, H.D., and Panda, S. (2009). Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression. Proc. Natl. Acad. Sci. USA *106*, 21453–21458.

Yamaguchi, K., Itoh, Y., Yokomizo, C., Nishimura, T., Niimi, T., Fujii, H., Okanoue, T., and Yoshikawa, T. (2010). Blockade of interleukin-6 signaling enhances hepatic steatosis but improves liver injury in methionine choline-deficient diet-fed mice. Lab. Invest. *90*, 1169–1178.

Yeh, H.H., Lai, W.W., Chen, H.H., Liu, H.S., and Su, W.C. (2006). Autocrine IL-6-induced Stat3 activation contributes to the pathogenesis of lung adenocarcinoma and malignant pleural effusion. Oncogene *25*, 4300–4309.

Yu, H., Lee, H., Herrmann, A., Buettner, R., and Jove, R. (2014). Revisiting STAT3 signalling in cancer: new and unexpected biological functions. Nat. Rev. Cancer *14*, 736–746.

Zhang, E.E., Liu, Y., Dentin, R., Pongsawakul, P.Y., Liu, A.C., Hirota, T., Nusinow, D.A., Sun, X., Landais, S., Kodama, Y., et al. (2010). Cryptochrome mediates circadian regulation of cAMP signaling and hepatic gluconeogenesis. Nat. Med. *16*, 1152–1156.

Zhao, L., Chen, Y., Tang, R., Chen, Y., Li, Q., Gong, J., Huang, A., Varghese, Z., Moorhead, J.F., and Ruan, X.Z. (2011). Inflammatory stress exacerbates hepatic cholesterol accumulation via increasing cholesterol uptake and de novo synthesis. J. Gastroenterol. Hepatol. *26*, 875–883.

Chapter 6

Circadian Web Portals – CiracadiOmics

6.1 Introduction

The software systems described in Chapters 2-5 essentially established an ecosystem of circadian analysis software, which is capable of generating a variety of results. These results contributed in a major way to our collaborative projects with other biologists. However, we believe our results can further benefit the circadian field by providing access to the general public. **CircadiOmics** is a project built for this goal. Not only does it incorporate results generated using our systems including `igb-pipelines`, it also integrates a vast amount of publicly available circadian datasets. The result is a powerful web portal boasting the most comprehensive repository of circadian data anywhere.

The motivation, implementation and impact of CircadiOmics are detailed in the paper appended below. I am a first co-author of this paper.

Nucleic Acids Research, 2018 **1**
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

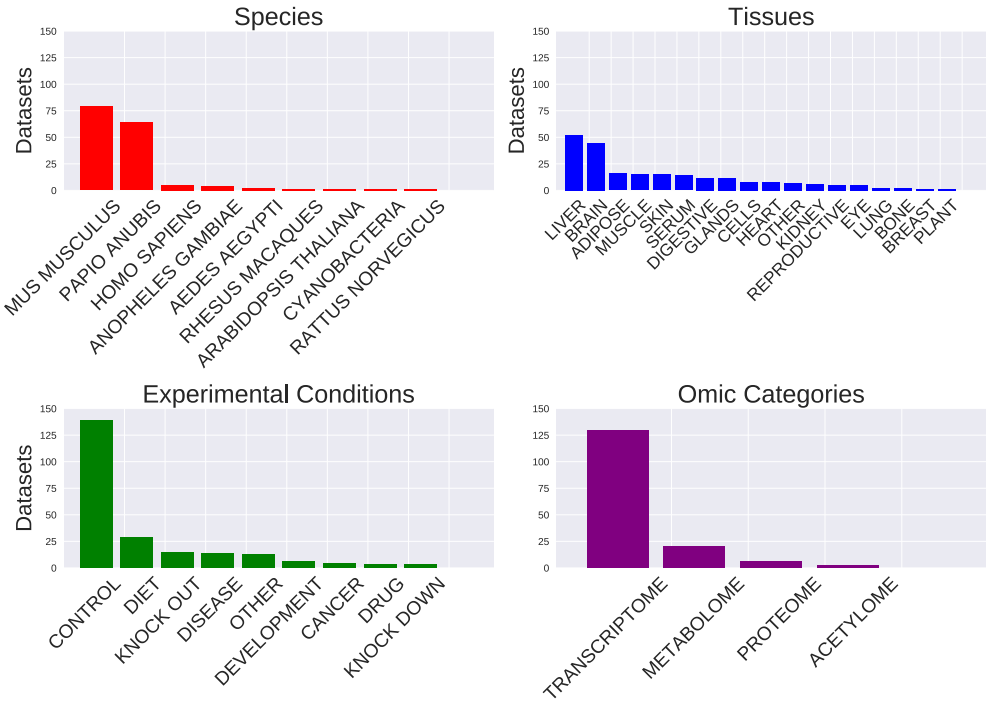


Figure 1. Dataset collection by species, tissues, experimental conditions and omic categories.

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

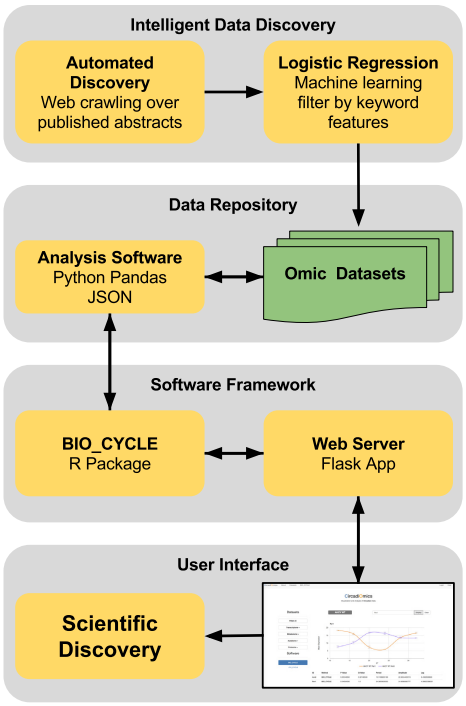


Figure 2. Three-tier Model-View-Controller architecture of the CircadiOmics web portal. Intelligent data discovery supplies candidate datasets for inclusion in the repository using a machine learning filter applied to key word features derived from web crawling published abstracts. BIO_CYCLE results 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.

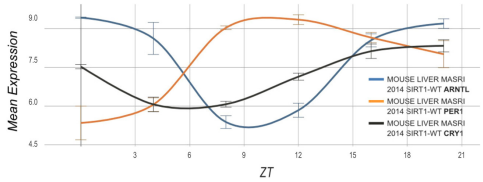


Figure 3. Visualization of queries for ARNTL, PER1 and CRY1 in a control mouse dataset. Any number of queries, across any number of datasets, can be displayed simultaneously.

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

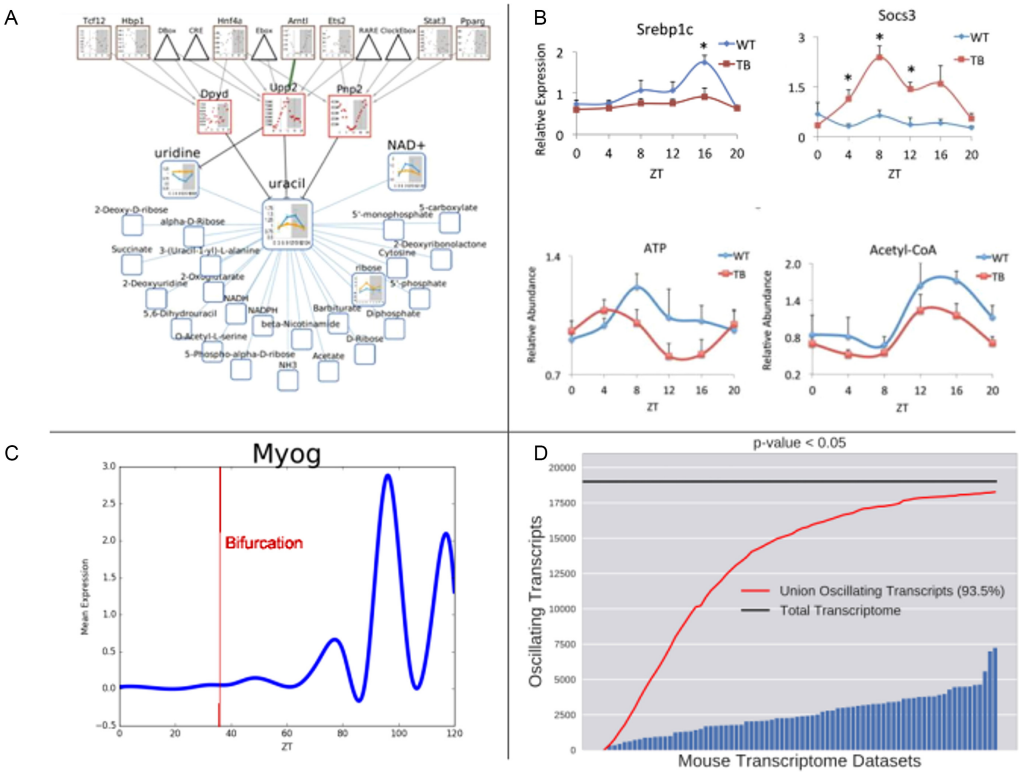


Figure 4. Selected examples of the impact of CircadiOmics. (**A**) CircadiOmics was used to link a multitude of circadian metabolites with functionally related circadian transcripts. Figure taken from Figure 5A of (17). (**B**) CircadiOmics was used to discover reprogrammed circadian transcripts and 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.

