Tissue-specific circadian transcriptional regulation

Présentée le 18 janvier 2019
à la Faculté des sciences de la vie
Unité du Prof. Naef
Programme doctoral en biotechnologie et génie biologique

pour l’obtention du grade de Docteur ès Sciences

par

JAKE YEUNG

Acceptée sur proposition du jury
Prof. B. Deplancke, président du jury
Prof. F. Naef, directeur de thèse
Prof. J. Hogenesch, rapporteur
Prof. E. van Nimwegen, rapporteur
Prof. J. Lingner, rapporteur
"In the fields of observation luck favors the prepared mind."
— Louis Pasteur

"Beware of finding what you're looking for."
— Richard Hamming

To my parents...
Acknowledgements

During my PhD, I have been fortunate to have worked with the right people at the right time. First of all, this thesis would not have been possible without my advisor Felix. He has been more than I could have hoped for as a mentor. I am grateful for all of our discussions over the last four years on wide topics of science and also the deep but important details. Felix, you have taught me that treasure can be found by getting the details right. Your critical eye and sound judgement have guided me toward paths of exciting discoveries. You inspired me to maintain a high technical level while making the broadest impacts in science. I was always eager to soak up knowledge and seek wisdom from your discussions. If I could have gleaned even a tiny fraction of your wisdom and insights through osmosis over the last few years, I would be delighted. I would also like to thank Fred and Paul for having confidence in me to work on a variety of projects. Paul, thank you for showing me the world of sleep and for broadening my view of biology. Fred, I benefited from your deep insights on metabolism, physiology, and the circadian clock. Your encyclopedic knowledge of biology is inspirational.

I would also like to thank my colleagues over the years who have contributed to my growth over the last few years, Jingkui, Cédric, Daniel, Benjamin, Saeed, Nick, Damien, Clémence, and Eric. Our discussions have inspired many ideas. Special thanks to Sophie for helping me with administrative tasks, taking them off my plate so I can focus on my work. I would like to thank the facilities, both experimental and computational, for providing expertise and insights to technical problems.

I would like to give special thanks to Jérome for a fruitful collaboration over the last few years. Your bold thinking and ability to sense the right experiments to do have benefited me greatly. I owe you a debt of gratitude for all of your work. Thank you also to Charlotte for letting me in on your experiments and inspiring new ways of thinking about dynamics with your experimental designs. Thank you Maxime for your work and analysis, which revealed treasure where I was unable to find on my own.

I would like to thank my family and friends both here and abroad for support over the years. Last but certainly not least, I would like to thank Bingqing for going on this journey with me these last several years. You have given me structure in my life and allowed me to focus on the important things. Thank you for interweaving your life with mine. Together this bond makes us stronger.

Lausanne, 23 August 2018

Jake Yeung
Abstract
Circadian rhythms in physiology and behavior evolved to resonate with daily cycles in the external environment. In mammals, organs orchestrate temporal physiology over the 24-hour day, which requires extensive gene expression rhythms targeted to the right tissue. Although a core set of gene products oscillate across virtually all cell-types, gene expression profiling across tissues over the 24-hour day showed that rhythmic gene expression programs are tissue-specific. We highlight recent progress in uncovering how the circadian clock interweaves with tissue-specific gene regulatory networks involving functions such as xenobiotic metabolism, glucose homeostasis, and sleep. This progress hinges on not only comprehensive experimental approaches but also computational methods for multivariate analysis of periodic functional genomics data. This thesis first explores how circadian gene expression is regulated across tissues. Second, we investigate how dynamic chromatin interactions underlie circadian gene transcription, core clock functions, and ultimately behavior. Third, we elucidate how the temporal transcriptome in mouse cortex responds to sleep deprivation. Finally, we discuss perspectives on extending the knowledge of the circadian clock in mice to human chronobiology.

Key words: systems chronobiology, chromatin interactions, gene regulation, feeding-fasting cycles, sleep-wake cycles
Résumé

Les rythmes circadiens de la physiologie ont évolué pour correspondre aux cycles quotidiens de l’environnement externe. Chez les mammifères, les organes orchestrent la physiologie temporelle sur une journée de 24 heures, ce qui nécessite des rythmes d’expression génique dans le bon tissu. Bien qu’un ensemble de base de produits génétiques oscillent dans pratiquement tous les types de cellules, le profil d’expression génétique dans différents tissus au cours des 24 heures a montré que certains programmes d’expression de gènes rhythmiques sont spécifique à des tissus. Nous mettons en évidence les progrès récents dans la découverte de l’imbrication de l’horloge circadienne avec les réseaux de régulation génique spécifiques aux tissus impliquant des fonctions telles que le métabolisme xénobiotique, l’homéostasie du glucose et le sommeil. Ce progrès repose non seulement sur des approches expérimentales complètes, mais aussi sur des méthodes informatiques pour l’analyse multivariée des données périodiques sur la génomique fonctionnelle. Cette thèse explore d’abord comment l’expression des gènes circadiens est régulée à travers les tissus. Deuxièmement, nous étudions comment les interactions de la chromatine dynamique sous-tendent la transcription des gènes circadiens, les fonctions de base de l’horloge et, en fin de compte, le comportement. Troisièmement, nous expliquons comment le transcriptome temporal dans le cortex de souris répond au traitement de privation de sommeil. Enfin, nous discutons des perspectives d’extension de la connaissance de l’horloge circadienne chez la souris à la chronobiologie humaine.

Mots clefs : systems chronobiology, chromatin interactions, gene regulation, feeding-fasting cycles, sleep-wake cycles
Contents

Acknowledgements i

Abstract (English/Français) iii

1 Introduction 1
   1.1 Circadian rhythms at all scales of biological organization ............... 1
      1.1.1 Mammalian circadian timing is organized as a hierarchy of distributed
clocks ......................................................... 2
      1.1.2 The molecular components of the circadian clock ...................... 2
   1.2 The function of the circadian clock is tissue-specific ..................... 3
   1.3 Regulatory mechanisms underlying tissue-specific rhythmic gene expression . 4
      1.3.1 Interactions between tissue-specific transcription factors and clock regu-
      lators ............................................................ 4
      1.3.2 Rhythmic systemic signals can drive tissue-specific rhythms in gene ex-
      pression .......................................................... 6
      1.3.3 Integrating temporal analysis of WT mice with clock-deficient mutants . 6
   1.4 Statistical analysis of rhythms ...................................................... 7
      1.4.1 Multivariate and modular analysis of rhythmic gene expression ...... 7
      1.4.2 Higher harmonics ................................................. 7
      1.4.3 Cell-type heterogeneity .............................................. 9
      1.4.4 Rhythmic gene expression in humans .................................... 9
      1.4.5 Dynamic responses to acute perturbations ................................ 9
      1.4.6 Mapping chromatin interactions ...................................... 10
      1.4.7 Chromatin interactions and the circadian clock ....................... 10
      1.4.8 Dynamic chromatin interactions in cell culture models ............... 10
      1.4.9 Tissue-specific and dynamic chromatin interactions in tissues ...... 11
      1.4.10 Dynamic chromatin interactions and transcription .................. 11
   1.5 Specific questions addressed in thesis .......................................... 13

2 Transcription factor activity rhythms and tissue-specific chromatin interactions ex-
plain circadian gene expression across organs 15
   2.1 Introduction ................................................................. 15
   2.2 Extended methods: model selection with Zellner’s g−priors ............... 15
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.1 Enumerating harmonic regression models to identify possible combinations of rhythms</td>
<td>15</td>
</tr>
<tr>
<td>across tissues</td>
<td></td>
</tr>
<tr>
<td>2.2.2 Selecting models by the Akaike Information Criterion (AIC), the Bayesian Information</td>
<td>17</td>
</tr>
<tr>
<td>Criterion (BIC), and Bayes Factors</td>
<td></td>
</tr>
<tr>
<td>2.3 Extended methods: complex-valued singular value decomposition</td>
<td>20</td>
</tr>
<tr>
<td>2.4 Extended methods: Gene ontology (GO) analysis around the clock</td>
<td>21</td>
</tr>
<tr>
<td>2.5 Contributions</td>
<td>22</td>
</tr>
<tr>
<td>2.6 Conclusion and perspectives</td>
<td>52</td>
</tr>
<tr>
<td>3 Clock-dependent chromatin topology modulates circadian transcription and behavior</td>
<td>53</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>53</td>
</tr>
<tr>
<td>3.2 Contributions</td>
<td>53</td>
</tr>
<tr>
<td>3.3 Conclusion and perspectives</td>
<td>86</td>
</tr>
<tr>
<td>4 Dynamic gene expression and regulation from circadian and sleep-wake processes in the</td>
<td>87</td>
</tr>
<tr>
<td>mouse cortex</td>
<td></td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>87</td>
</tr>
<tr>
<td>4.2 Contributions</td>
<td>87</td>
</tr>
<tr>
<td>4.3 Conclusion and perspectives</td>
<td>142</td>
</tr>
<tr>
<td>5 Conclusions and Perspectives</td>
<td>143</td>
</tr>
<tr>
<td>Bibliography</td>
<td>161</td>
</tr>
<tr>
<td>Curriculum Vitae</td>
<td>163</td>
</tr>
</tbody>
</table>
1 Introduction

Overview of the thesis

The thesis begins with an introduction that uses ideas and text that are discussed in a review paper in revision. References in this section to Yeung et al. and Mermet et al. highlight how the work in this thesis fits into the current challenges of chronobiology (Yeung et al., 2018; Mermet et al., 2018). This section concludes with three broad questions tackled in the thesis. After the introduction are three papers investigating regulation of gene regulatory dynamics. The first paper (Yeung et al., 2018) focuses on tissue-specific regulation. The second paper (Mermet et al., 2018) goes into details on the role of chromatin interactions in regulating the circadian clock. And the third paper, under preparation, explores how perturbing oscillatory gene expression by a short sleep deprivation treatment can reveal insights into how dynamic gene expression is regulated in the mouse cortex. The thesis concludes with the broad significance of my work and outstanding questions for the field.

1.1 Circadian rhythms at all scales of biological organization

Many organisms exhibit rhythms in behavior and physiology that are synchronized to the daily cycles in the environment (Bell-Pedersen et al., 2005). This internal circadian rhythm (period length of around one day) resonates with daily cycles, predicting and adapting to external cues such as light abundance and food availability (Schibler et al., 2015). Competition experiments in cyanobacteria suggest that this resonance enhances fitness (Ouyang et al., 1998). In mammals, a genetically encoded molecular clock oscillates in virtually every cell of the body (Takahashi, 2017). These clocks are synchronized to an internal body rhythm set by the master pacemaker, located in the suprachiasmatic nucleus (Hastings et al., 2018). Chronobiology researchers study how 24-hour periodicity in external signals is dynamically integrated at all scales of biological organization, from gene expression to behavior, promising to yield new insights on how temporally structured external signals such as artificial light, unconventional meal times, and drug therapies dynamically interact with circadian rhythms.
1.1.1 Mammalian circadian timing is organized as a hierarchy of distributed clocks

In mammals, the circadian timing system is organized as a hierarchy of oscillators, containing a master clock and peripheral clocks. The master clock, located in the suprachiasmatic nucleus (SCN), establishes rhythms in behavior by synchronizing peripheral clocks, located outside of the SCN, through systemic signals such as hormones, body temperature, and direct innervation (Mohawk et al., 2012). As insinuated by their name, circadian rhythms (’circa diem’ means ‘approximately a day’) do not have a period of exactly 24 hours but differ between species (Bell-Pedersen et al., 2005), individuals as well as age (Czeisler et al., 1999). For example, the free-running circadian period for humans is longer than 24 hours while for the mouse is shorter. Therefore, the circadian timing cannot measure 24 h with accuracy (although it is very precise) but is periodically synchronized to the environment by externals cues. These rhythmic cues are called zeitgebers. A prominent one is external sunlight light, which synchronizes the central clock in the SCN to geophysical time. Other zeitgebers include external temperature and feeding-fasting rhythms, which contribute to entrainment of clocks in tissues outside of the SCN (Damiola et al., 2000; Saini et al., 2011).

1.1.2 The molecular components of the circadian clock

Single-cell analyses have shown that circadian oscillations rely on a cell-autonomous clock that is genetically encoded, suggesting that potentially every cell in the body resonates with environmental time (Nagoshi et al., 2004). A widespread, though not necessarily definitive, model of the mammalian circadian clock consists of interlocked transcriptional and translational feedback loops that regulate the expression of core clock genes (Dibner et al., 2010). The genetic components and regulation of the circadian clock have been discussed in several excellent reviews (Takahashi, 2017; Mohawk et al., 2012; Dibner et al., 2010). Briefly, activating transcription factors (TFs) drive expression of many genes, including its own repressors, which decreases the accumulation of activating TFs and subsequently its repressors. The low accumulation of negative TFs derepress the activating TFs and a new cycle begins. A simplified diagram of two interlocked negative feedback loops is shown in Figure 1.1. Core clock transcription factors BMAL1 and CLOCK heterodimerize (BMAL1-CLOCK) and activate transcription of genes containing E-box (enhancer box) cis-regulatory elements. Important targets of BMAL1-CLOCK are the core clock genes, *Period* (*Per1, Per2, Per3*) and *Cryptochrome* (*Cry1, Cry2*), whose protein products (CRY-PER complexes) directly abrogates the transcriptional activity of BMAL1-CLOCK (Lowrey and Takahashi, 2004). Additionally, BMAL1-CLOCK activates transcription of another clock gene *Rev-erba*, whose protein product binds to ROREs (ROR response elements) and thus inhibits transcription of the *Bmal1* gene (Partch et al., 2014). These transcriptional and translational negative feedback loops establish the circadian clock (Yoo et al., 2004).
1.2. The function of the circadian clock is tissue-specific

Although this molecular clock ticks in virtually all cells of the body (Dibner et al., 2010), early transcriptome profiling studies around the clock in different mouse tissues have found that rhythmic gene expression is highly tissue-specific (Storch et al., 2002). How this clock is used

Extensive studies mutating different components of the clock in mouse have identified a wide range of period phenotypes. A comprehensive table in a review by Ko and Takahashi summarizes clock mutants and their period phenotypes (Ko and Takahashi, 2006). For example, Cry1 knockout (KO) mice show a 1-hour shorter period while Cry2 KO mice show a 1-hour longer period. Bmal1 KO mice are arrhythmic. These studies show that the circadian clock relies on a complex molecular architecture.

1.2. The function of the circadian clock is tissue-specific
Chapter 1. Introduction

in different cell types and tissues to regulate diverse physiological processes is at the core of contemporary chronobiology. Daily rhythms in cell-type specific functions are pervasive. For example, recent work showed that fibroblasts exhibit circadian rhythm in actin dynamics, allowing wound healing to be more efficient during the active versus resting phase (Hoyle et al., 2017). Size of liver cells, its protein accumulation, and ribosome number oscillate over the 24-hour day (Gerber et al., 2013; Sinturel et al., 2017), making metabolism and xenobiotic detoxification more efficient when animals are active and feed. Macrophages exhibit circadian rhythm in cytokine production (Keller et al., 2009), which may improve response against bacterial infections.

Intriguingly, disrupting the clock in different tissues can lead to opposite phenotypes (Bass and Lazar, 2016). For example, pancreas-specific ablation of the clock leads to hyperglycaemia (Marcheva et al., 2010) while liver-specific ablation leads to hypoglycaemia (Lamia et al., 2008). Alternatively, tissue-specific rescue of the clock in whole-body knockouts can also reveal surprises. For example, whole-body Bmal1 KO mice have increased non-rapid eye movement (NREM) sleep, and restoring Bmal1 expression specifically in skeletal muscles can rescue the amount of NREM sleep, although not the timing (Ehlen et al., 2017). These tissue-specific phenomena are likely underpinned by tissue-specific rhythmic gene expression signatures. Indeed, a high-resolution circadian gene expression atlas across twelve organs has shown that mRNA abundances oscillate mostly in an organ-specific manner (Zhang et al., 2014), indicating that the circadian clock regulatory network interweaves with tissue-specific gene regulatory mechanisms.