Chapter 7

Transcriptomic Organization of Circadian Rhythms

7.1 Introduction

The set of software systems described in this thesis provide us with a powerful set of tools for analyzing high throughput circadian omic data. Indeed, we have utilized this system in high impact research projects, some of which were listed above. CircadiOmics shares our results with the public and also gathers the largest set of circadian omic data available from the public domain. In particular it contains a large set of transcriptomic datasets, totaling more than 170 datasets across multiple species. This combination of tools and data uniquely positions us to address some of the most complex and systematic questions in circadian biology.

One of the most important questions of this kind is to understand the reprogramming and global organization of the circadian transcriptome, as mentioned in Chapter 1. To this end, we have engineered the `PyCircadiOmics` codebase. It incorporates results and tools from

individual pipelines in `igb-pipelines` and organizes them in a highly efficient and flexible python library. Using object oriented design, different methods can be readily combined to produce new results. Multiple parameters and input data can be processed in parallel to increase the robustness of results. By applying these methods on 81 mouse transcriptomic datasets, we have validated our methodology by rediscovering the core clock, and discovered dozens of novel circadian regulators. Last but not least, we have formulated a possible hierarchical model to explain large scale circadian reprogramming events in the transcriptome. Through our bioinformatics software systems, exemplified by PyCircadiOmics, we have conducted a systematic, data-driven and ground-breaking analysis of the circadian transcriptome in mouse.

The details of this study is shown in full in the appended paper. I am a first co-author of this paper (publication pending).

Transcriptomic Organization of Circadian Rhythms

Nicholas Ceglia[1,2,*], Yu Liu[1,2,*], Siwei Chen[1,2], Muntaha Samad[1,2]
Paolo Sassone-Corsi[2,3,4], & Pierre Baldi[1,2,3,4]

July 2, 2018

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. Center for Epigenetics and Metabolism, School of Medicine, University of California, Irvine, CA 92617, USA.
4. Department of Biological Chemistry, University of California, Irvine, CA 92617, USA.

Abstract

While many studies have highlighted the importance of the core circadian clock, a general model for the organization of the circadian transcriptome has yet to be described. To identify such a model, we developed the Circadian Regulatory Control (CRC) method and generated CRC graphs on 87 mouse transcriptomic datasets. Frequency analysis rediscovered the core circadian clock in both mouse and baboon. Node analysis on CRC graphs revealed more than 20 novel circadian transcription factors and RNA binding proteins. Edge analysis demonstrated that the core clock directly controls only a limited set of 334 TFs and RBPs. Graph analysis on the individual experiment specific CFC graphs as well as the aggregated CRC graph showed that the core clock directly controls about 35% of the oscillatory transcripts. However, at regulatory distance-3, the core clock can indirectly control more than 80% of the circadian transcripts. Furthermore, 95% of oscillatory transcripts can be reached from the core clock at some regulatory distance. Functional analysis of the remaining 5% non-oscillatory transcripts showed a significant enrichment of olfactory, sensory and G-protein coupled receptor pathways. Overall, we conclude that the circadian transcriptome is organized by a hierarchical regulatory system, where perturbation close to the core clock can cascade into large scale circadian reprogramming events.

1 Introduction

Circadian oscillations in the concentrations of molecular species play a fundamental role in many biological processes from metabolism, to cell cycle, and to neuronal function [1, 2, 3, 4]. To study the role of these

*These authors contributed equally to this work.

oscillations, an increasing amount of high through-put circadian omic data is being generated under diverse genetic, epigenetic, and environmental conditions. In any single circadian transcriptomic experiment, roughly 10% of measured transcripts are found to oscillate in a circadian manner [5, 6, 7, 8, 9]. However, the intersection of oscillating transcripts between any two experiments is typically small, only about 2% [10]. This small overlap between experiments suggests that the union of all oscillating transcripts across all experiments is large. Remarkably, we calculate that over 93% of all of protein coding transcripts in mouse are found to oscillate in at least one condition [11]. Previous studies have demonstrated specific mechanisms by which a cell can select different oscillating subsets of transcripts, an event known as circadian reprogramming [12, 13, 14, 9]. However, the question of how almost every transcript is capable of oscillating in a circadian manner remains unanswered. To address this question, we develop informatic methods and apply them to the largest collection of mouse circadian transcriptomic data.

To establish a framework, it must be noted first that the concentration of any molecular species cannot oscillate in isolation [15]. The fundamental unit of any such oscillation is a feedback loop of molecular interactions, such as transcriptional regulation, post-trascriptional modification, and protein-protein interactions [7, 16, 10], causing all species in the loop to oscillate at the same frequency. A very large number of such regulatory loops have been identified using informatics methods and large omic repositories [10, 17, 18, 19, 20, 21]. The empirically observed pervasiveness of circadian oscillations implies that a significant fraction of these loops is capable of oscillating with a 24 hour period. This 24 hour common period is most likely due to evolution given the importances of the differences between night and day for all biological life, the ~ 2 trillion night-day transitions [1] that have occurred since the origin of life 3.5 billion years ago, and the inherently circadian nature of the molecular circuitry of early photosynthetic life (cyanobacteria) [22]. Thus, in short, modern cells contain entire networks of circadian coupled oscillators. The question again is how specific subsets of oscillators are selected under specific genetic, epigenetic, and environmental conditions.

A key element of the answer to this question is the circadian core clock. The circadian core clock is genetically implemented by a relatively small set of genes whose transcripts are consistently found to oscillate in most circadian experiments [23, 24, 25]. The core clock regulates an extensive number of transcripts through a set of transcription factors (TF) including CLOCK-BMAL [26]. CLOCK-BMAL binds to E-box motifs that are found abundantly throughout the genome [27, 28]. A possible centralized model of organization is that the core clock directly orchestrates the selection of oscillators in the coupled network. While the importance of the core clock is undeniable [29, 26, 30, 31], additional findings have shown that knocking out elements of the core clock (including CLOCK-BMAL) does not lead to a complete loss of circadian oscillations [32, 33, 34, 35, 36]. Thus, at the other extreme, a completely decentralized model of circadian oscillations is also conceivable where oscillators compete and self organize. Here we seek to find where in this spectrum, from centrally orchestrated to completely decentralized, the cellular network of coupled-oscillators operates.

[1] This estimate accounts for the fact that the day and night cycle has been lengthening over time due to tidal influences.

To this end, using an intelligent discovery framework, we have aggregated the largest repository of high through-put circadian omic data on CircadiOmics (www.circadiomics.ics.uci.edu). Among others, CircadiOmics contains 161 transcriptomic datasets from 8 species and over 23 broad tissue categories. BIO_CYCLE, a deep learning based software available on CircadiOmics, is used to identify oscillating transcripts with statistical significance [37]. MotifMap and MotifMap-RNA are used to study transcription factors (TF) and RNA binding proteins (RBP) and their binding sites [18, 17, 38]. These binding sites can provide evidence for transcriptional and post-transcriptional regulation.

2 Results

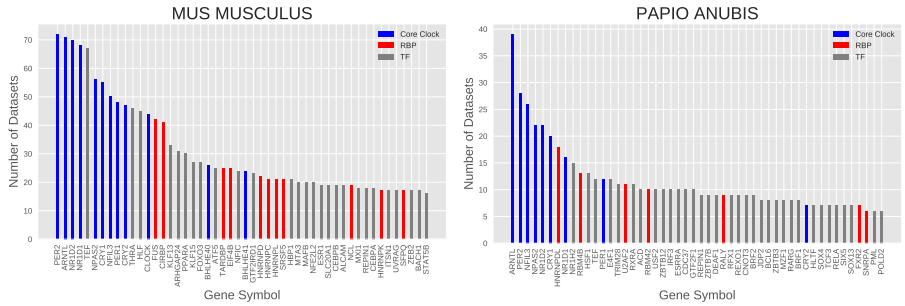
To identify a model of organization for transcriptomic circadian oscillations, we perform a series of analyses of increasing complexity using novel computational metrics. To achieve robustness and overcome noise in the data and incomplete knowledge, we present results obtained consistently at different statistical threshold as well as results that are supported by multiple lines of evidence. In total, 87 datasets from mouse were used to generate each set of results (Supplementary Table 1). The most represented tissues are liver (37 datasets), skin (14 datasets) and brain (13 datasets). In all cases, aggregated results, as well as tissue specific results, were generated.

2.1 Frequency Analysis

The frequency at which a TF or RBP is found to oscillate in a collection of datasets provides a simple metric for estimating its consistency in circadian oscillation. Figure 1 illustrates this frequency distribution for mouse at a BIO_CYCLE p-value < 0.01. Additionally, 64 datasets from Papio anubis (baboon) were used for comparison to validate the methods. Both analyses show that TFs involved in the circadian core clock are found to be the most frequently oscillating. This purely data-driven approach automatically discovers the circadian core clock. Furthermore, it identifies additional TFs and RBPs that must play an important role in circadian oscillation.

2.2 The Circadian Regulatory Control Graphs

Measuring the circadian regulatory influence of the TFs and RBPs identified in the previous analysis requires further investigation using more sophisticated computational methods. To this end, a novel computational method was used to identify and score directed regulatory edges in oscillating loops. The Circadian Regulatory Control (CRC) method can be understood as a proxy for circadian regulation between a TF or RBP source and a transcript target. There are three major components of the CRC method applied to each experiment. First, as a prerequisite, the source and target must be oscillating, as assessed by BIO_CYCLE. Second, the source must have at least one high quality binding site on the target for transcriptional or post-transcriptional regulation, as assessed by MotifMap and MotifMap-RNA [18, 17, 38]. For a TF, binding



(a) 87 Transcriptome Datasets

(b) 64 Transcriptome Datasets

Figure 1: Most Frequent Oscillating TFs and RBPs

sites were assessed at the promoter region of the target transcript. For an RBP, binding sites were assessed at the introns or UTRs of the target transcript. Third, there must be a correlative relationship between the phases of the source and the target. Recent studies have shown a significant lag between the transcript expression and the concentration of the corresponding protein [39]. We addressed this issue by computing and modeling the distribution of this lag, using transcriptomic and proteomic datasets produced from the same study on CircadiOmics (Methods 4.2).

After filtering on p-value for the first criteria, the remaining two criteria were combined into two different CRC scores. The *B-score* is a binary indicator of circadian regulation at various filtering thresholds for the number of high quality binding sites and the likelihood of phase correlation. The *E-score* is an exponentially weighted combination of these two criteria. In general, results generated using both scores tend to agree. However, as a binary indicator, the *B-score* is more convenient for large scale analysis of graph structures. In contrast, as a real valued metric, the *E-score* has more sensitivity and is used for ranking nodes and edges.

In either case, the CRC method yields a CRC graph, which is a directed weighted graph, where the vertices correspond to circadian protein-coding transcripts in a given experiment, and the directed edges correspond to regulatory relationships with $E\text{-}score > 0$, and the weights correspond to the *E-score* or the *B-score*. The CRC graphs obtained for each experiment can be superimposed, for instance, across tissues or across all experiments, by aggregating the scores of the corresponding edges.

2.3 Node Influence Analysis

For each source associated with a TF or RBP, an influence score was computed by aggregating all the CRC E-scores from all its outgoing edges either in all experiments or in tissue-specific experiments. The highest scoring TFs and RBPs are shown in Table 2 (Supplementary Table 2). When looking at aggregated results, core clock TFs such as CLOCK and BMAL1 were found to have the largest scores, a finding consistent with

Mouse All (n=81)		
TF/RBP	All Score	All Ranking
FUS	8.83	9
THRA	8.44	10
BHLHE40	8.30	11
NFIC	8.09	12
HNRPDL	7.56	15
CIRBP	7.42	16
MXI1	7.35	17
EIF4B	7.34	18
CEBPB	6.65	20
HNRNPK	5.61	21
TARDBP	5.40	22
KLF13	4.75	23
PPARA	4.73	24
FOXO3	4.64	25
RAD21	4.56	26
KHDRBS1	4.38	27
ALCAM	4.30	28
MTA3	4.25	29
ARHGAP24	4.23	30
YY1	4.22	31
HNRNPL	4.15	32
NFYA	4.14	33
MAFK	4.12	34
KLF15	4.12	35
PCBP4	4.06	36
HNRNPC	4.06	37
ESR1	4.00	38
SREBF2	3.92	39
BACH1	3.77	40

Mouse Brain (n=13)		
TF/RBP	Brain Score	Brain Rank
CIRBP	1.54	2
SFPQ	1.20	6
KLF15	0.96	8
FUS	0.94	9
ZC3H11A	0.88	10
MXI1	0.86	11
RBM28	0.81	14
EGR1	0.75	15
CHD1	0.71	16
CREB1	0.71	17
HIF1A	0.67	18
HNRPDL	0.66	19
CEBPG	0.65	20
BHLHE40	0.64	21
HNRNPK	0.63	22
SP2	0.63	23
RAD21	0.62	24
NFE2L2	0.60	25
EGR2	0.59	26
GTF2I	0.58	28
KLF12	0.58	29
HCFC1	0.58	30
CHD2	0.57	31
GTF2F1	0.56	32
SRPR	0.54	33
CEBPD	0.53	34
ETV1	0.49	35
GABPA	0.49	36
A1CF	0.49	37

Mouse Skin (n=14)		
TF/RBP	Skin Score	Skin Rank
NFIC	3.44	6
E2F1	2.68	7
MXI1	2.34	10
RUNX1	1.59	13
BRCA1	1.55	14
TCF4	1.46	15
HCFC1	1.41	16
MEF2A	1.32	17
ETV5	1.27	18
THRA	1.17	19
CHD1	1.16	20
FOXM1	1.12	21
NFATC1	1.12	22
FUS	1.12	23
ALCAM	1.11	24
HNRPDL	1.08	25
NFYA	1.03	26
ZFP161	1.00	27
CHD2	0.98	28
RBM5	0.97	29
TCF12	0.95	30
SREBF2	0.94	31
KHDRBS1	0.91	32
EIF4B	0.87	33
KLF13	0.85	34
ELK4	0.85	35
RFX5	0.83	36
SF1	0.81	37
ELAVL1	0.79	38

Mouse Liver (n=31)		
TF/RBP	Liver Score	Liver Rank
CEBPB	4.25	4
BHLHE40	4.22	5
FUS	3.92	7
HNRNPK	3.58	10
EIF4B	3.51	12
PCBP4	3.45	14
THRA	3.30	15
MXI1	2.83	18
YY1	2.80	19
MAFK	2.79	20
ATF5	2.75	21
MTA3	2.75	22
PPARA	2.63	23
BACH1	2.59	24
RXRA	2.58	25
ESR1	2.57	26
RFX4	2.54	27
HNRPDL	2.53	28
FOXO3	2.52	29
CIRBP	2.52	30
STAT5B	2.49	31
CRP	2.43	32
TARDBP	2.37	33
HNRNPC	2.35	34
ARHGAP24	2.33	35
DCTN2	2.28	36
NFIC	2.27	37
KLF1	2.25	38
USF2	2.25	39

Figure 2: Tables showing the ranking of circadian TFs and RBPs by CRC E-score in different tissue types. The leftmost table shows ranking in mouse transcriptome across all datasets. RBPs are labeled in red while TFs are labeled in black. Core clock TFs have been removed from the listing.

both the frequency of oscillation and previous literature [29]. Extended members of the core clock were also identified in the ranking including THRA and BHLHE40 [40, 41].

In the results across all datasets, additional TFs and RBPs were identified that seem to have a much broader regulatory role than reported in the literature. For instance, FUS and CIRBP have been reported to affect the core circadian factor PER2 via alternative splicing, but only in the mouse liver [42, 43, 44]. In contrast, we find that FUS and CIRBP are found to be high scoring also in both brain and skin. EIF4B has been identified in the circadian regulation of translation in mouse liver [45]. We find that EIF4B is also top scoring in skin. HNRPDL is listed as a potential target of circadian regulation via microRNA in the brain [46]. Strikingly, these RBPs and TFs are found to have very high CRC scores across all mouse datasets. This suggest that they play a broader, previously uncharacterized, role in circadian regulation.

When looking at tissue specific results, many additional TFs and RBPs with high CRC scores are discovered. Although literature evidence has shown that these factors interact with circadian pathways, they are not known to be regulators of oscillation. These TFs may explain tissue specific circadian reprogramming. Within brain tissue, SFPQ is functionally involved in the cell cycle pathway, which also includes NONO and PER2 [47]. EGR1 has been found to oscillate and under direct regulation by the core clock [48]. Within our results, EGR1 potentially regulates a large number of downstream transcripts in the brain. CHD1 is known to be involved in circadian chromatin remodeling in brain [49]. KLF15 is well known to be regulated by the peripheral clock in relation to circadian nitrogen homeostatis in liver and muscle [50]. Within skin

tissue, RUNX is a top TF and is known to be regulated in a circadian fashion in epidermal cells [51]. E2F1 is regulated by circadian factors SIRT1 and CLOCK[52]. BRCA1 is known to interact with core clock TFs such as PER2 [53]. Within liver tissue, CEBPB is top ranking excluding core clock TFs. This agrees with the literature finding that it interacts with the core clock through REV-ERB [54]. PCBP4 is known to be involved in circadian alternative splicing in the liver [55].

Additionally, there are many other novel findings that have been linked to very few circadian studies. These findings include: NFIC, RAD21, MXI1, and TARDBP across all tissues; ZC3H11A, RBM28, and CEBPG in brain; HCFC1 and ETV5 in skin; and HNRNPK, ATF5, and BACH1/MAFK in liver (Supplementary Table 2), and may provide leads for investigations of previously unknown circadian regulatory mechanisms.

While these results have focused on individual TF and RBP nodes, the identification of a model of transcriptomic organization requires analysis of regulatory edges found using the CRC method.