1.3 Regulatory mechanisms underlying tissue-specific rhythmic gene expression

The molecular circadian clock and systemic signals likely contribute to tissue-specific rhythmic gene expression (Hughes et al., 2012; Kornmann et al., 2007) (Figure 1.2A). What are the regulatory mechanisms? Liver has been a productive model to study temporal gene regulation across multiple omics data types, from transcriptional (Le Martelot et al., 2012; Sobel et al., 2017; Trott and Menet, 2018), post-transcriptional (Luck et al., 2014), translational (Atger et al., 2015), to post-translational (Mauvoisin et al., 2017; Robles et al., 2017; Wang et al., 2017). Extending such analyses to multiple tissues will uncover general principles of how tissues physiologically interact with each other and with the circadian clock.

1.3.1 Interactions between tissue-specific transcription factors and clock regulators

Recently, temporal RNA-seq analysis of multiple mouse tissues suggested that interactions between tissue-specific and a clock transcription factor (TF) underlies tissue-specific rhythmic transcription regulation (Yeung et al., 2018) (Figure 1.2B). The binding of a tissue-specific
1.3. Regulatory mechanisms underlying tissue-specific rhythmic gene expression

Figure 1.2 – **Generation of tissue-specific rhythms.** (A) Scheme of how external rhythmic cues entrain circadian rhythms across tissues. The central clock, located in the suprachiasmatic nucleus of the hypothalamus, takes light from the environment and synchronizes clocks in peripheral tissues through by regulating locomotor activity and feeding rhythms as well as through systemic signals such as hormones and metabolites. Peripheral tissues, such as kidney, heart, and liver, orchestrate rhythms in tissue physiology, such as sodium homeostasis, carbohydrate metabolism, and blood pressure, respectively. (B,C) Examples of how tissue-specific rhythmic gene expression can be generated transcriptionally. (B) Interactions between tissue-specific and clock transcription factors (TFs) can generate tissue-specific gene expression. For example, a gene can be rhythmically transcribed in one tissue but not in others by the presence or absence of a tissue-specific TF, which renders a nearby clock TF binding site accessible. (C) Different tissues can regulate rhythmic transcription of a gene by using different alternative promoters. In this example, the rhythmically transcribed promoter is used in liver but not in other tissues. A rhythmically transcribed gene is shown with a sinusoid, flat transcription is shown with a flat horizontal line.
TF may render the local chromatin region more accessible, increasing the affinity of clock TFs to their respective binding sites. For example, liver-specific rhythmic gene expression has been found to be enriched for liver-specific DNase-I hypersensitive sites containing binding sites for both liver-specific and clock TFs (Yeung et al., 2018). Direct protein-protein interactions between clock and other transcriptional regulators may also regulate rhythmic gene expression in a tissue-specific manner. For example, co-immunoprecipitation studies examining interactions of mouse nuclear receptors (NR) with CRY1 found that one-third of mouse NRs, some of which were expressed in a tissue-specific manner such as PXR and CAR (which regulates xenobiotic detoxification in liver) interact with CRY1 (Kriebs et al., 2017; Lamia et al., 2011). Another mechanism generating specificity in transcription dynamics is alternative promoter usage, where one transcript is rhythmically transcribed while the other is not, which explains a fraction of tissue-specific rhythmic transcript abundances (Figure 1.2C) (Yeung et al., 2018). Further analyses estimating the frequency of each mechanism that defines the rhythmic transcriptome will clarify which regulatory modes are most prevalent.

1.3.2 Rhythmic systemic signals can drive tissue-specific rhythms in gene expression

Generally, it is still unclear how rhythmic systemic signals such as hormones and metabolites can drive responses of varying magnitudes or amplitudes, and also activate different downstream pathways depending on the tissue context (Bass and Takahashi, 2010) (Figure 1.2A). One explanation is that signaling molecules such as hormones act in a tissue-specific manner; for example, fibroblast growth factor hormones regulate bile acid homeostasis in liver but contribute to thermogenesis in brown adipose tissue (Owen et al., 2015). Temporal profiles of serum concentrations could also differ across the body; for example, human IL-6 levels in cerebrospinal fluid versus plasma show distinct rhythms and peaked at different times of day (Agorastos et al., 2014), suggesting that dynamics in permeability of the blood-brain barrier can regulate distribution of cytokines across the body (Pan and Kastin, 2017). Thus tissue-specific decoding of systemic signals or local variation in ligand concentrations can induce transcription factor activities that oscillate in one tissue but not others.

1.3.3 Integrating temporal analysis of WT mice with clock-deficient mutants

Since rhythmic gene expression in organs can be generated through systemic cues as well as by the local clock (Hughes et al., 2012; Kornmann et al., 2007; Lamia et al., 2008), study designs incorporating wild-type (WT) and clock-deficient mutants allow clock-dependent and independent mechanisms to be distinguished (Atger et al., 2015), as demonstrated in a study of WT and Bmal1 KO clock mutants in liver and kidney (Yeung et al., 2018).
1.4 Statistical analysis of rhythms

1.4.1 Multivariate and modular analysis of rhythmic gene expression

Many genome-wide descriptions of circadian or diurnal rhythmicity have considered genes one at a time, and typically in one or few conditions (Hutchison et al., 2015). To assess rhythmic gene expression in a given condition, parametric (Atger et al., 2015; Fisher, 1929) and non-parametric (Hughes et al., 2010; Hutchison et al., 2015; Thaben and Westermark, 2014) methods have been used to statistically test, for each gene, whether a measured temporal pattern shows evidence for rhythmicity compared to a null model (e.g. flat gene expression in time).

Today, an increasing number of datasets include multiple conditions, such as multiple tissues, genotypes, feeding conditions, or combinations thereof (Korenčič et al., 2014; Atger et al., 2015; Zhang et al., 2014). Analysis methods that integrate oscillatory patterns across multiple conditions could reveal novel patterns not easily identified from standard analyses. One approach that extends classic tests to two conditions is the Chow test (Chow, 1960). Beyond two conditions, there is the model selection method, where multi-condition data is fit to a set of models representing combinations of rhythmic and non-rhythmic outcomes across the conditions. The best model is identified by balancing goodness of fit and model complexity, such as the Bayesian Information Criterion (Figure 1.3A) (Atger et al., 2015).

Another area of interest is constructing low-dimensional representations of periodic data across tissues. One method, complex-valued singular value decomposition (cv-SVD), extends conventional real-valued SVD analysis to periodic data. cv-SVD can identify gene sets with large 24-hour amplitudes and phase shifts across many conditions (Figure 1.3B, details in Section 2.3) (Yeung et al., 2018). For example, modules of genes with coherent phase and amplitude relations, which occurs often in chronobiology datasets, can be adequately captured in low-dimensional complex-valued representations. Other methods, such as Zeitzeiger (Hughey et al., 2016), can identify sparse sets of genes that oscillate across conditions which has been successfully applied to a NanoString assay, called BodyTime, to predict internal circadian time from blood monocytes (Wittenbrink et al., 2018). Interestingly, new methods such as Oscope and CYCLOPs can now reconstruct cyclic dynamics from unlabeled data. This enables identifying oscillating genes in datasets without explicit time labels, such as in large-scale human tissue gene expression datasets (Anafi et al., 2017; Leng et al., 2015; Ruben et al., 2018).

1.4.2 Higher harmonics

Although 24-hour rhythms contribute the largest temporal variance in circadian datasets across tissues, 12-hour rhythms (also classified as ultradian rhythms) are often the second largest (Hughes et al., 2009; Yeung et al., 2018). Methods that include linear combinations of higher harmonics can systematically analyze ultradian dynamics, defined as a period length
Figure 1.3 – **Computational methods integrating multiple genes and conditions.** (A) Model selection methods can integrate rhythms across multiple conditions (e.g., tissues). In this toy example, noisy data (shown as a roughened sinusoid) come from rhythmic gene expression from two tissues that oscillate in synchrony. The data is fit to each rhythmic model and the best model is selected. In this example, the best model is a model of intermediate complexity where rhythmic parameters are shared across two tissues. More complex models fit the data better, but are penalized for having more parameters. (B) Complex-valued singular-value decomposition (cv-SVD) allows rhythmic gene expression datasets to be factorized into gene space and tissue spaces. This representation identifies modules of genes with coherent phase and amplitude relationships across tissues. (i) First, the 3-dimensional matrix of real numbers are projected onto the 24-hour Fourier component, resulting in a 2-dimensional matrix of complex numbers (represented as arrows pointing to time of peak gene expression, length of arrow represent oscillation amplitudes). (ii) The 2D matrix is then decomposed into a gene space and tissue space by SVD (conventional implementations, for example in R, can handle complex numbers). The gene space shows the amplitude and phase for a reference tissue. Amplitude and phase of genes in other tissues are multiplied and added, respectively, based on the tissue loading. In the reality, the tissue space is often in phase. The antiphase example highlights the method.
shorter than a day (Costa et al., 2013; Zhu et al., 2017). In liver, clock-dependent (Cretenet et al., 2010; Westermark and Herzel, 2013) and independent (Zhu et al., 2017) mechanisms of 12-hour rhythms have been studied. Mechanisms include pairs of clock TFs (Westermark and Herzel, 2013), or ultradian regulation of stress response pathways such as the unfolded protein response in the endoplasmic reticulum (Cretenet et al., 2010). Some of these ultradian rhythms consist of two peaks of mRNA expression per day, which interestingly, revert to 24-hour rhythms in clock mutants (Atger et al., 2015; Cretenet et al., 2010; Hughes et al., 2009), indicating that the two peak times are differentially controlled by the clock and feeding cycles. Extending the analysis of higher harmonics to multiple tissues may reveal interactions between tissue-specific TFs with regulators of ultradian rhythmicity.

1.4.3 Cell-type heterogeneity

The degree of cell type heterogeneity and structure varies from tissue to tissue (Han et al., 2018; Yang et al., 2014), which needs to be considered if this heterogeneity changes over time (Scheiermann et al., 2013). Certain immune cell types, such as T-cells and macrophages, are recruited to tissues in a circadian manner (Keller et al., 2009; Scheiermann et al., 2012), suggesting that cellular heterogeneity within tissues could fluctuate over the day. When studying circadian oscillations in cytosine modifications, such cellular heterogeneity were taken into account as potential confounding factors (Oh et al., 2018).

1.4.4 Rhythmic gene expression in humans

Recently, an increasing number of gene expression datasets of human tissue samples have become available (Ardlie and Guigó, 2017), most of which are sampled without explicit regard to the time of day. Interestingly, it is possible to predict internal circadian time in gene expression in humans by computationally assigning time labels to unlabeled samples using different statistical learning approaches (Anafi et al., 2017; Hughey et al., 2016; Leng et al., 2015; Ruben et al., 2018; Wittenbrink et al., 2018). For example, knowledge in mice about which genes are rhythmically transcribed and in which tissues can provide prior information for analyzing rhythms in human tissue gene expression data, under the assumption that rhythmicity in mRNA expression profile is sufficiently conserved across mammalian evolution (Anafi et al., 2017).

1.4.5 Dynamic responses to acute perturbations

Novel experimental designs may motivate analyses beyond fitting periodic functions. For example, studies have shown that sleep deprivation alters expression of clock genes (Franken et al., 2007). But what are the dynamics of gene expression during and after sleep deprivation? How quickly does the system adapt to a new environment? Is the response tissue-specific? How do these transient dynamics apply to other environmental inputs such as feeding patterns...
Chapter 1. Introduction

(Acosta-Rodríguez et al., 2017)? Most circadian studies use periodic functions to fit gene expression data. However, the mentioned situations require us to think beyond periodic functions and investigate other dynamics that may occur during perturbations (we look at an example of this in Chapter 4). These temporal dynamics in gene expression may be accompanied by dynamics in other gene regulatory layers. We focus on chromatin interactions as a regulatory layer underlying circadian transcription and behavior.

1.4.6 Mapping chromatin interactions

The conformation and organization of chromatin in the nucleus is known to be important for the control of gene regulation (Pombo and Dillon, 2015), particularly gene transcription. Chromatin conformation capture (3C) technique (Dekker et al., 2002) and its high-throughput variants, such as 4C (Simonis et al., 2006), 5C (Nora et al., 2012), Hi-C (Lieberman-Aiden et al., 2009), and ChIA-PET (Zhang et al., 2013), as well as orthogonal techniques such as GAM (Beagrie et al., 2017) and SPRITE (Quinodoz et al., 2018), have revealed the spatial organization of the genome. An emerging picture is that interphase chromosomes are organized in a hierarchy of structural layers (Gibcus et al., 2018). At the largest scale, there are multi-megabase compartments A and B. A compartments are generally gene rich and transcriptionally active while B compartments are gene poor and transcriptionally repressed (van Steensel and Belmont, 2017). Generally, the compartment A occupies a central nuclear position while compartment B associates with the nuclear lamina (Bouwman and de Laat, 2015). Within these compartments, at sub-megabase scale, the chromatin is organized as topologically associating domains (TADs). At the TAD scale, genes show maximal enrichment of coexpression during differentiation from embryonic stem cells, compared to other genomic scales (Zhan et al., 2017). TAD boundaries are often demarcated by binding of cohesion and CTCF proteins oriented in a convergent manner (de Wit et al., 2015). At a fine scale, promoter-enhancer contacts allow transcription factors to regulate distal (often less than 100 kb away) target genes (Gibcus and Dekker, 2013). Generally, TAD interactions are fairly stable across cell-types (Dixon et al., 2012), whereas sub-TAD interactions can be cell-type specific as well as dynamic (Phillips-Cremins et al., 2013).

1.4.7 Chromatin interactions and the circadian clock

1.4.8 Dynamic chromatin interactions in cell culture models

What are the circadian dynamics in the spatial organization of the genome? The first studies applying 4C in the circadian context used cell line models. One study using mouse embryonic fibroblasts to evaluate chromatin contacts to the promoter of Dbp, a clock output gene, reported circadian fluctuations in inter-chromosomal contacts (Aguilar-Arnal et al., 2013). Another study found that the Pard3 locus was rhythmically recruited in the nuclear lamina in a human colon cancer cell line (HCT116) (Zhao et al., 2015). However, these models have not studied the dynamics of promoter-enhancer looping and whether such interactions may have
1.4. Statistical analysis of rhythms

1.4.9 Tissue-specific and dynamic chromatin interactions in tissues

Recent *in vivo* work showed that sub-TAD interactions play a role in circadian gene expression by connecting gene promoters with enhancers (Beytebiere et al., 2018; Kim et al., 2018; Mermet et al., 2018; Xu et al., 2016; Yeung et al., 2018). For example, tissue-specific promoter-enhancer interactions enable the circadian clock to regulate gene expression in a tissue-specific manner (Yeung et al., 2018) (Figure 1.4A), and such regulation can also be dynamic over the course of the day (Figure 1.4B). Moreover, 4C-seq performed around the clock focusing on a core clock gene, *Cry1*, and a clock-output gene, *Gys2*, demonstrated dynamic rhythms in promoter-enhancer contacts coinciding with rhythms in the active enhancer mark, H3K27ac (Mermet et al., 2018). As revealed by 4C-seq experiments in *Bmal1* KO, these contacts lost rhythmicity in mice without a functioning clock. Moreover, CRISPR-Cas9 deletion of the contacted intronic *Cry1* enhancer in mice indicated that the contact rhythms shortens the period of the locomotor activity (Mermet et al., 2018), revealing the dynamic function of a noncoding DNA element that propagates from chromatin interactions, gene expression, to locomotor activity. A Hi-C study performed at two time points in the liver reported chromatin interactions at the sub-TAD scale, which could be dynamic or stable over time (Kim et al., 2018). Consistently, a study employing ChIA-PET argued that promoters of rhythmically active genes form stable contact with enhancers (Beytebiere et al., 2018). It will be interesting to investigate why dynamic transcription of some genes is accompanied by dynamic chromatin looping, while other genes exhibit static looping.