2.4 Edge Influence Analysis

While the previous analysis was performed on the nodes of the CRC graph (aggregated or tissue-specific), to further understand the organization of circadian transcripts, it is necessary to study the regulatory edges (aggregated or tissue-specific). Because we are focused on the regulatory architecture, we restrict the results to edges where the source is either a TF or RBP and the target is also either a TF or RBP. This is a subgraph of the whole CRC graph. An edge analysis was performed using CRC E-score weighted edges in this restricted graph. We rank these edges on the aggregated *E-scores* (Figure 3 A). The results compiled from all 87 mouse datasets are shown in Figure 3. Additional tissue specific results can be found in Supplementary Figure 3. The heatmap demonstrates strong correlation within core clock TFs, as well as between the core clock and other top CRC scoring TFs and RBPs such as FUS and CIRBP. Figure 3 A lists some of the top interactions. Many of the findings are consistent with literature (Supplementary Table 2). However, some top interactions seem to be underestimated in the circadian literature, such as the potential regulation of PER2 by NFIX, or TEF by NFYA and RXRA. These novel interactions may be important in circadian regulation. Furthermore, we visualized the weighted edge scores between top TFs and RBPs, including the core clock, in Figure 2 B. It can be clearly seen that while there exists strong interaction between core clock TFs, the core clock does not strongly regulate a large number of additional TFs and RBPs. In fact, we find that using the CRC graph, the core clock only directly regulates 334 TFs and RBPs (Supplementary Material 3). Given the limited direct regulatory influence of the core clock, it is necessary to examine indirect regulatory influence beyond the immediate neighborhood of the core clock within the CRC graphs.

2.5 Graph Analysis

So far we have looked at direct influence through ranked node and edge scores, to further understand the architecture of circadian transcriptional oscillations, we must look also at indirect influences through

A

Regulation	Mouse Score	Ranking	Type
FUS->PER2	29.49	19	RBP->Core Clock TF
ARNTL->NFIC	24.13	28	Core Clock TF->TF
BHLHE40->PER2	23.64	29	TF->Core Clock TF
ARNTL->FOXO3	22.41	31	Core Clock TF->TF
NFIX->PER2	21.03	36	TF->Core Clock TF
MXI1->PER2	20.54	37	TF->Core Clock TF
CEBPB->PER2	17.59	48	TF->Core Clock TF
NFYA->TEF	17.42	50	TF->Core Clock TF
ARNTL->KLF13	16.61	52	Core Clock TF->TF
ARNTL->BHLHE41	15.91	53	Core Clock TF->TF
KLF15->ARNTL	15.59	55	TF->Core Clock TF
RXRA->TEF	15.37	56	TF->Core Clock TF
ARNTL->KLF15	14.84	60	Core Clock TF->TF
PER2->KLF15	14.76	61	Core Clock TF->TF
ARNTL->NCL	14.67	62	Core Clock TF->RBP
NFYA->ARNTL	14.63	63	TF->Core Clock TF
PER1->FOXO3	14.52	64	Core Clock TF->TF
ARNTL->BHLHE40	14.52	65	Core Clock TF->TF
REPIN1->TEF	14.09	66	TF->Core Clock TF
ARNTL->SLC20A1	13.94	68	Core Clock TF->TF
USF2->TEF	11.65	72	TF->Core Clock TF
ARNTL->XBP1	11.29	73	Core Clock TF->TF
CEBPB->NFIL3	10.63	74	TF->Core Clock TF
PER1->KLF15	9.40	75	Core Clock TF->TF
ARNTL->UVRAG	9.39	76	Core Clock TF->TF
CRY1->REPIN1	9.34	77	Core Clock TF->TF
NFYA->PER2	9.30	78	TF->Core Clock TF
ARNTL->MAFK	8.58	79	Core Clock TF->TF
BHLHE41->CRY1	8.55	80	TF->Core Clock TF

B

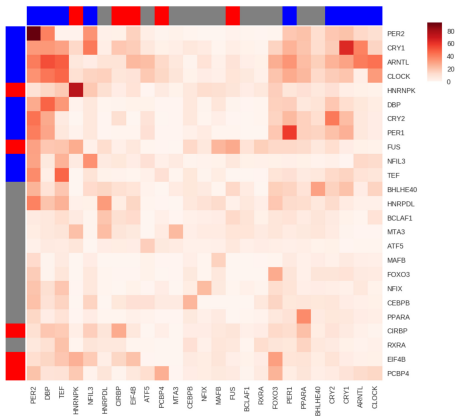


Figure 3: Edge influence analysis. **A**: Ranking of top edges between TFs and RBPs. Edges between core clock TFs have been omitted. **B: Edge score heatmap** Heatmap of inter-regulator (TF/RBP) circadian CRC score (*E-score* aggregates) in mouse. The score is calculated by aggregating CRC *E-scores* from the directed edges starting from row TF/RBP to the column TF/RBP across all datasets. Stronger colors in the heatmap indicate higher total scores (normalized for visualization). Color on row and column indicates the type of regulators: blue indicates core clock TF, red indicates RBP and gray indicates regular TF. The heatmap visualizes the 100 strongest edges by aggregate scores.

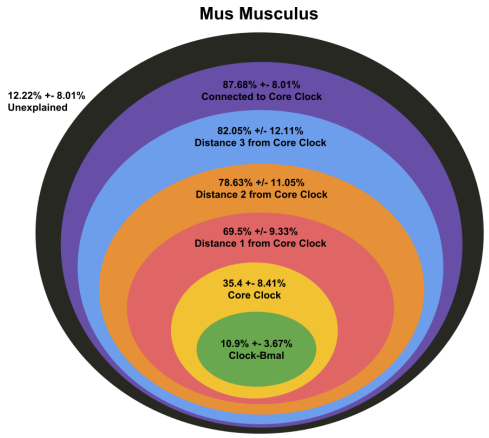


Figure 4: Mean percentages of transcriptome explained by TF/RBP at fixed regulatory distances from the core clock across mouse datasets.

regulatory paths within the CRC graphs.

Regulatory distance was computed as the length of the shortest directed path in a CRC graph from a source node to a sink node. The mean percentage of oscillating transcripts found at increasing regulatory distances across all mouse datasets is shown in Figure 4. The set of oscillating transcripts that are found to have a regulatory distance-one from the core clock were considered to be directly regulated by the core clock. The mean percentage of distance-one transcripts across all CRC graphs (individual, tissue-specific, or aggregated) is roughly 35% of all oscillating transcripts. While the majority of transcripts are not found to be directly regulated by the core clock, almost any transcript can be connected through a regulatory path in the CRC graph to the core clock. On average, in an individual experiment, 80% of all oscillating transcripts are within distance-three from the core clock (Figure 4).

Given that over 80% of oscillating transcripts can be connected to the core clock within a distance of three within experiment specific CRC graphs, we performed analysis of the aggregated CRC graph to quantify the total number of oscillating transcripts over the entire mouse transcriptome that can be connected to the core clock. We find that not only do 95% of protein-coding transcripts oscillate in some condition, but that this 95% can be connected to the core clock at some regulatory distance. The remaining 5% of non-oscillatory transcripts are found to be significantly enriched (p-value < 0.001, q-value < 0.01) in a few highly specific pathways including olfactory receptor activities, sensory perception and G-protein coupled receptor activity (Supplementary Material 4).

Within the aggregated CRC graph, the number of out-going edges for TFs and RBPs at increasing regulatory distances decreases. Core clock TFs have an average of 21 out-going edges. At distance-one,

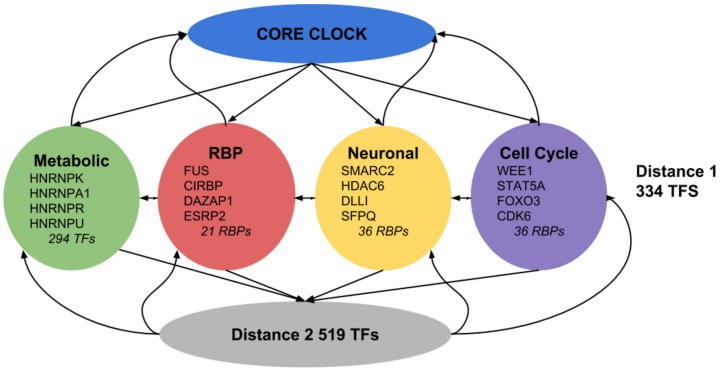


Figure 5: Network view of TFs and RBPs that are found at regulatory distance = 1. These TFs predominantly fall into three broad categories labeled from GO annotations that includes *Cell Cycle*, *Neuronal Function*, and *Metabolic Process*.

this number decreases to nine. For TFs and RBPs at a distance of three or greater, the average number of outgoing edges is less than one. The decreasing number of out-going edges at each regulatory distances illustrates the decreasing regulatory influence of these TFs and RBPs and highlights a cascading hierarchical organization of transcriptomic regulation.

Functional enrichment by Gene Ontology (GO) term was performed on the subset of transcripts found at each regulatory distance (Supplementary Table 3). Transcripts found at regulatory distance-1 from the core clock exhibited significant enrichment for terms related to Circadian Rhythms, Cell Cycle, Metabolic Processes, and Neuronal Processes. The sets of transcription factors found at this distance were grouped into these functional categories including the set of known RBPs. To observe the significant shifts in the set of oscillating transcripts found in reprogramming events, a perturbation must occur in the expression of TFs and RBPs with small regulatory distances and, consequently, a high CRC score. Perturbations affecting low CRC scoring TFs and RBPs at large regulatory distances can only induce changes in proportionally smaller sets of oscillating transcripts. Therefore, we conclude that experimental conditions that relate to changes in metabolism, neuronal function, cell cycle, or RBP modification likely instigate large reprogramming events through perturbations found in distance-one TFs or RBPs.

3 Discussion

The question of how specific subsets of oscillators are selected under specific genetic, epigenetic, and environmental conditions has remained an open question in the study of circadian rhythms. An organizational model that allows this selection within the network of coupled circadian oscillators must exist. This model must lie within a spectrum extending from a purely centralized model, entirely orchestrated by the core circadian clock, to a completely decentralized model. In order to determine where in this spectrum the network of coupled oscillators operates, we obtained results using the CRC method and CRC graph with the large collection of mouse transcriptomic data available on CircadiOmics.

Using 87 mouse transcriptomic datasets, we analyzed the frequency with which TFs and RBPs oscillate. With no prior information, this purely data driven results rediscovered the core circadian clock. Remarkably, these results were confirmed using the 64 Baboon datasets in CircadiOmics.

We formulated the Circadian Regulatory Control (CRC) method for identification of regulatory edges in circadian feedback loops. Two CRC scores, *B-Score* and *E-score*, incorporate multiple sources of evidence including the statistical significance of transcript oscillation and high quality predicted binding sites. These scores also take into account the delay between transcript and protein expression using available proteomic datasets included in CircadiOmics. CRC scores provide a measure circadian regulation from a TF or RBP to a target (promoter or transcript).

Using the CRC method, we introduced the concept of the CRC graph as a directed weighted graph where nodes are oscillating transcripts and regulatory edges are measured by CRC scores. We constructed CRC graphs from individual experiments and superimposed these graphs to generate tissue-specific and aggregated CRC graphs.

We performed a node influence analysis on the CRC graphs by aggregating all of the outgoing edges of a source node across experiments. The node influence analysis revealed more than 20 novel TFs and RBPs that are either previously underestimated or completely uncharacterized as important circadian regulators. Our analysis of directed edges within the CRC graph demonstrated the relatively small direct influence of the core clock. We found that core clock TFs only directly regulate 334 TFs and RBPs in the aggregated CRC graph.

We performed further graph analysis on both experiment specific CRC graphs and the aggregated CRC graph. We presented the average number of oscillating protein coding transcripts at increasing regulatory distances centering around the core clock. We found that 35% ($\pm$8%) of oscillating protein coding transcripts are directly regulated by the core clock. Additionally, we showed that on average 80% of oscillating transcripts can be connected within a regulatory distance of three to the core clock.

Analysis of the aggregated CRC graph revealed that not only do 95% of protein coding transcripts oscillate in some condition, but that any oscillating transcript can be connected at some regulatory distance to the core clock. Functional analysis of the remaining 5% non-oscillatory transcripts showed significant enrichment in highly specific pathways such as olfactory receptor activities, sensory perception and G-protein coupled

receptor activity. At increasing regulatory distances, we observed that TFs and RBPs have decreasing aggregated CRC edge scores and a decreasing number of out-going edges. This observation highlights the decreasing regulatory influence at increasing regulatory distances. Finally, functional annotation by GO terms revealed that at a regulatory distance of one, TFs can be grouped together into three categories: Cell Cycle, Metabolic Process, and Neuronal Function. These sets, along with the set of RBPs found at distance-one, are functionally related and highly influential circadian regulators. Reprogramming events, where the core circadian clock continues to oscillate, must originate from perturbations of expression in these TFs and RBPs. We conclude from these analyses that the organization of circadian transcriptional regulation is structured as a heirarchy of regulatory steps originating at the core circadian clock.

4 Online Methods

The methods presented provide tools for identifying the organization of circadian rhythms within the context of the transcriptome. This type of analysis is made possible by the large repository of circadian omic data in Circadiomics. CircadiOmics currently provides access to 231 omic datasets from eight species in a broad collection of tissue and conditions. To address bias in the repository of data related to tissue and species, separate results were computed from only mouse liver, skin, and brain are provided for each method (Supplementary Table 3).

4.1 Frequency Histograms

Transcript frequency is defined as the total number of datasets where a given protein coding transcript is found to be oscillating at a BIO_CYCLE p-value < 0.01. Histograms for results obtained at additional p-values is provided in Supplementary Figure 1. Protein-coding transcripts were identified from the BioMart ENSEMBL gene database [56].

4.2 Circadian Regulation Control Scoring

The Circadian Regulation Control method is a measure of circadian regulation from a source TF/RBP to a target (promoter or transcript), consisting of three components. First, both the source and the target must be oscillatory according to BIO_CYCLE [37] with p$\leq$ 0.01 (partial results with p$\leq$ 0.05 available in Supplementary Table 2). Second, if the source is a TF, there must exist at least one high quality binding site ($BBLS \geq 1, NLOD \geq 0.9$ and $FDR \leq 0.25$) in the promoter region of the target (defined as -5,000 to +500 bp of the transcription start site) according to MotifMap [18, 17]. For RBP sources, there must be high quality binding sites ($BBLS > 0, NLOD \geq 0.9$) predicted by MotifMap-RNA [38] in the introns or UTRs of its target transcripts.

Third, a measure of the correlative relationship between the phases of the source and the target is computed from the lag between peak expression. The distribution of lags between the transcript and the

protein product of 2,400 genes [57] was modeled as a beta distribution with $\mu = 0.35, SD = 0.25$. The empirical distribution and the fitted distribution are shown in Supplementary Figure 2. The lag between a circadian transcript and the protein product is roughly 4 $\pm$2 hours corresponding to the mode of the fitted distribution. CRC scores are computed from the value of the highest probability density interval (HPD) associated with a given lag between a source and its target.

The values of these two components are summarized in two ways. A binary edge indicator, the *B-score*, uses a threshold number of binding sites (≥ 1) and phase correlation as measured by HPD value (≥ 0.8). An exponentially weighted edge score, *E-score*, is defined as $E = 0.5 * HPD - 0.5 * exp(-NumberOfBindingSites)$.

4.3 Circadian Regulatory Control Graphs

CRC Graphs are constructed for each of the 87 mouse datasets. Nodes are defined as oscillating transcripts using a threshold BIO_CYCLE p-value ≤ 0.01. Edges with a CRC *B-score* = 1 are included. The shortest path length, used to define regulatory distance, between a TF and RBP and a target transcript and the number of out-going edges for each TF and RBP was computed by NetworkX (https://networkx.github.io/). Results for mouse liver, skin, and brain tissue are provided in Supplementary Table 4. GO enrichment analysis and functional annotation for transcripts found at each distance was completed using the Python library goatools (https://github.com/tanghaibao/goatools) using NCBI gene to GO term associations. The enrichment results provided in Supplementary Table 1 were performed at an adjusted p-value ≤ 0.01.

References

[1] Bass, J. Circadian topology of metabolism. *Nature* **491**, 348–356 (2012). [PubMed:23151577] [doi:10.1038/nature11704].

[2] Dibner, C., Schibler, U. & Albrecht, U. *The Mammalian Circadian Timing System: Organization and Coordination of Central and Peripheral Clocks*, vol. 72 (2010). URL http://www.annualreviews.org/doi/10.1146/annurev-physiol-021909-135821. [PubMed:20148687] [doi:10.1146/annurev-physiol-021909-135821].

[3] Gerstner, J. R. *et al.* Cycling Behavior and Memory Formation. *Journal of Neuroscience* **29**, 12824–12830 (2009). URL http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.3353-09.2009. [PubMed:19828795] [PubMed Central:PMC4077269] [doi:10.1523/JNEUROSCI.3353-09.2009].

[4] Menet, J. S. & Rosbash, M. When brain clocks lose track of time: cause or consequence of neuropsychiatric disorders. *Current Opinion in Neurobiology* **21**, 849–857 (2011). URL http://link.springer.com/10.1007/s10955-011-0269-9http://linkinghub.elsevier.com/retrieve/pii/S0959438811001073. [PubMed:21737252] [PubMed Central:PMC3252427] [doi:10.1016/j.conb.2011.06.008], NIHMS150003.

[5] Eckel-Mahan, K. L. *et al.* Coordination of the transcriptome and metabolome by the circadian clock. *Proceedings of the National Academy of Sciences* **109**, 5541–5546 (2012). URL http://www.pnas.org/cgi/doi/10.1073/pnas.1118726109. [PubMed:22431615] [PubMed Central:PMC3325727] [doi:10.1073/pnas.1118726109].

[6] Eckel-Mahan, K. L. *et al.* Reprogramming of the circadian clock by nutritional challenge. *Cell* **155**, 1464–1478 (2013). [PubMed:24360271] [PubMed Central:24360271] [doi:10.1016/j.cell.2013.11.034].

[7] Masri, S. *et al.* Circadian acetylome reveals regulation of mitochondrial metabolic pathways. *Proceedings of the National Academy of Sciences* **110**, 3339–3344 (2013). URL http://www.pnas.org/lookup/doi/10.1073/pnas.1217632110. [PubMed:23341599] [PubMed Central:PMC4908327] [doi:10.1073/pnas.1217632110].