1.4.10 Dynamic chromatin interactions and transcription

Recent studies in cell culture showed that promoter-enhancer looping influence transcriptional parameters such as burst fraction (fraction of active transcription sites in each nucleus, which is related to burst frequency) or burst size (Bartman et al., 2016; Kalo et al., 2015; Senecal et al., 2014). Combined live imaging and smRNA FISH studies to investigate molecular mechanisms found that targeted acetylation of promoter histones increases burst frequency while not affecting burst size. (Nicolas et al., 2018). In liver tissue, bursts in gene transcription have been characterized using smRNA FISH (Bahar Halpern et al., 2015). Combining smRNA FISH with functional genomics assays such as 4C-seq and measurements of chromatin H3K27ac have uncovered how promoter-enhancer looping influences transcriptional bursting in vivo (Mermet et al., 2018). In particular, deleting an enhancer element using CRISPR-Cas9 demonstrated that disrupting promoter-enhancer looping decreases burst fraction and shortens the period of locomotor activity (Figure 1.4C).

How promoter-enhancer looping in single-cells leads to bursts of gene transcription remains unknown. Are promoter-enhancer contacts preceding initiation of transcriptional bursts, or what are the time delays between the two events? Current techniques such as sm-
Figure 1.4 – Chromatin interactions as a regulatory layer underlying circadian gene expression and behavior. (A) Tissue-specific chromatin interactions can regulate clock TFs to bind in a tissue-specific manner. Here, the clock TF binding site is accessible in liver but not in kidney, shown as open and closed chromatin, respectively. (B) Dynamic chromatin interactions underlie circadian transcription of rhythmic genes such as the core clock gene Cry1. Clock object represents a clock TF. Rectangle represents an enhancer element. (C) Deleting a non-coding DNA element downstream of Cry1 in mouse (mutant named Cry1Δe) decreases chromatin interactions between promoter and enhancer, reduces burst fraction (fraction of active transcription start sites per nucleus, which is related to burst frequency), and shortens the period in locomotor activity. Simplified cartoon of two actograms indicate a shorter period of locomotor activity in Cry1Δe versus WT under free-running conditions. During entrained conditions (12-hours light, 12-hours dark, shown as white and black rectangles at the top), the mice are active (activity level in black) during lights off and inactive during lights on. During free-running conditions (in complete darkness), the period of locomotor activity depends on the genotype, which can be calculated by the slope of the red line.
RNA FISH can detect transcriptional bursting \textit{in vivo} by analyzing snapshots of RNA molecules across many cells (Bahar Halpern et al., 2015). But monitoring the same cells to determine the relative timing of promoter-enhancer contacts with transcriptional bursting remains challenging. In the context of molecular clocks, linking gene regulation and transcription at single cells will uncover how noisy gene expression is regulated to produce robust pacemakers at the tissue and organism level.

1.5 Specific questions addressed in thesis

This thesis tackles three questions related to the regulation of dynamic, and more specifically temporal gene expression. First two focus on circadian biology; the third relates to sleep.

1. How does the circadian clock regulate gene expression in a tissue-specific manner?
   - What fraction of variance in multi-tissue circadian gene expression data is of temporal or tissue-specific origin?
   - What transcriptional mechanisms underlie tissue-specific circadian gene expression?
   - How do tissue-specific chromatin landscapes regulate circadian gene expression?

2. What is the role of chromatin interactions in regulating circadian gene expression and circadian rhythms?
   - Are circadian dynamics in gene expression accompanied by dynamics in promoter-enhancer looping?
   - Do dynamics in promoter-enhancer looping require the circadian clock?
   - Does dynamics in promoter-enhancer looping regulate circadian gene transcription and period of locomotor activity?

3. How does diurnal gene expression respond to acute perturbations such as sleep deprivation?
   - What are the gene expression dynamics in response to acute sleep deprivation in mouse cortex?
   - How do the sleep homeostat and circadian processes contribute towards gene expression output?
   - What are possible transcriptional regulators underlying sleep-wake dynamics?
2 Transcription factor activity rhythms and tissue-specific chromatin interactions explain circadian gene expression across organs

2.1 Introduction

This chapter investigates transcriptional mechanisms underlying diurnal gene expression rhythms that can be regulated in a tissue-specific manner. The work has been published in *Genome Research* 2018 under the Creative Commons license and is reproduced here. The main text as well as supplemental figures are below. Due to the size of supplemental tables, these are not attached but can be found on the open online version of the *Genome Research* article. No changes were made to the main text or supplemental figures, which was downloaded from doi:10.1101/gr.222430.117.

The pursuit to uncover transcriptional mechanisms underlying tissue-specific circadian gene expression involved developing and applying a variety of computational and statistical methods. Some of these methods are specific to the problem I tackled while others may be broadly applied to other problems. Here, I will go into more detail of some methods which I think may be of use in other gene expression analysis projects: model selection for identifying different combinations of rhythms across conditions, complex-valued singular value decomposition, and GO term analysis around the clock.

2.2 Extended methods: model selection with Zellner’s $g-priors$

2.2.1 Enumerating harmonic regression models to identify possible combinations of rhythms across tissues

This chapter relies on identifying whether gene expression oscillates in no tissues, one tissue, or a combination of tissues. We use a model selection approach to select the combination of
tissues that fit the data while penalizing for model complexity (i.e., the number of parameters).

To illustrate an example, we take the case of two tissues ($|c|=2$) (Figure 2.1). The most complex model (i.e. most parameters) corresponds to each tissue having its own rhythmic parameters $\beta_{1,c}, \beta_{2,c}$ in addition to the intercepts $\alpha_c$.

$$M_f : \quad Y_{t,c_i} = \alpha_{c_i} + \beta_{1,c_i} \cos(\omega t) + \beta_{2,c_i} \sin(\omega t) + \epsilon, \quad (2.1)$$

where $i = 1, 2$ denote the index of tissues.

The restricted model, meaning both tissues are flat, would have only the intercept parameters $(\alpha_{c_1}, \alpha_{c_2})$:

$$M_f : \quad Y_{t,c_i} = \alpha_{c_i} + \epsilon. \quad (2.2)$$

For two tissues, there will be 3 intermediate models, each with 4 parameters. The first two models ($M_{I_1}, M_{I_2}$) capture the case where a gene is rhythmic in a single tissue (tissue-specific) and a shared model ($M_{I_3}$) captures the case where a gene is rhythmic in both tissues with the same rhythmic parameters

$$M_{I_1} : \quad Y_{t,c_i} = \begin{cases} \alpha_{c_i} + \beta_{1,c_i} \cos(\omega t) + \beta_{2,c_i} \sin(\omega t) + \epsilon & i = 1 \\ \alpha_{c_i} + \epsilon & i = 2 \end{cases}$$

$$M_{I_2} : \quad Y_{t,c_i} = \begin{cases} \alpha_{c_i} + \epsilon & i = 1 \\ \alpha_{c_i} + \beta_{1,c_i} \cos(\omega t) + \beta_{2,c_i} \sin(\omega t) + \epsilon & i = 2 \end{cases}$$

$$M_{I_3} : \quad Y_{t,c_i} = \alpha_{c_i} + \beta_1 \cos(\omega t) + \beta_2 \sin(\omega t) + \epsilon. \quad (2.3)$$

The shared model $M_{I_3}$ is notable because it is distinct from the full model $M_f$. $M_{I_3}$ forces both tissues to have the same amplitude and phase, whereas $M_f$ may have tissue-dependent rhythms. From the point of view of gene regulation, $M_{I_3}$ (shared model, same rhythm) has a simpler explanation than $M_f$ (full model, different rhythms), and we therefore explicitly distinguish these two cases.
2.2. Extended methods: model selection with Zellner’s \( g \)−priors

Increasing model complexity, larger penalization

\[
M_r
\]

\[
M_{I_{1,2,3}}
\]

\[
M_f
\]

Figure 2.1 – Illustration of number of models for the case of two tissues. For two tissues, there are five rhythmic models. Importantly, we allow the tissues to share rhythms, which reduces the number of parameters in the model if the rhythm is synchronized across tissues.

2.2.2 Selecting models by the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and Bayes Factors

After having enumerated the possible rhythmic models, the final step is to fit each model and identify which is the best model, based on the fit of the data as well as the complexity of the model. Typically, we seek to find a model \( M_\gamma \) that minimizes a penalized negative log-likelihood score

\[
l_n(\hat{L}_\gamma) = -2 \ln(\hat{L}_\gamma) + p_\gamma F
\]

where \( p_\gamma \) is the number of parameters for model \( \gamma \). \( F = 2 \) for AIC and \( F = \ln(n) \) for BIC. \( \ln(\hat{L}_\gamma) \) is the maximized value of the log likelihood. That is, \( \hat{L} = p(Y|\hat{\theta}, M_\gamma) \) where \( \hat{\theta} \) are the parameter values that maximize the likelihood function.

Although AIC and BIC provide straightforward approaches to selecting the best models, they both have their drawbacks. AIC can be shown to be an inconsistent estimator (i.e. probability of selecting true model does not go to 1 as sample size goes to infinity) (Yang, 2005). BIC can be derived as a Laplace approximation (obtained by finding the mode of the posterior distribution and then fitting a Gaussian centered at the mode) to Bayes factors, which may not always be appropriate, especially for small number of samples (Berger et al., 2003; George and Foster, 2000).

An alternative approach is to evaluate the posterior probabilities of the models, which can be expressed through the Bayes factor \( K \) between a candidate model \( \gamma \) and a base model \( b \). In linear regression models, this Bayes factor \( K \) form has a convenient closed-form solution in
Chapter 2. Transcription factor activity rhythms and tissue-specific chromatin interactions explain circadian gene expression across organs

the g-prior method (Equation 2.11).

$$p(M_f | Y) = \frac{p(M_f)p(Y | M_f)}{\sum_{M_f'} p(M_f')p(Y | M_f')} = \frac{p(M_f)K(M_f; M_b)}{\sum_{M_f'} p(M_f')K(M_f'; M_b)}$$ (2.5)

where the base model $M_b$ contains only the intercepts, $\alpha$. A key component from Equation 2.5 is the marginal likelihood of the data given the model:

$$p(Y | M_f) = \int_{\Theta} p(Y | \hat{\theta}_f, M_f) p(\hat{\theta}_f | M_f) d\hat{\theta}_f,$$ (2.6)

which requires integrating over the parameters of the model and setting a prior on each of the parameters. For parameters that are common to all models, we may set an improper prior, such as Jeffrey’s prior, which is defined up to an arbitrary multiplicative constant. For all other parameters, it is advised to avoid improper priors because the arbitrary constant does not factor out in the posterior calculation.

In linear regression models, the model parameters $\hat{\theta}$ can be written as intercept and regression coefficients, $\alpha$ and $\hat{\beta}$, respectively. The data $Y$ with $n$ data points can be modeled in a linear regression model with Gaussian noise $\epsilon \sim N(0, \sigma^2)$:

$$Y = \Gamma_n \alpha + X_\gamma \beta_\gamma + \epsilon$$ (2.7)

A computationally efficient prior for linear regression models is the "g-prior" (Zellner, 1986), defined as

$$p(\alpha, \sigma^2 | M_f) = \frac{1}{\sigma^2}, \quad \beta_\gamma | \sigma \sim N \left(0, g \sigma^2 \{X^T X^{-1}\} \right).$$ (2.8)

The scalar $g$ controls the spread of the prior in the parameters. Larger $g$ tends to favor parsimonious models (i.e. few large parameters), while smaller $g$ tends to favor saturated models. Note that this prior is related to more standard Bayesian linear regression with Gaussian distribution for coefficients and inverse gamma distribution for the variance:

$$p(\hat{\beta}, \sigma^2) \sim N \left(\hat{\beta}_0, \sigma^2 V_0 \right) IG \left(\sigma^2 | a_0, b_0 \right).$$ (2.9)
2.2. Extended methods: model selection with Zellner’s $g$–priors

and setting $a_0 = b_0 = 0$, corresponding to an uninformative prior for $\sigma^2$ and to set $\nu_0 = 0$ and $V_0 = g(X^TX)^{-1}$. This reduces the problem to a single hyperparameter, $g$, and has straightforward closed-form solutions for calculating the Bayes factor, which is helpful for model selection (Liang et al., 2008).

There are many strategies for specifying $g$ as extensively discussed by Liang et al. (2008). We will select a specific value for $g$ based on how we expect the data should behave based on assumptions from circadian biology.

The marginal likelihood is then given in closed form:

$$p(\bar{Y}|M_\gamma, g) = \frac{\Gamma((n-1)/2)}{\sqrt{\pi(n-1)/2}} \frac{\parallel\bar{Y} - \hat{Y}\parallel^{-(n-1)}}{(1+g)^{\frac{n-1}{2}}}.$$  \hspace{1cm} (2.10)

The Bayes factor $K(M_\gamma : M_b)$ of model $M_\gamma$ with respect to a base model $M_b$ can expressed as the ratio of their marginal likelihoods:

$$K(M_\gamma : M_b)|g = \frac{(1+g)^{\frac{n-1}{2}}}{[1+g(1-R^2_\gamma)]^{\frac{n-1}{2}}}.$$  \hspace{1cm} (2.11)

where $R^2_\gamma = 1 - \sum_i (y_i - \bar{y})^2 / \sum_i (y_i - \hat{y})^2$ is the coefficient of determination for model $\gamma$, a goodness of fit measurement that can easily be calculated. $n$ is the number of data points.

In this chapter, we apply Equation 2.10 directly by selecting $g$, which controls the penalty on model complexity. To guide our choice of $g$, we make an assumption pertinent to circadian biology: slight differences in the amplitude and phase of core clock genes between tissues should still be considered to yield shared rhythms. To achieve this, we choose $g$ with sufficient tolerance to amplitude and phase differences (which may depend on the experiment and technology) such that most of the condition-wide oscillations are incorporated into the shared model (rather than the full model). Plotting the 24h spectral power of genes in the shared model as a function of $g$ allows us to identify the beginnings of a plateau. We interpret this point to be the $g$ value whereby most core clock genes are included in the shared model (Figure 2.2). In our liver and kidney RNA-seq dataset, we found that $g = 1000$ provided biologically interpretable outputs.
Figure 2.2 – Spectral variance as a function of $g$. For each $g$, we calculated the 24-hour variance by taking the sum of the 24-hour variance across genes assigned to model that shares a rhythm across all conditions. Note that near $g = 1000$, the variance begins to plateau, suggesting the additional genes that are included $g > 1000$ may likely be noise rather than robust amplitudes.

2.3 Extended methods: complex-valued singular value decomposition

Here, we seek to project temporal gene expression across tissues onto lower projections in order to visualize and summarize high-dimensional data across tissues and time. Circadian datasets across tissues often contain correlation structures that can be represented in lower dimensions. For example, many core circadian clock and clock outputs genes oscillate synchronously in nearly all tissues. This projection is useful as a genome-wide exploration technique (analogous to principal component analysis) or as a downstream analysis of clusters of genes identified using model selection. We found that complex-valued singular value decomposition (cv-SVD) to have useful properties that were biologically interpretable. Chronobiology datasets that sample multiple tissues or conditions over the 24-hour day can be represented by a matrix of complex values where the values represent amplitude (we define it here as min-to-max magnitude rather than mean-to-peak to simplify biological interpretation) and
phase (time at peak expression) for each gene and each condition. This complex number can be calculated by the 24-hour Fourier component corresponding to an angular frequency of \( \omega = \frac{2\pi}{24} \) (we checked that the 24-h represents the largest temporal component in Yeung et al. by breaking down the temporal variation into all Fourier components, which showed that the largest temporal component comes from 24-hour rhythms, followed by 12-hours). This approximation works well for genes that are can be modeled by cosine and sines such as in models \( M_f, M_{I1}, M_{I2}, M_{I3} \) in Equation 2.1 and 2.3.

Since we expect the temporal variance to be well approximated by the 24-hour Fourier space \( \left( \sum_{t} y_t(t) - \bar{y}(t) \right)^2 \approx |Y_{\omega_{24}}|^2 \), we project

\[
Y_{g,c,\omega_{24}} = \sum_{t} Y_{g,c,t}(t) e^{i\omega_{24} t},
\]

where \( \omega_{24} = \frac{2\pi}{24} \). This projection works well on genes that have temporal signals that can be modeled well by a sine wave of period 24 hours. This approximates the dataset to a complex-valued matrix of \( G \) rows and \( C \) columns. Each complex-valued element \( Y_{g,c,\omega_{24}} \) corresponds to the amplitude and phase of gene \( g \) in condition \( c \). Because many rhythms may be shared across conditions, we can decompose \( Y_{g,c,\omega_{24}} \) by standard singular value decomposition (SVD) into sample space (eigentissues) and tissue space (eigensamples), which are also complex values, and the first few singular values should provide a low-dimensional representation of the data. We can then visualize and summarize the eigentissues and eigensamples in the 24-hour Fourier space by noting that \( |Y_{g,c,\omega_{24}}| \) corresponds to the amplitude and \( \text{Arg}(Y_{g,c,\omega_{24}}) \) corresponds to the phase of gene \( g \) and tissue \( c \). The amplitude and phase can be naturally visualized in polar coordinates; the radial distance representing amplitude and the phase angle representing time at peak expression.

The SVD technique factorizes \( Y_{g,c,\omega_{24}} = U \Sigma V^* \), where \( V^* \) is the complex conjugate of \( V \). \( U \) and \( V \) are unique eigensamples and eigengenes, and are defined up to a unit-phase factor \( e^{i\phi} \). To simplify interpretation, we set the phase of the tissue space such that the tissue with the largest amplitude has phase 0 and amplitude 1. Correspondingly, the gene space would be interpreted as the amplitude and phase of the largest amplitude tissue. To get the amplitude and phase of genes in another tissue, one would simplify add and multiply the phase and amplitude, respectively, of the loadings for that tissue with the gene space.

### 2.4 Extended methods: Gene ontology (GO) analysis around the clock

From our model selection process, we identify a set of genes (a module) that have rhythms in a specific combination of tissues (e.g., genes that oscillate in liver but not in kidney). This module have gene expression rhythms that peak at different times of day. One useful downstream analysis is to ask whether certain Gene Ontology (GO) terms are enriched at different times
of day. To extend GO term analysis to the 24-hour day, I used foreground genes as genes in the module that are within a time window \([t - 3, t + 3]\) and expressed genes in the dataset as background genes. For each \(t \in \{1, 2, 3, ..., 24\}\) I calculated a p-value for enrichment using classic Fisher’s exact test, as implemented by TopGO package in R. This outputs p-value enrichment over time, which can then be visualized in a polar plot.

Example code can be found on GitHub: https://github.com/naef-lab/CyclicGO.

### 2.5 Contributions

I did all the computational methods and analyses in this project. Jérome Mermet did all the 4C-seq experiments. RNA-seq experiments were performed by Julien Marquis, Aline Charpagne, and Céline Jouffe.
Transcription factor activity rhythms and tissue-specific chromatin interactions explain circadian gene expression across organs

Jake Yeung,1,5 Jérôme Mermet,1,5 Céline Jouffe,2 Julien Marquis,3 Aline Charpagne,3 Frédéric Gachon,2,4 and Felix Naef1

1Institute of Bioengineering, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, CH-1015, Switzerland; 2Department of Diabetes and Circadian Rhythms, Nestlé Institute of Health Sciences, CH-1015 Lausanne, Switzerland; 3Functional Genomics, Nestlé Institute of Health Sciences, CH-1015 Lausanne, Switzerland; 4Faculty of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

Temporal control of physiology requires the interplay between gene networks involved in daily timekeeping and tissue function across different organs. How the circadian clock interweaves with tissue-specific transcriptional programs is poorly understood. Here, we dissected temporal and tissue-specific regulation at multiple gene regulatory layers by examining mouse tissues with an intact or disrupted clock over time. Integrated analysis uncovered two distinct regulatory modes underlying tissue-specific rhythms: tissue-specific oscillations in transcription factor (TF) activity, which were linked to feeding-fasting cycles in liver and sodium homeostasis in kidney; and colocalized binding of clock and tissue-specific transcription factors at distal enhancers. Chromosome conformation capture (4C-seq) in liver and kidney identified liver-specific chromatin loops that recruited clock-bound enhancers to promoters to regulate liver-specific transcriptional rhythms. Furthermore, this looping was remarkably promoter-specific on the scale of less than 10 kilobases (kb). Enhancers can contact a rhythmic promoter while looping out nearby nonrhythmic alternative promoters, confining rhythmic enhancer activity to specific promoters. These findings suggest that chromatin folding enables the clock to regulate rhythmic transcription of specific promoters to output temporal transcriptional programs tailored to different tissues.

[Supplemental material is available for this article.]
2015), we analyzed these data sets separately. We performed principal component analysis (PCA) on the entire set of conditions (11 tissues × 24 time points) to obtain an unbiased overview into the contributions of tissue- and time-specific variance. Most of the variance concerned differences in expression between tissues (Fig. 1A; Supplemental Fig. S1A–D). Temporal variance, in particular, 24-h periodicity, was present among a group of principal components carrying lower amounts of variance (Fig. 1A; Supplemental Fig. S1E–G). Focusing on genome-wide temporal variation within each tissue, we found that 24-h rhythms constituted the largest contribution of temporal variance, followed by 12-h rhythms, which were close to background levels for many tissues (Fig. 1B; Hughes et al. 2009). We thus focused the rest of our analysis on 24-h rhythms.

We analyzed the peak-to-trough amplitudes (hereafter also referred to as fold change) of 24-h rhythmic transcripts. Metabolic tissues, notably liver, brown fat, and skeletal muscle exhibited more (on the order of 100 transcripts) intermediate- to high-amplitude (between two- and 10-fold) transcript rhythms. Brain tissues showed virtually no rhythmic transcripts above fourfold (Fig. 1C). In liver and kidney of Bmal1 KO mice, the number of rhythmic mRNAs was reduced by threefold compared to WT littermates. This effect increased for larger amplitudes. Few transcripts in tissues of Bmal1 KO oscillated by more than 10-fold (Fig. 1D). Thus, a functional circadian clock is required for high-amplitude transcript rhythms across diverse tissues, while systemic signals regulate lower amplitude rhythms that persist in clock-deficient liver (Hughes et al. 2012; Atger et al. 2015; Sobel et al. 2017) and kidney (Nikolaeva et al. 2012).

Combinatorics of rhythmic transcript expression across tissues and genotypes

We reasoned that identifying sets of genes with shared rhythms across subsets of tissues would allow finding underlying regulatory mechanisms. We developed a model selection (MS) algorithm extending harmonic regression (Fisher 1929) to classify genes into modules sharing rhythmic mRNA profiles across subsets of tissues (Fig. 2A; Methods). Phase (time of peak) and amplitude (log2 fold change) relationships between genes and tissues are summarized using complex-valued singular value decomposition (SVD) (Fig. 2B; Methods). We applied MS to the 11 tissues, which identified gene modules involving rhythmic mRNA accumulation in nearly all tissues (tissue-wide) (Fig. 2C), in single tissues (tissue-specific), or in several tissues (tissue-restricted) (examples shown in Fig. 2D; Supplemental Fig. S2A; Supplemental Table S3).

The tissue-wide module contained a set of both clock- and system-driven rhythmic mRNAs, as determined by comparing Bmal1 KO data in liver and kidney (Fig. 2C, left). Moreover, these transcripts oscillated in synchrony across all tissues and peaked at fixed times of day,
Figure 2. Combinatorics of rhythmic transcript expression across tissues and genotypes. (A) Schema for the model selection (MS) algorithm to identify rhythmic gene expression modules across tissues. Temporal transcriptomes of different tissues represented as a three-dimensional array (left). Gene modules are probabilistically assigned among different combinations of 24-h rhythms across tissues (e.g., tissue-specific or tissue-wide rhythms schematically shown on right). (B) Gene modules are summarized by the first component of complex-valued singular value decomposition (SVD) to highlight phase (peak time shown as the clockwise angle) and amplitude (log2 fold change shown as the radial distance) relationships between genes (gene space) and between tissues (tissue space). SVD representation is scaled such that the genes show log2 fold changes, while tissue vectors are scaled such that the highest amplitude tissue has length of 1 and a phase offset of 0 h. (C–E) MS applied to 11 WT tissues. (C) SVD representation of tissue-wide mRNA rhythms from the 11 tissues. Genes are labeled as system-driven (blue) or clock-driven (red) according to the comparison of the corresponding temporal profiles in Bmal1 KO and WT littermates. (D) Examples of anti-phasic rhythms (brown fat and muscle, n = 20, first SVD component explains 81% of variance), and tissue-specific rhythms (liver, n = 846, first SVD component explains 59% of variance). Representative genes with large amplitudes are labeled. (E) Number of transcripts showing rhythms (P-value < 0.01, harmonic regression) in different numbers of tissues, in function of increasing peak to trough amplitudes on the x-axis. -axis: average log2 fold change calculated from the identified rhythmic tissues. (F) MS applied to Bmal1 KO and WT littermates in liver and kidney. (F) SVD representation of clock- (top, n = 991, 83% of variance) and system-driven (bottom, n = 1395, 84% of variance) liver-specific rhythms. (G) Number of transcripts showing clock- (solid) or system-driven (dotted) rhythms (P-value < 0.01, harmonic regression) in liver (red), kidney (blue), or both (magenta).
although their amplitudes varied between tissues, with brain regions showing the smallest amplitudes (Fig. 2C, right). The clock drove synchronized oscillations at high amplitudes, notably clock genes (e.g., \textit{Arntl}, \textit{Npas2}, \textit{Nr1d1}, \textit{2}; note that \textit{Arntl} and \textit{Nr1d1}, \textit{2} are also named \textit{Bmal1} and \textit{Rev-erba}, \textit{b}), respectively, clock output genes (e.g., \textit{Dbp}, \textit{Nfil3}), and cell cycle regulators (\textit{Cdkn1a} and \textit{Wee1}) (Matsuo et al. 2003; Gréèchez-Cassiau et al. 2008). Interestingly, clock genes \textit{Per1,2} continued to oscillate in \textit{Bmal1} KO in multiple tissues, extending previous studies in liver (Kornmann et al. 2007). Other clock-independent oscillations included mRNAs of heat- and cold-induced genes, such as \textit{Hspa8} and \textit{Cbp} (Mort et al. 2012; Gotic et al. 2016), that peaked 12 h apart near CT18 and CT6 (CT: circadian time; CT0 corresponds to subjective dawn and start of the resting phase; CT12 corresponds to subjective dusk and start of the activity phase), concomitantly with highs and lows in body temperature rhythms (Refinetti and Menaker 1992).

Tissue-restricted modules contained rhythmic transcripts that peaked in synchrony, such as in liver and kidney, or with fixed offsets, such as the nearly 12-h shifted rhythms in brown fat and skeletal muscle (Supplemental Fig. S3A). Overall, transcripts with large amplitudes (FC: >3) oscillated in either a few tissues (three or less) or tissue-wide (eight or more) (Fig. 2E).

To distinguish clock- and system-driven mRNA rhythms, we applied the MS algorithm to the liver and kidney transcriptomes in WT and \textit{Bmal1} KO mice (Fig. 2F; Supplemental Fig. S3B; Supplemental Table S4). This separation identified clock- and system-driven modules that oscillated in liver but were flat in kidney (Fig. 2F), as exemplified by mRNAs of \textit{Lig2} and \textit{Lpin1} (Supplemental Fig. S2B). Indeed, both transcripts oscillated in WT liver with robust amplitudes, peaking near CT11, but were flat in kidney (ZT: Zeitgeber time; ZT0 corresponds to onset of lights-off; ZT12 corresponds to onset of lights-on). However, in \textit{Bmal1} KO, \textit{Lpin1} continued to oscillate, while \textit{Lig2} was flat. Summarizing, we found that shared clock-driven mRNA rhythms, which contained core clock and clock-controlled genes, oscillated with significantly larger amplitudes than system-driven genes (Fig. 2G, magenta solid versus dotted). Similarly, clock-driv-en liver-specific mRNA rhythms also oscillated at higher amplitudes compared with system-driven mRNA rhythms (Fig. 2G, red solid versus dotted). On the other hand, kidney-specific clock- and system-driven transcripts oscillated with comparable amplitudes (Fig. 2G, blue solid versus dotted) and were less numerous overall, which could reflect the distinct cell types constituting the kidney (Lee et al. 2015). The uncovered diversity of clock- and system-driven mRNA rhythms involving distinct combinations of tissues hints at complex transcriptional or post-transcriptional regulation.

\textbf{Oscillatory TF activity in one tissue but not others can drive tissue-specific mRNA rhythms}

We focused on WT and \textit{Bmal1} KO liver and kidney to identify rhythmic TF activities underlying clock- and system-driven tissue-specific mRNA rhythms. We first analyzed liver-rhythmic genes driven by systemic signals (n = 1395, MS) (Fig. 3A), which were associated with feeding and fasting rhythms (Gene Ontology analysis around the clock) (Methods). Indeed, ribosome biogenesis was up-regulated most strongly during the first 6 h of the feeding phase (from ZT12 to ZT18) (Jouffe et al. 2013; Chauvin et al. 2014), while insulin signaling was down-regulated during first 6 h of the fasting phase (from ZT0 to ZT6) (Ravnskjæer et al. 2013), consistent with daily responses to nutrient fluctuations in liver (Sinturel et al. 2017).

To infer rhythmic TF activities that may underlie these mRNA rhythms, we applied a penalized regression model (MARA) (Balwietz et al. 2014) that integrates TF binding site predictions near promoters with mRNA accumulation. This approach identified ROREs in liver-specific mRNA rhythms and our previously described temporal nuclear protein data set (Wang et al. 2017), that fluctuated in rhythmic activity of MBF and EGR1 and were flat in kidney, consistent with loss of REV-ERB expression and our previously described temporal nuclear protein data set (Wang et al. 2017), that fluctuated in rhythmic activity of MBF and EGR1 and were flat in kidney, consistent with loss of REV-ERB expression.

Next, we analyzed clock-driven transcripts oscillating specifically in the kidney (n = 156, MS) (Fig. 3C), among which sodium ion and organic anion transporters peaked near ZT12 and ZT0, respectively. The up-regulation of sodium ion transporters in kidney during the behaviorally active phase may underlie clock-depend-ent increase of sodium excretion (Nikolaeva et al. 2012). Similarly, the up-regulation of organic anion transporters during the resting phase may explain increased transport activity for precursors of gluconeogenesis, such as pyruvate and lactate, during fasting (Stumvoll et al. 1998; Ekberg et al. 1999). TF analysis predicted TFCP2 to regulate mRNAs that peaked during the resting phase (Fig. 3D; Supplemental Fig. S4C). The predicted TFCP2 activity was anti-phasic with \textit{Tfcp2} mRNA abundance, suggesting a repressive activity, consistent with the ability of TFCP2 to recruit histone deacetylase HDAC1 (Kim et al. 2016).

Finally, liver-specific clock-driven rhythmic transcripts (n = 991, MS) were comprised of genes associated with glucose metabolism (enriched at ZT18), such as \textit{Gck} and \textit{Ppp1r3b} (Kelsall et al. 2009; Oosterveer and Schoonjans 2014), as well as lip, cholesterol, and bile acid metabolism genes (enriched at ZT2), such as \textit{Elolv3}, \textit{Insig2}, \textit{Hsd3b7}, and \textit{Cyp8b1} (Fig. 3E; Shea et al. 2007; Le Martelot et al. 2009; Guillou et al. 2010; Sayin et al. 2013). Predicted activity of ELF1 oscillated and peaked near ZT3 in WT liver but was flat in \textit{Bmal1} KO (Fig. 3F; Supplemental Fig. S4F; Fang et al. 2014). Interestingly, mRNA abundance of ELF1, as well as its nucle-ar protein abundance, also oscillated in WT, supporting ELF1 as a potential regulator of oscillating transcripts peaking near ZT6.