[8] Masri, S. *et al.* Lung Adenocarcinoma Distally Rewires Hepatic Circadian Homeostasis. *Cell* **165**, 896–909 (2016). URL http://dx.doi.org/10.1016/j.cell.2016.04.039. [PubMed:27153497] [PubMed Central:PMC5373476] [doi:10.1016/j.cell.2016.04.039].

[9] Murakami, M. *et al.* Gut microbiota directs PPARγdriven reprogramming of the liver circadian clock by nutritional challenge. *EMBO reports* **17**, 1292–1303 (2016). URL http://embor.embopress.org/lookup/doi/10.15252/embr.201642463. [PubMed:27418314] [PubMed Central:PMC5007574] [doi:10.15252/embr.201642463].

[10] Patel, V. R. *et al.* The pervasiveness and plasticity of circadian oscillations: The coupled circadian-oscillators framework. *Bioinformatics* **31**, 3181–3188 (2015). [PubMed:26049162] [PubMed Central:PMC4592335] [doi:10.1093/bioinformatics/btv353].

[11] Nicholas Ceglia, S. C. F. A. K. E.-M. P. S.-C. P. F. B., Yu Liu. CircadiOmics: Circadian Omic Web Portal. *Nucleic Acid Research* (2018).

[12] Li, X.-M. *et al.* Cancer inhibition through circadian reprogramming of tumor transcriptome with meal timing. *Cancer research* **70**, 3351–3360 (2010).

[13] Li, X. M. *et al.* Cancer inhibition through circadian reprogramming of tumor transcriptome with meal timing. *Cancer Research* **70**, 3351–3360 (2010). [PubMed:20395208] [doi:10.1158/0008-5472.CAN-09-4235].

[14] Yagita, K. *et al.* Development of the circadian oscillator during differentiation of mouse embryonic stem cells in vitro. *Proceedings of the National Academy of Sciences* **107**, 3846–3851 (2010). URL http://www.pnas.org/cgi/doi/10.1073/pnas.0913256107. [PubMed:20133594] [PubMed Central:PMC2840478][doi:10.1073/pnas.0913256107], arXiv:1408.1149.

[15] Bell-Pedersen, D. *et al.* Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nature Reviews Genetics* **6**, 544 (2005).

[16] Robles, M. S., Cox, J. & Mann, M. In-Vivo Quantitative Proteomics Reveals a Key Contribution of Post-Transcriptional Mechanisms to the Circadian Regulation of Liver Metabolism. *PLoS Genetics* **10** (2014). [PubMed:24391516] [PubMed Central:PMC2840478] [doi:10.1371/journal.pgen.1004047].

[17] Daily, K., Patel, V. R., Rigor, P., Xie, X. & Baldi, P. MotifMap: integrative genome-wide maps of regulatory motif sites for model species (2011).

[18] Xie, X., Rigor, P. & Baldi, P. MotifMap: A human genome-wide map of candidate regulatory motif sites. *Bioinformatics* **25**, 167–174 (2009).

[19] Szklarczyk, D. *et al.* String v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic acids research* **43**, D447–D452 (2014).

[20] Kanehisa, M. & Goto, S. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research* **28**, 27–30 (2000).

[21] Rual, J.-F. *et al.* Towards a proteome-scale map of the human protein–protein interaction network. *Nature* **437**, 1173 (2005).

[22] Panda, S., Hogenesch, J. B. & Kay, S. A. Circadian rhythms from flies to human. *Nature* **417**, 329–335 (2002). [PubMed:12015613] [doi:10.1038/417329a].

[23] Ko, C. H. & Takahashi, J. S. Molecular components of the mammalian circadian clock. *Human Molecular Genetics* **15**, 271–277 (2006). [PubMed:16987893] [doi:10.1093/hmg/ddl207], 284.

[24] Koike, N. *et al.* Transcriptional Architecture and Chromatin Landscape of the Core Circadian Clock in Mammals. *Science* **338**, 349–354 (2012). URL http://www.sciencemag.org/cgi/doi/10.1126/science.1226339. [PubMed:22936566] [PubMed Central:PMC3694775] [doi:10.1126/science.1226339].

[25] Takahashi, J. S., Hong, H.-K., Ko, C. H. & McDearmon, E. L. The genetics of mammalian circadian order and disorder: implications for physiology and disease. *Nature Reviews Genetics* **9**, 764 (2008).

[26] Ripperger, J. A. & Schibler, U. Rhythmic clock-bmal1 binding to multiple e-box motifs drives circadian dbp transcription and chromatin transitions. *Nature genetics* **38**, 369 (2006).

[27] Oishi, K. *et al.* Genome-wide expression analysis of mouse liver reveals clock-regulated circadian output genes. *Journal of Biological Chemistry* **278**, 41519–41527 (2003).

[28] Zhang, E. E. *et al.* A genome-wide rnai screen for modifiers of the circadian clock in human cells. *Cell* **139**, 199–210 (2009).

[29] Ueda, H. R. *et al.* System-level identification of transcriptional circuits underlying mammalian circadian clocks. *Nature genetics* **37**, 187 (2005).

[30] Kondratov, R. V., Kondratova, A. A., Gorbacheva, V. Y., Vykhovanets, O. V. & Antoch, M. P. Early aging and age-related pathologies in mice deficient in bmal1, the core componentof the circadian clock. *Genes & development* **20**, 1868–1873 (2006).

[31] Rudic, R. D. *et al.* Bmal1 and clock, two essential components of the circadian clock, are involved in glucose homeostasis. *PLoS biology* **2**, e377 (2004).

[32] Kondratov, R. V., Kondratova, A. A., Gorbacheva, V. Y., Vykhovanets, O. V. & Antoch, M. P. Early aging and age-related pathologies in mice deficient in bmal1, the core componentof the circadian clock. *Genes & development* **20**, 1868–1873 (2006).

[33] Liu, A. C. *et al.* Intercellular coupling confers robustness against mutations in the scn circadian clock network. *Cell* **129**, 605–616 (2007).

[34] Alvarez, J. *et al.* The circadian clock protein bmal1 is necessary for fertility and proper testosterone production in mice. *Journal of biological rhythms* **23**, 26–36 (2008).

[35] Yamaguchi, S. *et al.* Role of dbp in the circadian oscillatory mechanism. *Molecular and cellular biology* **20**, 4773–4781 (2000).

[36] DeBruyne, J. P., Weaver, D. R. & Reppert, S. M. Clock and npas2 have overlapping roles in the suprachiasmatic circadian clock. *Nature neuroscience* **10**, 543 (2007).

[37] Agostinelli, F., Ceglia, N., Shahbaba, B., Sassone-Corsi, P. & Baldi, P. What time is it? Deep learning approaches for circadian rhythms. *Bioinformatics* **32**, i8–i17 (2016). [PubMed:27307647] [PubMed Central:PMC4908327] [doi:10.1093/bioinformatics/btw243].

[38] Liu, Y. *et al.* MotifMap-RNA: a genome-wide map of RBP binding sites. *Bioinformatics* **33**, 2029–2031 (2017). URL https://academic.oup.com/bioinformatics/article/3045023/MotifMap-RNA: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btx087.

[39] Robles, M. S., Cox, J. & Mann, M. In-vivo quantitative proteomics reveals a key contribution of post-transcriptional mechanisms to the circadian regulation of liver metabolism. *PLoS genetics* **10**, e1004047 (2014).

[40] Storch, K.-F. *et al.* Extensive and divergent circadian gene expression in liver and heart. *Nature* **417**, 78 (2002).

[41] Li, J. Z. *et al.* Circadian patterns of gene expression in the human brain and disruption in major depressive disorder. *Proceedings of the National Academy of Sciences* **110**, 9950–9955 (2013).

[42] Kornmann, B., Schaad, O., Bujard, H., Takahashi, J. S. & Schibler, U. System-driven and oscillator-dependent circadian transcription in mice with a conditionally active liver clock. *PLoS biology* **5**, e34 (2007).

[43] Morf, J. *et al.* Cold-inducible rna-binding protein modulates circadian gene expression posttranscriptionally. *Science* **338**, 379–383 (2012).

[44] Oster, H., Damerow, S., Hut, R. A. & Eichele, G. Transcriptional profiling in the adrenal gland reveals circadian regulation of hormone biosynthesis genes and nucleosome assembly genes. *Journal of biological rhythms* **21**, 350–361 (2006).

[45] Jouffe, C. *et al.* The circadian clock coordinates ribosome biogenesis. *PLoS biology* **11**, e1001455 (2013).

[46] Cheng, H.-Y. M. *et al.* microrna modulation of circadian-clock period and entrainment. *Neuron* **54**, 813–829 (2007).

[47] Guillaumond, F. *et al.* Chromatin remodeling as a mechanism for circadian prolactin transcription: rhythmic nono and sfpq recruitment to hltf. *The FASEB Journal* **25**, 2740–2756 (2011).

[48] Lee, J. H. & Sancar, A. Circadian clock disruption improves the efficacy of chemotherapy through p73-mediated apoptosis. *Proceedings of the National Academy of Sciences* **108**, 10668–10672 (2011).