\textbf{Colocalized binding of clock- and liver-specific TFs drives liver-specific mRNA rhythms}

To further dissect liver-specific clock-driven rhythms, we reasoned that accessible chromatin regions specific to the liver could harbor regulatory sites for clock TFs, which could then regulate mRNA rhythms liver-specifically. Comparing DNase I hypersensitive sites (DHSs) in liver and kidney (DNase-seq data from ENCODE) (Yue et al. 2014), we found that liver-specific clock-driven genes were enriched with liver-specific DHSs (within 40 kb of promoters), compared to system-driven as well as nonrhythmic genes (Fig. 4A). Using TF binding site predictions underlying these liver-specific DHSs, we applied MARA to predict rhythmic TF activities that explain gene expression of this module (Supplemental Fig. S5A). In WT liver, the predicted activity of RORE oscillated with robust amplitudes and peaked near ZT21. RORE activity became high and flat in \textit{Bmal1} KO liver, consistent with loss of REV-ERB expression and...
consequently de-repression of REV-ERB target genes (Fig. 4B, top; Bugge et al. 2012). Activity of E-box in WT liver peaked at ZT7, consequently de-repression of REV-ERB target genes (Fig. 4B, top; Bugge et al. 2012). Activity of E-box in WT liver peaked at ZT7, consistent with BMAL1:CLOCK activity (Rey et al. 2011), albeit with weaker amplitudes compared to RORE activity, likely reflecting consistent with BMAL1:CLOCK activity (Rey et al. 2011), albeit with weaker amplitudes compared to RORE activity, likely reflecting fewer E-box target genes compared to RORE in this module. In Bmal1 KO mice, E-box activity was low and flat in liver, as expected. We hypothesized that cooperativity of liver-specific and clock TFs at liver-specific DHSs can regulate liver-specific mRNA rhythms. Pairwise analysis of TF binding sites at liver-specific DHSs found enrichment of co-occurrence between RORE and liver-specific TF motifs, FOXA2, ONECUT, and CUX2 (Fig. 4C). Enrichment of both CUX2 and ONECUT (also named HNF6) is consistent with ONECUT binding to both ONECUT and CUX2 motifs (Conforto et al. 2015). miRNAs of genes with co-occurrence of RORE and liver-specific TF motifs peaked near ZT1, consistent with peak RORL activity (near ZT21) preceding peak mRNA abundance of REV-ERB targets (Supplemental Fig. S5B). Analysis of ChIP-exo data sets targeting FOXA2, ONECUT1, and REV-ERBa in liver (Wang et al. 2014; Zhang et al. 2015; Iwafuchi-Doi et al. 2016) confirmed colocalization of TF binding at liver-specific DHSs distal from clock-driven liver miRNAs such as Insm2 and Slc44a4 (Fig. 4D).

Liver-specific chromatin loops regulate liver-specific mRNA rhythms

To test whether distally located liver-specific DHSs can contact promoters of clock-driven liver-rhythmic genes, we selected the promoters of Mrg, Pik3ap1, and Slc44a1 as baits for 4C-seq experiments in liver and kidney harvested at the time of peak mRNA accumulation for the selected genes (Methods; Fig. S5A; Supplemental Figs. S6A, S7A). Upstream of Mrg, the 4C-seq signal, which measures frequency of promoter-enhancer contacts (van de Werken et al. 2012), decayed rapidly to background level in both liver and kidney (Fig. S5B, top). Downstream from Mrg, however, the 4C-seq signal showed a tissue-dependent pattern, decaying slowly in the liver but more rapidly in the kidney. This difference in decay suggests increased frequency of promoter-enhancer contacts in the liver compared to the kidney. Indeed, differential analysis identified liver-specific chromatin contacts 40 kb downstream from the promoter (Fig. S5B, bottom). Overlaying the contact data with DNase-seq, we found that liver-specific chromatin contacts downstream from Mrg connected liver-specific DHSs with the Mrg promoter (Fig. S5C). Furthermore, ChIP-exo showed colocalization of REV-ERBa and FOXA2 binding at liver-specific DHSs contacting the promoters (Fig. S5C). In contrast, accessible regions upstream of the Mrg promoter did not show liver-specific chromatin contacts. The 4C-seq data thus suggest that liver-specific chromatin loops can recruit clock-bound distal elements to promoters to regulate liver-specific transcriptional rhythms. Other liver-specific rhythmic transcripts, Pik3ap1 and Slc44a1, also displayed liver-specific chromatin loops between promoter and liver-specific open chromatin regions (Supplemental Figs. S6, S7). In sum, tissue-specific chromatin looping can drive tissue-specific mRNA rhythms.
Figure 4. Colocalized binding of clock- and liver-specific TFs underlies liver-specific mRNA rhythms. (A) The fraction of genes containing liver-specific DNase I hypersensitive sites (DHSs) in the clock-driven liver-specific module is higher compared with both nonrhythmic and system-driven liver-specific modules. Error bars and *P*-values calculated from 10,000 bootstrap iterations. (B) Predicted temporal activities of RORE (top) and E-box (bottom) TF motifs located within liver-specific DHSs. Error bars show standard deviation of the estimated activities. (C) Co-occurrence of RORE with all other TFs in the SwissRegulon database (Pachkov et al. 2007) (189 TF motifs). Positive log10 odds ratios (ORs) represent pairs of motifs enriched in the clock-driven liver-specific module compared to the flat module. *P*-values for the motif pairs were calculated from χ² tests applied to three-way contingency tables (Myšicková et al. 2012). Selected pairs are in bold. (D) DNase I hypersensitivity in liver, kidney, and the corresponding differential signal (in log2 fold change) near two representative genes (top: Insig2; bottom: Slc4a4). RORE, ONECUT1, and FOXA2 TF binding motifs (posterior probability > 0.5, MotEvo) co-occur at liver-specific DHSs (red boxes). Predicted TF binding sites correspond to experimentally observed TF binding in publicly available ChIP-exo data sets for REV-ERBa, ONECUT1, and FOXA2 (bottom).
Precise promoter-enhancer contacts underlie liver-specific mRNA rhythms

To test whether distinct chromatin loops would form at alternative nearby gene promoters with distinct temporal mRNA profiles, we searched for candidate genes where one promoter was rhythmically transcribed while the alternative one was nonrhythmic (Supplemental Fig. S8). Slc45a3 has two alternative transcripts using different promoters 8 kb apart. The shorter pre-mRNA oscillated in the liver (rhythmic promoter, Slc45a3-short), while the longer did not (flat promoter, Slc45a3-long). In kidney, neither Slc45a3-short nor Slc45a3-long showed robust transcript rhythms (Supplemental Fig. S9). Targeting the Slc45a3-short promoter with 4C-seq in liver and kidney showed liver-specific chromatin interactions that differed in liver and kidney (Fig. 5A). In liver, the 4C-seq short promoter showed 4C-contact signals specific to liver (Fig. 5B). In kidney, neither Slc45a3-short nor Slc45a3-long showed robust transcript rhythms (Supplemental Fig. S9). Targeting the Slc45a3-long promoter with 4C-seq in liver and kidney showed liver-specific chromatin loops that differed in liver and kidney (Fig. 6A). These same regions did not form liver-specific chromatin loops in kidney, suggesting that promoters 8 kb apart can contact distinct enhancers.

Figure 5. Liver-specific chromatin loops regulate liver-specific mRNA rhythms. (A) Temporal mRNA profile for Mreg, a clock-driven liver-rhythmic gene. Error bars are SEM (n = 2). (B) 4C-seq profiles (summarized by two replicates, each pooling two different mice) using the Mreg promoter as a bait in liver and kidney at ZT20. Data are shown in a window of ±250 kb from the bait (top). Profiles of differential contacts between liver and kidney (bottom) represented as signed log P-values (regularized t-test, positive values denote liver-enriched 4C contacts). (C) Tracks of differential 4C contacts (signed log P-values), log, fold change of DNase I hypersensitivity between liver and kidney, and ChIP-exo of REV-ERBa and FOXA2. Regions of significant differential 4C contacts correspond to liver-specific DNase I hypersensitive regions and REV-ERBa binding sites.

Precise promoter-enhancer contacts underlie liver-specific mRNA rhythms

To test whether distinct chromatin loops would form at alternative nearby gene promoters with distinct temporal mRNA profiles, we searched for candidate genes where one promoter was rhythmically transcribed while the alternative one was nonrhythmic (Supplemental Fig. S8). Slc45a3 has two alternative transcripts using different promoters 8 kb apart. The shorter pre-mRNA oscillated in the liver (rhythmic promoter, Slc45a3-short), while the longer did not (flat promoter, Slc45a3-long). In kidney, neither Slc45a3-short nor Slc45a3-long showed robust transcript rhythms (Supplemental Fig. S9). Targeting the Slc45a3-short promoter with 4C-seq in liver and kidney showed liver-specific chromatin loops that differed in liver and kidney (Fig. 5A). In liver, the 4C-seq short promoter showed 4C-contact signals specific to liver (Fig. 5B). In kidney, neither Slc45a3-short nor Slc45a3-long showed robust transcript rhythms (Supplemental Fig. S9). Targeting the Slc45a3-long promoter with 4C-seq in liver and kidney showed liver-specific chromatin loops that differed in liver and kidney (Fig. 6A). These same regions did not form liver-specific chromatin loops in kidney, suggesting that promoters 8 kb apart can contact distinct enhancers. Overlapping 4C-seq with DNase-seq showed that these chromatin loops link liver-specific DHSs specifically to the Slc45a3-short promoter (Fig. 6C). These liver-specific DHSs were bound by liver-specific TFs, FOXA2 and ONECUT1, and clock TF, REV-ERBa, as shown in ChIP-seq. The 4C experiments suggest that enhancers can contact a rhythmic promoter while looping out nearby nonrhythmic alternative promoters, confining rhythmic enhancer activity to specific promoters (Fig. 6D). Furthermore, rhythmically active enhancers can contact promoters in a tissue-specific manner. Thus, chromatin folding not only regulates tissue-specific rhythms but also differentiates between closely spaced promoters to control rhythmic transcription with spatial precision.

Discussion

The mammalian genome encodes transcriptional programs that allow the molecular clock to robustly oscillate across diverse tissue transcriptomes while maintaining flexibility to regulate distinct clock outputs in different combinations of tissues. Here, we identified two regulatory modes underlying tissue-specific transcript rhythms: Regulatory sequences can recruit individual TFs bearing rhythmic activity; coordinated binding of clock- and tissue-specific TFs can generate tissue-specific rhythms. Moreover, we found that clock- and tissue-specific TFs bound at distal enhancers can be recruited to promoters through precise chromatin loops.

Several of our predictions of transcription regulators and regulated genes (e.g., Egr1, Por, Upp2) corroborated with previous analyses of independent data sets (Yan et al. 2008; Bozek et al. 2009; Bhargava et al. 2015). Furthermore, functional significance of colocalization could be gained, for example, by using inducible knockout models for tissue-specific regulators. Moreover, the observed colocalization does not exclude other cooperative modes, such as tethering of REV-ERBa to ONECUT1 through protein-protein interactions (Zhang et al. 2015).

Our 4C analysis showed that chromatin looping might mediate interaction between clock- and tissue-specific transcriptional programs by recruiting clock-bound distal elements to promoters in a tissue-specific manner. Such loops can surgically discriminate between nearby promoters as close as 8 kb apart, suggesting a way to separate temporal regulation of neighboring promoters. A previous 4C study on a core clock gene enhancer proposed that colocalization was involved in promoter-enhancer looping and compartmentalize rhythmic gene expression within genomic regions spanning 150 kb (Xu et al. 2016). Here, chromatin interactions that differed between tissues were localized to a small genomic region (<10 kb) near promoters (<100 kb). Future studies integrating temporal data across tissues with large-scale promoter-enhancer networks may reveal regulatory sequences that encode promoter-enhancer compatibility and elucidate whether this compatibility is tissue-specific (Li and Noll 1994; Merli et al. 1996; Zahidi et al. 2014; Nguyen et al. 2016).

While our work focused on transcriptional mechanisms, studying other mechanisms such as post-transcriptional,
translational, and post-translational processes using PRO-seq, Ribo-seq, and proteomics data may provide additional insights. Expanding our 24-h analysis to 12-h or other harmonics would broaden the view of tissue-specific temporal gene expression but may require experimental designs of higher temporal resolution (Hughes et al. 2009; Krishnaiah et al. 2017). In sum, integrating the temporal axis into tissue-specific gene regulation offers an integrated understanding of how tissue physiology resonates with daily cycles in the environment.

Methods

Animal experiments

Eight- to 14-wk-old C57Bl/6 mice have been purchased from Charles River Laboratory. Bmal1 KO mice have been previously described (Jouffe et al. 2013). Without further indications, mice are kept under a 12-h light/12-h dark regimen and ad libitum feeding. All animal care and handling was performed according to the Canton de Vaud (Fred Gachon, authorization no. VD 2720) laws for animal protection.

RNA-seq experiments

To complement the mouse liver WT and Bmal1 KO RNA-seq data (GSE73554) (Atger et al. 2015), transcriptomes of kidneys from Bmal1 KO and WT littermates (12-h light/12-h regimen; night-restricted feeding) were measured from poly(A)-selected mRNA using single-end reads of length 100. mRNA levels were quantified using kallisto version 0.42.4 (mm10) (Bray et al. 2015).

Global temporal variance

For each tissue, we estimated the contribution of temporal variance for each gene, broken down by its Fourier components. We calculated the background level assuming temporally unstructured data (white noise), whose magnitude (strength of the white noise) was estimated from the mean of squared magnitudes of Fourier coefficients that were not submultiples of 24 h (i.e., the mean of 48-, 16-, 9.6-, 6.9-, 5.3-, 4.4-h components).

Model selection

We fitted harmonic regression models that integrated temporal gene expression across different combinations of rhythms in different conditions (Atger et al. 2015). We used a g-prior for the rhythmic parameters $\beta$ as a penalization scheme (Liang et al. 2008). We set $g = 1000$, which we found to maximize temporal variations captured in the shared rhythms model while minimizing temporal variations captured in the flat model.

Figure 6. Precise promoter-enhancer contacts underlie liver-specific mRNA rhythms. (A, B) 4C-seq profiles for the (A) Scl45a3-short and (B) Scl45a3-long isoforms within ±250 kb around baits targeting the two TSSs (top). Signed log P-values for differential contacts between liver and kidney (bottom) as in Figure 5B. TSSs for Scl45a3-short and Scl45a3-long are 8 kb apart. (C) Differential 4C contacts (signed log P-values), log$_2$ fold change of DNase I hypersensitivity between liver and kidney, and ChIP-exo signal of REV-ERBa, FOXA2, and ONECUT1. Regions of significant differential contacts in Scl45a3-short correspond to liver-specific DNase I hypersensitive regions. Yellow arrowheads in A and C show liver-specific distal contacts recruited to the Scl45a3-short TSS. These contacts are absent for Scl45a3-long TSS (B). (D) Schematic model illustrating enhancer-promoter interactions in liver and kidney that may generate liver-specific rhythms. Yellow circles illustrate liver-active enhancers contacting the rhythmic promoter (red arrow) but not the alternative nonrhythmic promoter (gray arrow). In kidney, the enhancer is not accessible, and both promoters are nonrhythmic.
Complex singular value decomposition representation of gene and tissue module

We first transformed the time domain to the frequency domain corresponding to 24-h rhythms. The resulting matrix was decomposed using SVD; the first left-singular and first right-singular values were visualized in separate polar plots (Supplemental Methods).

Functional analysis by GO terms

We used Fisher’s exact test to assess statistical significance of gene enrichment for each GO term. Foreground genes were genes with phases within a 6-h window. Background genes were genes assigned to a model. For each GO term, we slid the 6-h window with a step size of 1 h and calculated the P-value enrichment. GO terms were chosen by visualizing significant GO terms in the tree and choosing GO terms that were comparably deep in the tree.

Chromatin conformation experiments and analysis

C57Bl/6 mice were sacrificed at ZT08 and ZT20 to extract liver and kidneys. Liver and kidney nuclei were prepared as previously described (Rippeger and Schüller 2006) with some minor changes. 4C-seq assays were performed as in Gheldof et al. (2012). Raw read counts for each sample were normalized by library size by the sum of the read counts on the cis-chromosome (excluding 10 fragments around the bait). We used a weighted linear model to fit the log_{10} signal around each fragment f. A Gaussian window was used to smooth the signal. A weighted FDR was used to account for multiple testing when correcting for a given number of fragments.

Data access

Raw and processed RNA-seq and 4C-seq data generated from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE100457.

Competing interest statement

C.J., J. Marquis, A.C., and F.G. are employees of Nestlé Institute of Health Sciences SA.