[49] Belden, W. J., Lewis, Z. A., Selker, E. U., Loros, J. J. & Dunlap, J. C. Chd1 remodels chromatin and influences transient dna methylation at the clock gene frequency. *PLoS genetics* **7**, e1002166 (2011).

[50] Jeyaraj, D. *et al.* Klf15 orchestrates circadian nitrogen homeostasis. *Cell metabolism* **15**, 311–323 (2012).

[51] Janich, P. *et al.* Human epidermal stem cell function is regulated by circadian oscillations. *Cell stem cell* **13**, 745–753 (2013).

[52] Nakahata, Y. *et al.* The nad+-dependent deacetylase sirt1 modulates clock-mediated chromatin remodeling and circadian control. *Cell* **134**, 329–340 (2008).

[53] Winter, S. L., Bosnoyan-Collins, L., Pinnaduwage, D. & Andrulis, I. L. Expression of the circadian clock genes pert, per2 in sporadic, familial breast tumors. *Neoplasia* **9**, 797–800 (2007).

[54] Huang, G., Zhang, F., Ye, Q. & Wang, H. The circadian clock regulates autophagy directly through the nuclear hormone receptor nr1d1/rev-erbα and indirectly via cebpb/(c/ebpβ) in zebrafish. *Autophagy* **12**, 1292–1309 (2016).

[55] McGlincy, N. J. *et al.* Regulation of alternative splicing by the circadian clock and food related cues. *Genome biology* **13**, R54 (2012).

[56] Smedley, D. *et al.* The biomart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Research* **43**, W589–W598 (2015). URL http://dx.doi.org/10.1093/nar/gkv350. /oup/backfile/content_public/journal/nar/43/w1/10.1093_nar_gkv350/2/gkv350.pdf.

[57] Dallmann, R., Viola, A. U., Tarokh, L., Cajochen, C. & Brown, S. A. The human circadian metabolome. *Proceedings of the National Academy of Sciences* **109**, 2625–2629 (2012). URL http://www.pnas.org/cgi/doi/10.1073/pnas.1114410109. [PubMed:22308371] [PubMed Central:22308371] [doi:10.1073/pnas.1114410109].

Chapter 8

Summary and Conclusion

The study of circadian rhythms increasingly relies on high throughput circadian omic data. The processing, analyzing and integrating of such data requires a set of efficient and powerful bioinformatics methods. We have implemented a collection of such methods and organized them into complete software systems. Using these systems, we produced large scale genomics data for circadian analysis. We created pipelines for circadian statistical and machine learning analyses. We also created pipelines for large scale functional annotations and enrichment analyses. Furthermore, we converted large scale functional data to functional graphs. For individual projects, we implemented a visualization and web hosting server for circadian data. For all publicly available circadian data, we have created CircadiOmics, which the largest circadian omic web portal to date. Thanks to these efforts, we have published numerous circadian papers. Finally, using our software systems and data, we have studied the organization of the circadian transcriptome and produced significant results: we found that, in mouse, the core clock directly controls about 35% of the circadian transcripts, while it can indirectly affect more than 80% of the circadian transcripts in any given condition. Furthermore, more than 95% of the protein coding transcripts can be connected to the core clock by some regulatory distance. The remaining 5% of non-oscillatory transcripts are enriched for

specific pathways such as olfactory receptors. Along with these results, we have identified more than 20 novel circadian TFs and RBPs in terms of their role in circadian regulation.

Overall, we have created a ground breaking framework for studying circadian rhythms and obtained significant novel discoveries. This framework should continue to benefit the whole of the circadian field in the future.

Bibliography

[1] H. Adser, J. F. P. Wojtaszewski, A. H. Jakobsen, K. Kiilerich, J. Hidalgo, and H. Pilegaard. Interleukin-6 modifies mRNA expression in mouse skeletal muscle. *Acta Physiologica*, 202(2):165–173, 2011. [PubMed:21352507] [doi:10.1111/j.1748-1716.2011.02269.x].

[2] F. Agostinelli, N. Ceglia, B. Shahbaba, P. Sassone-Corsi, and P. Baldi. What time is it? Deep learning approaches for circadian rhythms. *Bioinformatics*, 32(12):i8–i17, 2016. [PubMed:27307647] [PubMed Central:PMC4908327] [doi:10.1093/bioinformatics/btw243].

[3] A. Alyass, M. Turcotte, and D. Meyre. From big data analysis to personalized medicine for all: Challenges and opportunities. *BMC Medical Genomics*, 8(1):1–12, 2015.

[4] D. F. Andrews. Plots of high-dimensional data. *Biometrics*, pages 125–136, 1972.

[5] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, et al. Gene ontology: tool for the unification of biology. *Nature genetics*, 25(1):25, 2000.

[6] G. Asher and P. Sassone-Corsi. Time for food: The intimate interplay between nutrition, metabolism, and the circadian clock. *Cell*, 161(1):84–92, 2015. [PubMed:25815987] [doi:10.1016/j.cell.2015.03.015].

[7] A. Azzi, R. Dallmann, A. Casserly, H. Rehrauer, A. Patrignani, B. Maier, A. Kramer, and S. A. Brown. Circadian behavior is light-reprogrammed by plastic DNA methylation. *Nature Neuroscience*, 17(3):377–382, 2014. [PubMed:24531307] [doi:10.1038/nn.3651].

[8] P. Baldi and A. D. Long. A bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. *Bioinformatics*, 17(6):509–519, 2001.

[9] J. Bass. Circadian topology of metabolism. *Nature*, 491(7424):348–356, 2012. [PubMed:23151577] [doi:10.1038/nature11704].

[10] N. L. Bray, H. Pimentel, P. Melsted, and L. Pachter. Near-optimal probabilistic rna-seq quantification. *Nature biotechnology*, 34(5):525, 2016.

[11] E. G. Cerami, B. E. Gross, E. Demir, I. Rodchenkov, Ö. Babur, N. Anwar, N. Schultz, G. D. Bader, and C. Sander. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Research*, 39(SUPPL. 1):685–690, 2011.

[12] K. Daily, V. R. Patel, P. Rigor, X. Xie, and P. Baldi. MotifMap: integrative genome-wide maps of regulatory motif sites for model species, 2011.

[13] R. Dallmann, A. U. Viola, L. Tarokh, C. Cajochen, and S. A. Brown. The human circadian metabolome. *Proceedings of the National Academy of Sciences*, 109(7):2625–2629, 2012. [PubMed:22308371] [PubMed Central:22308371] [doi:10.1073/pnas.1114410109].

[14] C. Dibner, U. Schibler, and U. Albrecht. *The Mammalian Circadian Timing System: Organization and Coordination of Central and Peripheral Clocks*, volume 72. 2010. [PubMed:20148687] [doi:10.1146/annurev-physiol-021909-135821].

[15] K. L. Eckel-Mahan, V. R. Patel, S. De Mateo, R. Orozco-Solis, N. J. Ceglia, S. Sahar, S. A. Dilag-Penilla, K. A. Dyar, P. Baldi, and P. Sassone-Corsi. Reprogramming of the circadian clock by nutritional challenge. *Cell*, 155(7):1464–1478, 2013. [PubMed:24360271] [PubMed Central:24360271] [doi:10.1016/j.cell.2013.11.034].

[16] K. L. Eckel-Mahan, V. R. Patel, R. P. Mohney, K. S. Vignola, P. Baldi, and P. Sassone-Corsi. Coordination of the transcriptome and metabolome by the circadian clock. *Proceedings of the National Academy of Sciences*, 109(14):5541–5546, 2012. [PubMed:22431615] [PubMed Central:PMC3325727] [doi:10.1073/pnas.1118726109].

[17] D. Feng and M. A. Lazar. Clocks, Metabolism, and the Epigenome. *Molecular Cell*, 47(2):158–167, jul 2012. [PubMed:22841001] [PubMed Central:PMC3408602] [doi:10.1016/j.molcel.2012.06.026].

[18] L. Fu and C. C. Lee. The circadian clock: Pacemaker and tumour suppressor. *Nature Reviews Cancer*, 3(5):350–361, 2003. [PubMed:12724733] [doi:10.1038/nrc1072].

[19] J. R. Gerstner, L. C. Lyons, K. P. Wright, D. H. Loh, O. Rawashdeh, K. L. Eckel-Mahan, and G. W. Roman. Cycling Behavior and Memory Formation. *Journal of Neuroscience*, 29(41):12824–12830, oct 2009. [PubMed:19828795] [PubMed Central:PMC4077269] [doi:10.1523/JNEUROSCI.3353-09.2009].

[20] C. S. Greene, J. Tan, M. Ung, J. H. Moore, and C. Cheng. Big data bioinformatics. *Journal of Cellular Physiology*, 229(12):1896–1900, 2014.

[21] J. A. Haspel, S. Chettimada, R. S. Shaik, J. H. Chu, B. A. Raby, M. Cernadas, V. Carey, V. Process, G. M. Hunninghake, E. Ifedigbo, J. A. Lederer, J. Englert, A. Pelton, A. Coronata, L. E. Fredenburgh, and A. M. Choi. Circadian rhythm reprogramming during lung inflammation. *Nature Communications*, 5(May):1–15, 2014. [PubMed:25208554] [PubMed Central:PMC4162491] [doi:10.1093/10.1038/ncomms5753].