Acknowledgments

We thank Eric Paquet for critical reading and Saeed Omidi for help and discussions in bioinformatics. This work was supported by Swiss National Science Foundation Grant 31003A-153340, European Research Council Grant ERC-2010-StG-260667, and the Ecole Polytechnique de Lausanne. J.Y. benefits from the Natural Sciences and Engineering Research Council of Canada Postgraduate Studies Doctoral scholarship.

Author contributions: Conceptualization, J.Y., J. Mermet, and F.N.; formal analysis, J.Y. and F.N.; investigation, J. Mermet, C.J., J.Y., and F.G.; supervision, F.N. and F.G.; funding acquisition, F.N. and F.G.

References

Tissue-specific circadian transcription regulation


Received June 23, 2017; accepted in revised form December 11, 2017.
Transcription factor activity rhythms and tissue-specific chromatin interactions explain circadian gene expression across organs

Jake Yeung, Jérôme Mermet, Céline Jouffe, et al.

Genome Res. 2018 28: 182-191 originally published online December 18, 2017
Access the most recent version at doi:10.1101/gr.222430.117

Supplemental Material
http://genome.cshlp.org/content/suppl/2018/01/10/gr.222430.117.DC1

References
This article cites 67 articles, 20 of which can be accessed free at:
http://genome.cshlp.org/content/28/2/182.full.html#ref-list-1

Open Access
Freely available online through the Genome Research Open Access option.

Creative Commons License
This article, published in Genome Research, is available under a Creative Commons License (Attribution 4.0 International), as described at http://creativecommons.org/licenses/by/4.0/.

Email Alerting Service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

To subscribe to Genome Research go to:
http://genome.cshlp.org/subscriptions

© 2018 Yeung et al.; Published by Cold Spring Harbor Laboratory Press
Supplemental Figure S1

A. Principal component analysis showing the fraction of total variance explained by each principal component.

B. Tissue and temporal variance loadings for different tissues over time.

C. PC2 gene loadings for specific genes over time.

D. PC1 gene loadings for specific genes over time.

E. PC1 tissue loadings for different tissues over time.

F. Temporal variance loadings for different periods.

G. Bar graph showing -log10(P-value) for rhythmicity.
Supplemental Figure S1 - Contribution of tissue and temporal variance across 11 tissues

(A) Fraction of total variance explained by principal components 1 to 12. Colors represent the contribution of different temporal periods (Fourier coefficients) in each principal component. Components 1 to 12 show predominantly tissue differences.

(B) Sample loadings of 11 tissues over time in PC1. Contribution of tissue variance to the principal component is calculated by the sum of the squared difference between tissue mean and the global mean. Variance in PC1 consists of mostly tissue variance.

(C) Scatterplot of gene loadings from PC1 and PC2. Genes with large PC1 loadings are mainly brain-specific genes.

(D) mRNA abundance across tissues and time of a brain-specific GABA transporter, Slc6a1.

(E) idem as A, for principal components 13 to 100.

(F) Liver loadings over time for PCs 13 to 21. Contribution of temporal variance to the principal component is calculated by the sum of the squared difference between each time point and the tissue mean.

(G) Analysis of 24h rhythmicity of liver loadings over time (PC1 to PC25, harmonic regression). Inset: Amplitude (radial coordinate) and phase (clockwise angle) of liver loadings in PC1 and PC25.
Supplemental Figure S2

A

**Lipg**

Log mRNA abundance (centered)

| Adr, Aorta, BFAT, Cere, Kidney, Liver, Lung | BS, Heart, Hypo, Mus |

| Aorta, Kidney, Liver, Lung | Adr, BFAT, Heart, Mus | BS, Cere, Hypo |

| Liver | Flat |

| Liver | Kidney |

| Liver | Kidney |

**Dbp**

Log mRNA abundance (centered)

| Adr, Aorta, BFAT, Cere, Kidney, Liver, Lung | BS, Heart, Hypo, Mus |

| Aorta, Kidney, Liver, Lung | Adr, BFAT, Heart, Mus | BS, Cere, Hypo |

| Liver | Kidney |

| Liver | Kidney |

**Cirbp**

Log mRNA abundance (centered)

| Adr, Aorta, BFAT, Cere, Kidney, Liver, Lung | BS, Heart, Hypo, Mus |

| Aorta, Kidney, Liver, Lung | Adr, BFAT, Heart, Mus | BS, Cere, Hypo |

| Liver | Kidney |

| Liver | Kidney |

**ZT**

WT  Bmal1 KO
Supplemental Figure S2 – Temporal profiles of mRNA accumulation across tissues and in clock-disrupted conditions

(A) Examples of genes whose expression is rhythmic in different combinations of tissues. Flat indicates nonrhythmic tissues. Rhythms can be further classified as clock- or system-driven by analysis in WT and Bmal1 KO data, shown in (B).

(B) Examples of tissue-specific (Lipg, Lpin1) and tissue-wide (Dbp, Cirbp) genes whose oscillations can be clock-dependent (Lipg, Dbp) or -independent (Lpin1, Cirbp) manner.
Supplemental Figure S3 – Modules of rhythmic gene expression across different subsets of tissues

(A) Heatmaps for the modules corresponding to genes rhythmic in different subsets of tissues. Each rectangle, demarcated by white vertical lines, represents normalized mRNA accumulation over 48 hours (red denotes low expression, green denotes high) for each gene in module (y-axis) in a single tissue. Schematic of the combination of tissues in which genes are rhythmic are shown above each heatmap. For clarity, genes in tissue-wide module (bottom) are filtered for amplitudes with average log2 fold change greater than 0.8.

(B) Heatmaps representing modules of liver- and kidney-rhythmic genes that are driven by the local clock or systemic cues.
Supplemental Figure S4

A

B

C

D

Gene: Egr1
TF: EGR

Abundance or motif activity (scaled)

Gene: Tgif1
TF: TGIF1

Abundance or motif activity (scaled)
Supplemental Figure S4 - TF regulators associated with tissue-specific rhythmic gene expression

(A) Predicted activities of TF motifs associated with system-driven liver-rhythmic module (left).

(B) Predicted TF activity (solid line), nuclear protein abundance (if available, triangles), and mRNA accumulation (dotted) oscillate in both WT and Bmal1 KO. Error bars in nuclear protein, mRNA, and TF activity are SEM (n=2).

(C,D) Predicted activities of TF motifs associated with clock-driven kidney-rhythmic (C), and clock-driven liver-rhythmic (D) module represented by the first component of complex-valued SVD. Motifs with z-score > 1.25 are shown. Candidate TFs shown in Figure 4B,D,F are displayed in bold.
Supplemental Figure S5

A

- RORE
- CDC5L
- AR
- E-box
- POUSF1
- EV1
- NR3C1
- HAND1.2
- NFKB1_REL_RELA
- ZEB1

Activity [A.U.]

0.0 1.5 6

12 18 24

ZT

B

- RORE
- RORE + liver-specific TF motifs

Log2 fold change

0 1.5 3

12 18 24

ZT
Supplemental Figure S5 – Liver-specific accessible regions harboring clock TF binding sites underlie clock-driven liver-specific rhythms

(A) Activities of TF motifs associated with clock-driven liver-rhythmic module, predicted using TF binding site occurrences at liver-specific DHSs.

(B) Genes containing RORE (red) or co-occurrence of RORE and liver-specific TF (ONECUT1, CUX2, or FOXA2, dark red) in clock-driven liver-rhythmic module. mRNA abundances of genes with liver-specific DHSs harboring RORE motifs peak approximately 3 hours after peak RORE activity.
Supplemental Figure S6

A  

Pik3ap1

WT  

Bmal1 KO

Log2 mRNA abundance

Liver  

Kidney

B

ZT

Log2 mRNA abundance

Liver  

Kidney

C

D

log_{10} 4C signal

-log_{10} (pvalue)

(liver - kidney)

Position relative to bait (kb)

Position relative to bait (kb)

Position relative to bait (kb)

Position relative to bait (kb)

Tm9sf3  

Pik3ap1  

AK087395

DNase-I

4C

DNase-I

4C
**Supplemental Figure S6 - Chromatin loops between liver-specific enhancers and promoter of Pik3ap1**

(A) Temporal mRNA abundance profiles of Pik3ap1 in liver and kidney of mice with (left) and without (right) a functioning clock. Nuclei were extracted from WT liver and kidney from 4 mice at ZT8 to perform 4C-Seq.

(B) 4C-Seq profiles (summary from 2 samples, each pooled from 2 mice) using the Pi3kap1 promoter as a bait in liver and kidney within a window of 500 kb.

(C) Profiles of differential contacts for liver versus kidney, shown as signed log p-values (regularized t-test, positive values show liver-enriched 4C contacts).

(D) Tracks of differential contacts (signed log p-values), DNAse-I hypersensitivity in liver, kidney, and their log2 fold change. Regions of significant differential contacts correspond to liver-specific DHS regions.
Supplemental Figure S7 - Chromatin loops between liver-specific enhancers and promoters of *Slc44a1* transcript

(A) Temporal mRNA abundance profiles of *Slc44a1* in liver and kidney of mice with (left) and without (right) a functioning clock. Nuclei were extracted from WT liver and kidney from 4 mice at ZT20 to perform 4C-Seq.

(B) 4C-Seq profiles (summary from 2 samples, each pooled from 2 mice) using the *Slc44a1* promoter as a bait in liver and kidney within a window of 500-kb.

(C) Profiles of differential contacts between liver versus kidney, shown as signed log p-values (regularized t-test, positive values show liver-enriched 4C contacts).

(D) Tracks of differential contacts (signed log p-values), DNase-I hypersensitivity in liver, kidney, and their log2 fold change. Regions of significant differential contacts correspond to liver-specific DHS regions.
Supplemental Figure S8 - Correlations of alternative transcript start site (TSS) usage and tissue-specific rhythms in mRNA accumulation

(A) Scatterplot of alternative promoter usage versus log2 fold-change (FC) of rhythmic tissues. Alternative TSS usage defined by calculating the Euclidean distance of transcript expression in tissues with rhythmic transcript versus tissues with nonrhythmic transcripts.

(B) Example of alternative TSS usage in rhythmic gene expression. RNA-Seq of Ddc at CT34 shows that rhythmic tissues (liver and kidney) use an upstream promoter whereas nonrhythmic tissues use a downstream promoter.
Supplemental Figure S9

**Slc45a3**

- **Tissue**: Liver, Kidney
- **Genotype**: WT, Bmal1 KO
- **Slc45a3**: short, long

Log2 mRNA abundance over ZT (Zoe time) for liver and kidney tissues under WT and Bmal1 KO genotypes.
Supplemental Figure S9 — Liver and kidney use different TSSs in *Slc45a3*, a clock-driven liver transcript

Temporal abundance profiles of two transcripts of *Slc45a3* in liver and kidney of mice with and without a functioning clock. *Slc45a3*-short isoform is rhythmic specifically in the liver; *Slc5a3*-long isoform is not robustly rhythmic in liver or kidney.
Chapter 2. Transcription factor activity rhythms and tissue-specific chromatin interactions explain circadian gene expression across organs

2.6 Conclusion and perspectives

This paper investigates transcriptional regulatory modes that allow diurnal gene expression to be tissue-specific. Our study reveals that the circadian clock is embedded into tissue-specific regulatory networks. Tissue-specific chromatin interactions can establish this regulation, allowing circadian clock transcription factors to regulate a gene in a tissue-specific manner. In fact, genes that oscillate with 24-hour rhythm in all tissues are a small minority compared tissue-specific or tissue-restricted oscillations.

One mode by which the clock can embed into tissue-specific regulatory networks is through tissue-specific chromatin interactions. Our data suggest that circadian clock transcription factors (TFs) can bind to tissue-specific enhancers, which allow clock TFs to regulate gene expression oscillations in a tissue-specific manner through promoter-enhancer looping. Further questions and extensions for understanding tissue-specific circadian gene expression include:

1. Incorporate mean expression to comparisons of oscillatory gene expression across tissues.
2. Extend method to 12-hour rhythms.
3. Investigate post-transcriptional regulation underlying tissue-specific gene expression
3 Clock-dependent chromatin topology modulates circadian transcription and behavior

3.1 Introduction

This chapter looks into the circadian dynamics of promoter-enhancer looping and the function of rhythmically active enhancers in gene transcription as well as circadian locomotor activity. The work has been published in Genes & Development, 2018 under the Creative Commons license and is reproduced here. The main text as well as supplemental figures are attached. Due to the size of supplemental tables, they are not attached but can be found on the online version of the Genes & Development article. No changes were made to the main text or supplemental figures, which were downloaded from doi:10.1101/gad.312397.118.

3.2 Contributions

Jérome Mermet did the 4C-seq and RNA-seq experiments, with help from Céline Jouffe, Damien Nicolas, and Daniel Mauvoisin. I performed the analysis of the 4C-seq, RNA-seq, and ChIP-seq data, with help from Kyle Gustafson and Felix Naef in the initial stages of the project. Clémence performed and analyzed the smRNA-FISH experiments, with some help from me in the analysis. Yann Emmenegger from the Paul Franken lab performed the locomotor activity experiments.
Clock-dependent chromatin topology modulates circadian transcription and behavior

Jérôme Mermet,1,4 Jake Yeung,1,4 Clémence Hurni,1 Daniel Mauvoisin,1 Kyle Gustafson,1 Céline Jouffe,7 Damien Nicolas,1 Yann Emmenegger,3 Cédric Gobet,1,5 Paul Franken,7 Frédéric Gachon,1,2 and Félix Naef1

1School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland; 2Nestle Institute of Health Sciences, CH-1015 Lausanne, Switzerland; 3Center for Integrative Genomics, University of Lausanne, CH-1015 Lausanne, Switzerland

The circadian clock in animals orchestrates widespread oscillatory gene expression programs, which underlie 24-h rhythms in behavior and physiology. Several studies have shown the possible roles of transcription factors and chromatin marks in controlling cyclic gene expression. However, how daily active enhancers modulate rhythmic gene transcription in mammalian tissues is not known. Using circular chromosome conformation capture (4C) combined with sequencing (4C-seq), we discovered oscillatory promoter–enhancer interactions along the 24-h cycle in the mouse liver and kidney. Rhythms in chromatin interactions were abolished in arrhythmic Bmal1 knockout mice. Deleting a contacted intronic enhancer element in the Cryptochrome 1 (Cry1) gene was sufficient to compromise the rhythmic chromatin contacts in tissues. Moreover, the deletion reduced the daily dynamics of Cry1 transcriptional burst frequency and, remarkably, shortened the circadian period of locomotor activity rhythms. Our results establish oscillating and clock-controlled promoter–enhancer looping as a regulatory layer underlying circadian transcription and behavior.

[Keywords: circadian rhythms; chromatin topology; promoter–enhancer loops; DNA regulatory elements; transcriptional bursting]

Supplemental material is available for this article.

Received January 26, 2018, revised version accepted March 2, 2018.
regulation of core clock function and control of circadian gene expression, remain an open question. Indeed, rhythmic transcription could be regulated over an established static promoter–enhancer network (Ghavi-Helm et al. 2014; Xu et al. 2016), or, conversely, the clock could drive dynamic promoter–enhancer looping for high-amplitude daily oscillations in transcription.

Here we monitored promoter–enhancer contacts of a core clock and metabolic clock output gene across time and genotypes in mouse tissues and discovered that contact frequencies oscillated along the 24-h cycle. In arrhythmic Binall knockout animals, these oscillations were abolished. Deletion in mice of an enhancer that was rhythmically recruited to the Cryptochrome 1 (Cry1) promoter led to a short period phenotype in locomotor activity. Moreover, this deletion compromised rhythmic chromatin topology in the liver and led to reduced peak Cry1 mRNA expression levels. Finally, single-molecule RNA fluorescent in situ hybridization (smRNA-FISH) showed that the abolished rhythmic chromatin contact reduced the daily dynamics of Cry1 transcriptional burst frequency.

Results

Rhythmic local chromatin interactions in mouse livers

We focused on two genes representing key temporally regulated hepatic functions: a gene essential for the core circadian oscillator, Cry1 (Griffin et al. 1999; van der Horst et al. 1999), and a liver-specific clock-controlled gene, Glycogen Synthase 2 (Gys2) (Doi et al. 2010), which encodes the rate-limiting enzyme in hepatic glycogen synthesis (Irimia et al. 2010). These transcripts are rhythmically expressed in the liver at opposite times of day, Cry1 peaking during the night at Zeitgeber time 20 (ZT20) and Gys2 peaking during the day at ZT08 (with ZT0 corresponding to lights on and ZT20 corresponding to lights off) (Supplemental Fig. S1A). Using circular chromosome conformation capture (4C) combined with sequencing (4C-seq) (Gheldof et al. 2012), we estimated the interaction frequencies of DNA bait fragments placed near the transcription start sites (TSSs) of Cry1 and Gys2 versus the entire genome in livers of wild-type mice collected at ZT08 and ZT20 (n = 4 per time point). 4C-seq signals around the Cry1 and Gys2 TSSs decayed to background levels following a power law (Supplemental Fig. S1B,C; Supplemental Table S1; Sanborn et al. 2015) and did not exceed background on trans chromosomes (Supplemental Fig. S1D,E; Supplemental Table S1). The high proportion of chromatin interactions within the first 2 Mb surrounding the baits on the cis chromosome [Cry1 TSS: 41% of total cis contacts at ZT08 and 46% at ZT20, Gys2 TSS: 54% at ZT08 and 57% at ZT20] indicated that Cry1 and Gys2 regulatory contacts were contained within this signal-rich region (Sanyal et al. 2012). To compare 4C-seq profiles across conditions, we normalized the data and applied locally weighted multilinear regression (LWMR), which uses a Gaussian window (σ = 2500 kb) centered on each fragment for local smoothing (Materials and Methods). For Cry1, the 4C-seq profiles after LWMR were similar between ZT08 and ZT20 except in a region downstream from the Cry1 promoter, where the contact frequency was increased at ZT20 (Fig. 1A). While the
differential signal covered the entire Cry1 locus, the largest difference was localized—peaking 26 kb downstream from the TSS in the first Cry1 intron—and highly significant ($P < 5.5 \times 10^{-17}$ at the peak) [Fig. 1B, bottom tracks, vertical dotted line at the left]. A secondary peak was observed near 8.7 kb below the TSS [Fig. 1C, bottom tracks, vertical dotted line at the right; Supplemental Fig. S2A; Supplemental Table S1]. Thus, these 4C-seq data in the liver suggested dynamic contacts between the Cry1 promoter and the +26-kb intronic site as well as the −7-kb upstream site. Since Cry1 mRNA accumulated rhythmically in the kidney [Supplemental Fig. S3A], we also performed 4C-seq in kidneys. Consistent with the liver data, these sites were also recruited to the Cry1 promoter more frequently at ZT20 than at ZT08 [Supplemental Fig. S3B,C].

Opposite to Cry1, the Gys2 promoter contacted an intragenic region more frequently at ZT08 versus ZT20 [Fig. 1C], with a peak 21 kb downstream from the TSS in exon 8 ($P < 2.3 \times 10^{-19}$ at peak) [Fig. 1C, bottom tracks, brown solid line, vertical dotted line at the left], consistent with its anti-phasic rhythmic mRNA accumulation [Supplemental Fig. S1A]. This significant differential signal was validated by reciprocal 4C-seq using the exon 8 as bait [Fig. 1C, bottom tracks, black solid line, vertical dotted line at the right; Supplemental Fig. S2B]. In the kidney, where Gys2 mRNA accumulation was constant and low, this differential signal was absent [Supplemental Fig. S3D-F]. Thus, both gene promoters formed DNA loops with neighboring intragenic regions in cis that coincided with the timing of the respective peaks in Cry1 and Gys2 mRNA expression.

The dynamics of chromatin topology depend on BMAL1

To test whether these dynamic contacts depended on a functional circadian clock, we performed 4C-seq in the livers of clock-deficient animals (Bmal1 knockout) in which Cry1 and Gys2 lost rhythmic expression and were constantly expressed at high and low levels, respectively [Supplemental Fig. S4A,B]. In Bmal1 knockout, the Cry1 +26-kb intronic and −7-kb upstream regions contacted the promoter at comparable frequencies at ZT20 and ZT08, suggesting static chromatin loops [Fig. 2A,B]. For Gys2, the profile between the exon 8 region and the promoter was also static [Fig. 2C,D]. Comparing wild-type and Bmal1 knockout at both time points revealed that for Cry1, the loop was locked in a closed conformation [Supplemental Fig. S4C, constitutively high frequencies], and for Gys2, it was locked in an open conformation [Supplemental Fig. S4D, constitutively low frequencies]. Thus, the closed and open states of DNA loops concurred with high and low transcription, respectively [Supplemental Fig. S4, cf. A, C and B, D]. We note that these 4C profiles suggested a BMAL1-independent interaction upstream of Gys2 [Fig. 2C, lower panels], but this effect was less robust compared with the BMAL1-dependent intragenic looping. As a negative control, we tested the Hoxd4 locus, which is a transcriptionally silent region in the adult liver. As expected, chromatin contact profiles at the Hoxd4 locus remained static over time in both wild-type and Bmal1 knockout livers [Fig. 2E, Supplemental Fig. S4E]. These data thus showed that rhythmic loops in Cry1 and Gys2 depended on the clock TF BMAL1.

Rhythmic DNA loops connect gene promoters with daily active enhancers

To characterize the interacting genomic regions, we integrated temporal data on DNase-I hypersensitivity sites [DHSs] with ChIP-seq [chromatin immunoprecipitation [ChIP] combined with high-throughput sequencing] data for RNA polymerase II [Pol II], the activity-related chromatin mark H3K27ac [Sobel et al. 2017], and rhythmically active TFs [Rey et al. 2011; Zhang et al. 2015]. This allowed us to assess whether the rhythms in DNA contacts coincided with rhythms in activity-related chromatin marks. For Cry1, RNA Pol II and H3K27ac signals peaked near ZT20 [Fig. 3], while, for Gys2, they peaked near ZT08 [Fig. 4]. However, while RNA Pol II signals extended throughout the gene bodies, H3K27ac signals were spatially confined around the largest differential contact precisely at sites marked with DHSs. Furthermore, both the 26-kb downstream intronic site and the −7-kb upstream site of the Cry1 TSS contained a RORE-responsive element [RRE] and were bound by the circadian TFs REV-ERBa and RORγ [Fig. 3; Supplemental Table S4; Zhang et al. 2015]. In mouse fibroblasts, the intronic RRE is required for proper timing of Cry1 expression [Ukai-Tadenuma et al. 2011]. The interacting Gys2 exon 8 site was bound by the clock regulator Bmal1 at ZT06 [Rey et al. 2011] and by REV-ERBa at ZT10 [Fig. 4]. This indicated that DNA contacts connected local rhythmically active enhancer elements with the promoters of Cry1 and Gys2.

Deleting the Cry1 intronic enhancer in mice shortens the circadian locomotor period

To study the function of the rhythmic chromatin interactions, we generated a mouse strain [Cry1Δe] with a 300-base-pair [bp] deletion covering the Cry1 intronic enhancer [Supplemental Fig. S5A,B]. We measured spontaneous locomotor activity in constant darkness and observed that Cry1Δe animals had an endogenous circadian period that was significantly shorter ($P < 1.1 \times 10^{-5}$, t-test) by 15 min compared with wild-type littermates [Fig. 5A; Supplemental Fig. S5C,D]. Such period shortening is in the range of classic short period core clock mutants such as Per1 [Cermakian et al. 2001] and Clock [Debruyne et al. 2006]. As Cry1 loss of function shortens the circadian period by 1.2 h [van der Horst et al. 1999], our noncoding DNA deletion suggests a Cry1 hypomorph.
Expression of Cry1, clock, and clock output genes is perturbed in Cry1Δe

To investigate the link between the deletion, promoter-enhancer looping, and Cry1 expression in livers and kidneys, we first generated temporal RNA sequencing (RNA-seq) data in kidneys, we first generated temporal RNA sequencing data in Cry1Δe and wild-type littermates under an entraining light-dark cycle. The transcriptomes in Cry1Δe and wild-type littermates were comparable overall in both tissues (Supplemental Fig. S6A). While Cry1 mRNA levels remained rhythmic in both genotypes, likely driven by further regulatory sites (e.g., the TSS and −7-kb sites), the peak expression at ZT20 was significantly reduced by 27% in the livers (15% in the kidneys) of Cry1Δe animals compared with wild type [Supplemental Fig. S6B]. Quantifying the intronic reads as a proxy for transcription showed that Cry1 transcription was also phase-advanced in Cry1Δe animals [Supplemental Fig. S6C]. Moreover, Cry1 protein abundance in the liver was lower in Cry1Δe compared with wild type, consistent with a reduction in mRNA levels [Supplemental Fig. S6D,E].

As is known in chronobiology, entraining a short period circadian oscillator by an external light-dark cycle leads to a phase advance of internal timing markers (Aschoff and Pohl 1978). This prediction was confirmed in the transcriptome data. Indeed, core clock and clock-controlled genes [Supplemental Table S5] were phase-advanced by, on average, 30 min in the livers of Cry1Δe animals compared with wild type [P < 0.01 binomial test] [Supplemental Fig. S6F], with Cry1 showing the largest phase advance [P = 0.011 for livers; P = 0.047 for kidneys, bootstrap test] [Supplemental Fig. S6B].

The Cry1Δe mutation disrupts rhythmic chromatin topology

Next, we explored the dynamics of chromatin topology along the 24-h cycle in liver sampled every 4 h in wild type and Cry1Δe [n = 3 per time point]. First, we confirmed oscillatory chromatin interactions in Gys2 in wild type. Indeed, the Gys2 promoter rhythmically recruited the +21-kb enhancer, peaking near ZT08 in both the TSS bait and exon 8 bait (P < 10−6 at the peak harmonic regression) [Supplemental Fig. S7A–C]. As negative control, the Hoxd4 bait measured around the clock did not show oscillatory contacts [Supplemental Fig. S7D]. For Cry1 wild type, the frequency of contacts between the promoter and the +26-kb enhancer significantly oscillated, peaking near ZT30 [P < 10−8 at the peak] [Fig. 5B–D; Supplemental Fig. S8A,B]. In contrast, in Cry1Δe, the contact frequencies in this region were lower at all time points compared with wild type, and the oscillation was compromised [Fig. 5B–D; Supplemental Fig. S8A,B]. Finally, we also estimated chromatin contacts for a bait placed at the −7-kb upstream enhancer [Fig. 1B, bottom tracks, vertical dotted line at the right], showing oscillation in contact frequency peaking around ZT20 with the +26-kb intronic enhancer in wild type but...
nonrhythmic and overall lower contact frequency in Cry1Δe (Supplemental Fig. S8C,D). Decreased contact frequency in Cry1Δe mice indicates that the RRE-containing 300-bp fragment drives the promoter–enhancer loop.

Overall, these data demonstrate robust rhythmic chromatin topology for Cry1 and Gys2, where the frequency of enhancer–promoter contacts is modulated with time of day. Furthermore, deleting a localized noncoding DNA enhancer element (300 bp) in the Cry1 gene could disrupt such rhythms.

**The Cry1 intronic enhancer modulates transcriptional burst frequency**

To analyze whether the Cry1 intronic enhancer modulates transcription, we estimated transcriptional parameters by smRNA-FISH against Cry1 pre-mRNA in the livers of wild-type and Cry1Δe animals at ZT08 and ZT20 (Fig. 6A). Mammalian promoters are irregularly transcribed (transcriptional bursting), as characterized by the burst size and burst frequency (Suter et al. 2011; Bahar Halpern et al. 2015). Taking into account the ploidy of cry1Δe mice...
liver nuclei (Supplemental Fig. S9A–D), smRNA-FISH showed that Cry1 burst fraction (fraction of active transcription sites in each nucleus, which is proportional to the burst frequency per allele) was 2.2-fold higher at ZT20 compared with ZT08 in wild type (Fig. 6B). Importantly, the burst fraction was reduced by 28% in Cry1Δe animals at ZT20 (Fig. 6B). In contrast, the burst intensity (proportional to the burst size) was similar in all conditions (Fig. 6C). Thus, the lowered Cry1 mRNA levels in Cry1Δe at ZT20 can be quantitatively explained by the reduced burst fraction. In sum, dynamic enhancer loops modulate transcriptional bursting in mammalian tissues (Bartman et al. 2016; Fukaya et al. 2016); in particular, rhythmic DNA loops involving clock enhancers control burst frequency while maintaining burst size.

Discussion

In animals, developmental transitions occurring on the time scales of days have been shown to involve remodeled DNA contacts and promoter–enhancer loop formation [Noordermeer et al. 2014]. While such dynamics are typically irreversible, we here discovered that chromatin topology in mouse tissues can be locally (100 kb, in cis) plastic, exhibiting temporal dynamics that are regulated by daily time and the circadian oscillator and thus recur within a 24-h period. While previous work in cell culture reported dynamic chromatin contacts on larger genomic scales, notably between the Dbp gene and DNA regions on trans chromosomes [Aguilar-Arnal et al. 2013], the genes analyzed here did not show rhythmic chromatin...
interactions on such scales. We then showed genetically that these rhythmic DNA contacts depend on the clock protein BMAL1 and, in the case of Cry1, a 300-bp intronic RRE-containing enhancer sequence.

How is BMAL1 involved in the formation of these dynamic loops? In the case of Gys2 in Bmal1 knockout mice, the loop is constitutively open, and Gys2 mRNA expression is constitutively low. Combined with the binding of BMAL1 at the looping site, these data strongly argue for a direct involvement of BMAL1. For Cry1, the activator RORs and the repressor REV-ERBα bind to the Cry1 intronic enhancer at the expected peak (ZT20) and trough (ZT08) activities, as is typical of functional RREs. We note that while the expression of the RRE-binding repressors Rev-Erbα/β is low in Bmal1 knockout, the corresponding activator RORγ is constitutively high (Atger et al. 2015). Therefore, the constitutively closed Cry1 loop in the Bmal1 knockout most likely reflects an indirect effect via perturbed REV-ERB and ROR activities. This is further corroborated by the constitutively open state of the Cry1 promoter–enhancer loop in Cry1Δe mice, showing chromatin interactions that are constantly below wild-type trough levels, indicating that loop-promoting factors (for example, RORs) act within the 300-bp element. Therefore, our data suggest a canonical mechanism of enhancer–promoter looping by which sequence-specific TFs help recruit transcription complexes, which facilitate the function of Pol II at core promoters (Levine and Tjian 2003).

To investigate the effects of the dynamic looping on transcriptional parameters, we complemented bulk 4C-seq and RNA-seq experiments with single-molecule transcript analysis in situ, which revealed that the abolished rhythmic chromatin contact in Cry1Δe mice reduced Cry1 transcriptional burst frequency. These results in mammalian tissues contribute to our current understanding of how enhancer loops modulate transcriptional bursting (Bartman et al. 2016; Fukaya et al. 2016). In particular, we showed that rhythmically active clock enhancers can increase burst frequency while not changing burst size.

The ablation of the Cry1 noncoding regulatory element even led to a short period phenotype in locomotor activity. While noncoding genetic variation in humans has been associated recently with circadian clock-related and sleep phenotypes (Allebrandt et al. 2010; Hu et al. 2016), no demonstration of such variation on circadian transcription or behavior has yet been provided. Indeed, previously characterized mutations impacting mammalian circadian behavior have concerned protein-coding regions (Vitaterna et al. 1994; Toh et al. 2001). Here, we provided evidence that noncoding regulatory elements within the
core circadian regulatory network can drive dynamic pro-
moter–enhancer looping, modulate temporal transcrip-
tion, and regulate circadian locomotor behavior.

Materials and methods

Animal and ethics statement

All animal care and handling were performed according to Can-
ton de Vaud laws for animal protection [authorization VD2801
[Frédéric Gachon] and VD3109 [Félix Naef]]. All experiments
were performed on males between 8 and 10 wk old. Bmal1 knock-
out animals were described previously in Jouffe et al. (2013).