[22] D. Howe and S. Yon. The future of biocuration. *Nature*, 455(7209):47–50, 2008.

[23] D. W. Huang, B. T. Sherman, and R. A. Lempicki. Systematic and integrative analysis of large gene lists using david bioinformatics resources. *Nature protocols*, 4(1):44, 2008.

[24] M. E. Hughes, K. C. Abruzzi, R. Allada, R. Anafi, A. B. Arpat, G. Asher, P. Baldi, C. De Bekker, D. Bell-Pedersen, J. Blau, et al. Guidelines for genome-scale analysis of biological rhythms. *Journal of biological rhythms*, 32(5):380–393, 2017. [PubMed:29098954],[PubMed Central:PMC5692188][doi:10.1177/0748730417728663].

[25] M. E. Hughes, J. B. Hogenesch, and K. Kornacker. JTK_CYCLE: An Efficient Nonparametric Algorithm for Detecting Rhythmic Components in Genome-Scale Data Sets. *Journal of Biological Rhythms*, 25(5):372–380, oct 2010. [PubMed:19772347] [PubMed Central:PMC2761978] [doi:10.1177/0748730410379711].

[26] A. Kamburov, U. Stelzl, H. Lehrach, and R. Herwig. The consensuspathdb interaction database: 2013 update. *Nucleic acids research*, 41(D1):D793–D800, 2012.

[27] M. Kanehisa and S. Goto. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1):27–30, 2000.

[28] M. Kanehisa and S. Goto. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1):27–30, 2000. [PubMed:10592173], [PubMed Central: PMC102409],[doi:10.1093/nar/28.1.27].

[29] H. Kashyap, H. A. Ahmed, N. Hoque, S. Roy, and D. K. Bhattacharyya. Big Data Analytics in Bioinformatics: A Machine Learning Perspective. 13(9):1–20, 2015.

[30] M. A. Kayala and P. Baldi. Cyber-t web server: differential analysis of high-throughput data. *Nucleic acids research*, 40(W1):W553–W559, 2012.

[31] C. H. Ko and J. S. Takahashi. Molecular components of the mammalian circadian clock. *Human Molecular Genetics*, 15(SUPPL. 2):271–277, 2006. [PubMed:16987893] [doi:10.1093/hmg/ddl207].

[32] N. Koike, S.-H. Yoo, H.-C. Huang, V. Kumar, C. Lee, T.-K. Kim, and J. S. Takahashi. Transcriptional Architecture and Chromatin Landscape of the Core Circadian Clock in Mammals. *Science*, 338(6105):349–354, oct 2012. [PubMed:22936566] [PubMed Central:PMC3694775] [doi:10.1126/science.1226339].

[33] S. Y. Krishnaiah, G. Wu, B. J. Altman, J. Growe, S. D. Rhoades, F. Coldren, A. Venkataraman, A. O. Olarerin-George, L. J. Francey, S. Mukherjee, S. Girish, C. P. Selby, S. Cal, U. ER, B. Sianati, A. Sengupta, R. C. Anafi, I. H. Kavakli, A. Sancar, J. A. Baur, C. V. Dang, J. B. Hogenesch, and A. M. Weljie. Clock Regulation of Metabolites Reveals Coupling between Transcription and Metabolism. *Cell Metabolism*, 25(5):1206, may 2017. [PubMed:28380384],[doi:10.1016/j.cmet.2017.03.019].

[34] X. M. Li, F. Delaunay, S. Dulong, B. Claustrat, S. Zampera, Y. Fujii, M. Teboul, J. Beau, and F. Lévi. Cancer inhibition through circadian reprogramming of tumor transcriptome with meal timing. *Cancer Research*, 70(8):3351–3360, 2010. [PubMed:20395208] [doi:10.1158/0008-5472.CAN-09-4235].

[35] Y. Liu, S. Sun, T. Bredy, M. Wood, R. C. Spitale, and P. Baldi. MotifMap-RNA: a genome-wide map of RBP binding sites. *Bioinformatics*, 33(13):2029–2031, jul 2017.

[36] L. Maaten. Learning a parametric embedding by preserving local structure. In *Artificial Intelligence and Statistics*, pages 384–391, 2009.

[37] S. Masri, T. Papagiannakopoulos, K. Kinouchi, Y. Liu, M. Cervantes, P. Baldi, T. Jacks, and P. Sassone-Corsi. Lung Adenocarcinoma Distally Rewires Hepatic Circadian Homeostasis. *Cell*, 165(4):896–909, 2016. [PubMed:27153497] [PubMed Central:PMC5373476] [doi:10.1016/j.cell.2016.04.039].

[38] S. Masri, V. R. Patel, K. L. Eckel-Mahan, S. Peleg, I. Forne, A. G. Ladurner, P. Baldi, A. Imhof, and P. Sassone-Corsi. Circadian acetylome reveals regulation of mitochondrial metabolic pathways. *Proceedings of the National Academy of Sciences*, 110(9):3339–3344, 2013. [PubMed:23341599] [PubMed Central:PMC4908327] [doi:10.1073/pnas.1217632110].

[39] J. S. Menet and M. Rosbash. When brain clocks lose track of time: cause or consequence of neuropsychiatric disorders. *Current Opinion in Neurobiology*, 21(6):849–857, dec 2011. [PubMed:21737252] [PubMed Central:PMC3252427] [doi:10.1016/j.conb.2011.06.008].

[40] M. Müller. Dynamic time warping. *Information retrieval for music and motion*, pages 69–84, 2007.

[41] M. Murakami, P. Tognini, Y. Liu, K. L. EckelMahan, P. Baldi, and P. SassoneCorsi. Gut microbiota directs PPARγdriven reprogramming of the liver circadian clock by nutritional challenge. *EMBO reports*, 17(9):1292–1303, 2016. [PubMed:27418314] [PubMed Central:PMC5007574] [doi:10.15252/embr.201642463].

[42] S. C. F. A. K. E.-M. P. S.-C. P. F. B. Nicholas Ceglia, Yu Liu. CircadiOmics: Circadian Omic Web Portal. *Nucleic Acid Research*, 2018.

[43] S. Panda, J. B. Hogenesch, and S. A. Kay. Circadian rhythms from flies to human. *Nature*, 417(6886):329–335, 2002. [PubMed:12015613] [doi:10.1038/417329a].

[44] C. L. Partch, C. B. Green, and J. S. Takahashi. Molecular architecture of the mammalian circadian clock. *Trends in Cell Biology*, 24(2):90–99, 2014. [PubMed:23916625] [PubMed Central:PMC3946763] [doi:10.1016/j.tcb.2013.07.002].

[45] V. R. Patel, N. Ceglia, M. Zeller, K. Eckel-Mahan, P. Sassone-Corsi, and P. Baldi. The pervasiveness and plasticity of circadian oscillations: The coupled circadian-oscillators framework. *Bioinformatics*, 31(19):3181–3188, 2015. [PubMed:26049162] [PubMed Central:PMC4592335] [doi:10.1093/bioinformatics/btv353].

[46] M. S. Robles, J. Cox, and M. Mann. In-Vivo Quantitative Proteomics Reveals a Key Contribution of Post-Transcriptional Mechanisms to the Circadian Regulation of Liver Metabolism. *PLoS Genetics*, 10(1), 2014. [PubMed:24391516] [PubMed Central:PMC2840478] [doi:10.1371/journal.pgen.1004047].

[47] T. Roenneberg and M. Merrow. The circadian clock and human health. *Current Biology*, 26(10):R432–R443, 2016. [PubMed:27218855][doi:10.1016/j.cub.2016.04.011].

[48] N. H. Shah and J. D. Tenenbaum. The coming age of data-driven medicine: Translational bioinformatics' next frontier. *Journal of the American Medical Informatics Association*, 19(E1):25–27, 2012.

[49] P. Tognini, M. Murakami, Y. Liu, K. L. Eckel-Mahan, J. C. Newman, E. Verdin, P. Baldi, and P. Sassone-Corsi. Distinct circadian signatures in liver and gut clocks revealed by ketogenic diet. *Cell metabolism*, 26(3):523–538, 2017.

[50] C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, and L. Pachter. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, 7(3):562–578, 2012.

[51] J. Ule, G. Stefani, A. Mele, M. Ruggiu, X. Wang, B. Taneri, T. Gaasterland, B. J. Blencowe, and R. B. Darnell. An RNA map predicting Nova-dependent splicing regulation. *Nature*, 444(7119):580–586, 2006.

[52] X. Xie, P. Rigor, and P. Baldi. MotifMap: A human genome-wide map of candidate regulatory motif sites. *Bioinformatics*, 25(2):167–174, 2009.

[53] K. Yagita, K. Horie, S. Koinuma, W. Nakamura, I. Yamanaka, A. Urasaki, Y. Shigeyoshi, K. Kawakami, S. Shimada, J. Takeda, and Y. Uchiyama. Development of the circadian oscillator during differentiation of mouse embryonic stem cells in vitro. *Proceedings of the National Academy of Sciences*, 107(8):3846–3851, 2010. [PubMed:20133594] [PubMed Central:PMC2840478][doi:10.1073/pnas.0913256107].