Mouse genome editing by direct knockout using CRISPR–Cas9

Px-330 plasmids targeting upstream of and downstream from the
Cry1 intron1 regulatory region were injected into pronuclei and
then transplanted into B6D2F1 pseudopregnant mice at the Ecole
Polytechnique Fédérale de Lausanne (EPFL) Transgenic Core Fa-
cility (http://tcf.epfl.ch). Pups from the first generation (F0) were
then screened for the deletion using the PCR primers indicated
in Supplemental Table S2. F0 animals of interest were back-
planted into B6D2F1 pseudopregnant mice at the Ecole
Polytechnique Fédérale de Lausanne (EPFL) Transgenic Core Fa-
cility (http://tcf.epfl.ch). Pups from the first generation (F0) were
then screened for the deletion using the PCR primers indicated
in Supplemental Table S2. F0 animals of interest were back-
planted into B6D2F1 pseudopregnant mice at the Ecole
Polytechnique Fédérale de Lausanne (EPFL) Transgenic Core Fa-
cility (http://tcf.epfl.ch). Pups from the first generation (F0) were
then screened for the deletion using the PCR primers indicated
in Supplemental Table S2. F0 animals of interest were back-

Nucleus purification and fixation

Immediately after sacrifice, 5 mL of 1× PBS was perfused through
the spleen to flush blood from the liver. Livers and kidneys from
individual animals were homogenized and fixed in 4 mL of 1×
PBS, including 1.5% formaldehyde, for 10 min at room tempera-
ture. The cross-linking reaction was stopped by adding 25 mL of
ice-cold stop reaction buffer (2.2 M sucrose, 150 mM glycine, 10
mM HEPES at pH 7.6, 15 mM KCl, 2 mM EDTA, 0.15 mM sper-
midine, 0.5 mM DTT, 0.5 mM PMSF) to the ho-
mogenates and was kept for 5 min on ice. Homogenates were then
loaded on top of 10 mL of cushion buffer (2.05 M sucrose, 10%
glycerol, 125 mM glycine, 10 mM HEPES at pH 7.6, 15 mM KCl,
2 mM EDTA, 0.15 mM spermidine, 0.5 mM spermidine, 0.5 mM
DTT, 0.5 mM PMSF) and centrifuged at 105,000 × g for 4 h at 16°C.
Nuclei were washed twice in 1× PBS and immediately frozen.

4C-seq

4C template preparation 4C templates were prepared as in Ghel-
dol et al. (2012). Nuclei were resuspended in 1 mL of a buffer con-
taining 10 mM Tris-HCl (pH 8.0), 10 mM NaCl, 0.2% NP-40,
and 1× protease inhibitor cocktail (Complete Mini EDTA-free
protease inhibitor cocktail, Sigma-Aldrich), kept for 15 min on
ice, and washed twice with 1× DpnII buffer [New England Biol-
abs]. Thirty million nuclei were resuspended in 1× DpnII buffer
[New England Biolabs] containing 0.1% SDS and incubated for
10 min at 65°C. Triton X-100 was added to 1% final concentra-
tion. Chromatin was digested overnight with 400 U of DpnII
[New England Biolabs] at 37°C with shaking. After heat inactiva-
tion, digestion efficiency was evaluated by both DNA visualiza-
tion on agarose gels and quantitative PCR using primer pairs
covering multiple restriction sites. Chromatin was then ligated
with 3000 U of T4 DNA ligase [New England Biolabs] in an 8-
µL final volume for 4 h at 16°C plus 1 h at room temperature.
The cross-linking reaction was reverted by the addition of
50 µL of 10 mg/mL proteinase K and incubation overnight at
65°C. DNA was purified by multiple phenol/chloroform extrac-
tions, resuspended in TE buffer [pH 8.0] containing RNase A,
and incubated for 30 min at 37°C. Ligation efficiency was evalu-
ated by loading DNA on an agarose gel. Libraries were digested
with 1 U of NlaIII per microgram of template [New England Biol-
abs] overnight at 37°C, and digestion was controlled by visualiza-
tion on an agarose gel. After heat inactivation, digested products
were ligated with 2000 U of T4 DNA ligase [New England Biol-
abs] for 4 h at 16°C in a 14-µL final volume. Circularized pro-
ducts were purified and resuspended in TE buffer [pH 8.0]. 4C
templates were prepared in four biological replicates in wild-
type mouse livers and kidneys and three biological replicates in
the livers of Bmal1 knockout and Cry1ΔΔe and wild-type litter-

Figure 6. The oscillatory Cry1 promoter–enhancer loop modulates
Cry1 transcriptional bursting. (A) smRNA-FISH against Cry1 pre-
mRNA in the livers of wild-type (top) and Cry1ΔΔe (bottom) animals at ZT08 [left] and ZT20 [right]. Burst fractions (B) and burst intensities
[C] measured from images of smRNA-FISH performed against Cry1 pre-mRNA in Cry1ΔΔe (dashed) and wild-type (solid) livers at ZT08 (red)
and ZT20 (blue). Burst fraction is the number of active transcription sites in each nucleus divided by the ploidy. (B, C) Shown are the means
and standard errors over nuclei collected and pooled from two animals in each of the four conditions (individual animals are analyzed in
Supplemental Fig. S8C,D). n = 2191 wild-type ZT08 nuclei; n = 983 Cry1ΔΔe ZT08 nuclei; n = 2150 wild-type ZT20 nuclei; n = 1473 Cry1ΔΔe
ZT20 nuclei. In B, [*] P < 0.05, [**] P < 0.001, t-test. In C, differences between genotypes are not significant.
Inverse PCR and sequencing in wild-type and Bmal1 knockout mouse livers and kidneys Six-hundred nanograms of 4C template was used for PCR amplification using Sigma-Aldrich long-template PCR system with bait-specific inverse primers conjugated to Illumina sequencing adaptors (primer sequences are in Supplemental Table S3) in a final volume of 50 µL in the following PCR program: 2 min at 94°C followed by 30 cycles of 15 sec at 94°C, 1 min at 55°C, and 3 min at 68°C and a final extension of 7 min at 68°C. PCR products were performed in parallel reactions with 6 x 100 ng of template for each sample. PCR products were purified with the AMPure XP beads system (Beckman Coulter), and amplification profiles were analyzed by fragment analyzer and then sequenced on Illumina HiSeq 2000 machines using single-end 100-bp read length.

Inverse PCR and sequencing in the livers of Cry1ae and wild-type littermates Six-hundred nanograms of 4C template was used for PCR amplification using Sigma-Aldrich long-template PCR system with two-step PCR system from Illumina. Bait-specific inverse primers conjugated to Illumina sequencing adaptors (primer sequences are in the Supplemental Table S3) were used in a first PCR reaction in a final volume of 50 µL with the following program: 2 min at 94°C followed by 20 cycles of 15 sec at 94°C, 1 min at 55°C, and 3 min at 68°C and a final extension of 7 min at 68°C. PCRs were performed in parallel reactions with 6 x 100 ng of template for each sample. PCR products were purified with the AMPure XP beads system (Beckman Coulter) and then sequenced on NextSeq 500 machines using single-end 75-bp read length.

4C-seq analysis

Preprocessing computational methods Demultiplexed Fastq files were mapped to the mouse genome (mm9) using Bowtie2 with default HTSstation parameters (http://htsstation.epfl.ch). Since each restriction fragment contained two mapping sites (two ends of the fragment), the fragment score was computed as the average of the number of reads per mapping site.

Quality control of 4C-seq data Samples with ≥75% of restriction fragments without any counts in a window of ±1 Mb upstream of and downstream from each bait were not analyzed (Supplemental Table S1). The first five fragments upstream of and downstream from the bait (10 total) were not considered in the analysis because they mostly contained partially digested and self-ligated products.

Normalization and LWMR We follow a method developed recently in Yeung et al. (2018) with minor modifications. Briefly, raw read counts for each sample were library size-rescaled by the normalized sum of the read counts on the cis chromosome (excluding 10 restriction fragments around the bait). To control the variability of low signals, in subsequent analyses, the fragment counts c in each sample were log transformed using the variable

\[ Y = \log_{10}(c + 1) \]

with \( P = 500 \). A weighted linear model was then fit locally using a Gaussian window [\( w_2 = 2500 \text{ bp} \)] centered on the fragment of interest. For each position, nearby 4C-seq signals \( Y \) were modeled with fragment effects \( a \), and condition effects \( b \) (which can be time, tissue, or genotype). In LWMR, these parameters were estimated by minimizing the weighted sum of squared residuals across replicates \( r \):

\[ S = \arg\min\left(\sum_{i,j,r} W_{ijr} Y_{ijr} - a_i - b_j r^2\right) \]

with weights \( W_{ijr} \) defined as \( W_{ijr} = w_{ijr} \times v_{ijr} \), where \( w_{ijr} \) is the Gaussian smoothing kernel at position \( i \), and \( v_{ijr} \) is a condition weight based on the number of samples with nonzero counts on fragment \( i \). Specifically, we used \( w_{ijr} = 0.5, 1, 2.5, 3.5, \) or 4.5 for fragments with zero, one, two, three, or four replicates showing nonzero counts, which down-weights positions with high dropout rates. To estimate the statistical significance for differential contacts (for example, ZT20 vs. ZT08), we propagated the estimated uncertainty (standard errors for locally weighted regression) in the corresponding \( b \) values to calculate Z-scores and used regularized \( t \) statistics with \( n - p \) degrees of freedom (DOF, \( n \) is the number of data points within window, and \( p \) is the number of parameters). For the analysis of 24-h rhythmicity in contacts (weighted harmonic regression), we proceeded analogously by propagating the uncertainty in the \( b_s \) for the six time points to that in the squared 24-h Fourier coefficient and used the \( \chi^2 \) test with two DOF (owing to the real and imaginary parts). For each set of 15 samples, we computed the regularized residual variance as

\[ \sigma^2 = \sigma^2 + \sigma_{\text{min}}^2 \exp\left(-\frac{\hat{Y}_i}{\overline{Y}}\right) \]

with \( \sigma^2 \) as the estimator of the squared residuals, \( \hat{b} \) as the estimated signal across samples, and \( \sigma_{\text{min}} = \log(q(2)). \sigma_{\text{min}} \) prevents artificially small variance from positions of high dropout rates and is estimated from the distribution of \( \sigma^2 \) across all fragments. \( \overline{Y} \) ranges from 0.06 to 0.16 (same units as \( Y \)), depending on the bait (Supplemental Table S1).

Rhythmic DNA loops tune transcription and behavior

Rhythmic DNA loops tune transcription and behavior

H3K27ac and RNA Pol II ChIP-seq and DNase-I-seq analysis

Ras files from GSE60578 (Sohle et al. 2017) were analyzed in genomic regions ±1 Mb from the 4C-seq baits. There, read counts were binned in 500-bp intervals and normalized by the library size. The amplitude and phase of the log2 read counts of each of the three signals were calculated for each bin after applying a running average of seven bins (three bins upstream, three bins downstream, and one bin in the center) to smooth the signal. Obtained rhythmic amplitudes and phases were compared with differential 4C-seq signals. The rhythmic signal in each bin [phase, amplitude, and \( \log\text{cal} [p] \)] was mapped to a color using the hue, saturation, and value (HSV) color scheme. Hue \( h \) was defined by the phase of the oscillation, with blue as ZT0. The saturation \( s \) was set to 1. The value \( v \) was set to a color if both amplitude \( X_v \) and \( \log\text{cal} [p] X_s \) were beyond thresholds \( k_v = 1.3, k_s = 4.5 \); otherwise, the color was set to black. To obtain smooth transitions, \( v \) was calculated using a Hill function with Hill coefficient \( n = 5 \) and

\[ v = \min_{\mu} e^{\mu - \frac{\mu}{\mu}} \cdot \frac{\log(10)^{\mu}}{\log(10)^{\mu}} \]

For TT-binding site predictions [Supplemental Table S4], we used weight matrices of TFs defined by SwissRegulon (Pachkov et al. 2007, http://swissregulon.unibas.ch/ecfp/ww/downloads).

RNA-seq in the livers and kidneys of Cry1ae and wild-type littermates

Parts of the livers and kidneys from the animals used for temporal 4C-seq experiments were frozen in liquid nitrogen immediately after sacrifice. Organs were homogenized in 4 M guanidine thiocyanate, 25 mM sodium citrate, 1% β-mercaptoethanol, and 0.2
M sodium acetate. Nucleic acids were extracted with phenol:chloroform-isoamylalcohol, and RNA was precipitated with 4 M LiCl. RNA concentration and purity were measured using nanodrop, and the quality was controlled by fragment analyzer. Poly-A-selected RNA was sequenced on NextSeq 500 machines using single-end 75-bp read length. mRNA levels were quantified using kallisto version 0.42.4 [mm10] [Bray et al. 2016].

**RNA-seq in the livers and kidneys of Bmal1 knockout and wild-type mice**

To complement the mouse liver wild-type and **Bmal1** knockout RNA-seq data [GSE75554], transcriptomes of kidneys from wild-type animals were measured following the same protocol as in Ager et al. [2015]. mRNA levels were quantified using the same method as in Ager et al. [2015].

**Circadian period estimation in Cry1Δe animals and wild-type littermates**

Estimation of the circadian period was performed as in Diessler et al. [2017]. Briefly, 8- to 10-wk-old males were single-caged and kept under 12 h/12 h light/dark cycle for 14 d and switched to constant darkness for 21 d. During the 5 wk of the experiment, the locomotor activity was recorded with passive infrared sensors. Data were sampled with 5-min resolution and analyzed using the χ² periodogram function in the ClockLab software [ActiMetrics]. Food and water were available ad libitum during the entire experiment.

**Western blotting**

Liver cytoplasmic extracts were prepared as described previously [Jouffe et al. 2013]. Protein extract concentrations were quantified using a BCA protein assay kit [Thermo Fisher Scientific], and 20 µg of liver protein extract was resolved by SDS-PAGE using standard procedures. Densitometry analyses of the blots were performed using the ImageJ software. Naphtol blue and black staining of the membranes was used as a loading control and served as a reference for normalization of the quantified values. CRY1 antibody [1/500] was from Abcam [ab104736].

**smRNA-FISH on mouse liver sections**

Parts of the livers from the same animals used in the 4C-seq and RNA-seq were collected, immediately embedded in O.C.T. compound [Tissue-Tek, Sakura-Finetek USA], and snap-frozen. The RNA-FISH was done on 8-µm cryosections using a RNAscope probe for Cry1 pre-mRNA [Cry1_intron1, catalog no. 5002531] according to the manufacturer’s instructions for the RNAscope fluorescent multiplex assay [Advanced Cell Diagnostics]. Nuclei were counterstained with DAPI, and sections were mounted with ProLong Gold anti-fade mountant [Molecular Probes].

**Microscope image acquisition, quantification, and ploidy assignment**

The sections were imaged using a Leica DM5500 wide-field microscope equipped with a CCD camera [DFC 3000] for fluorescence [Leica Microsystems] and a motorized stage. Z-stacks were acquired (0.2 µm between each Z position, 40 images per frame) with an oil immersion 63× objective. The images were quantified using ImageJ. To detect the fluorescent RNA-FISH spots, a Laplacian filter was applied on a maximal projection, and local maxima were computed. Transcription site fluorescent intensities [burst size] were quantified on the sum projection of the nine best-focused stacks per image. Total transcription site signals were computed using a mask of 3 × 3 pixels. Nuclei were detected using filters, thresholding, and watershed transformation. Ploidy (2N, 4N, or 8N) was assigned to the nuclei based on their diameter [Bahar Halpem et al. 2015]. A four-component Gaussian mixture model was fitted to the diameter distribution [package “mixtools” in R]. Nuclei with a probability of >0.7 to belong to one of the three inferred populations with the smallest means were assigned to 2N, 4N, and 8N, respectively. The Gaussian distribution with the largest variance captured outliers in nucleus diameters (>15–18 µm) and were discarded. Burst fraction was calculated as the number of active transcription sites in each nucleus divided by its estimated ploidy, and these fractions were then averaged over the entire populations of nuclei [Fig. 5B,C].

**Data availability**

Raw and processed sequencing data generated from this study [4C-seq and RNA-seq] have been submitted to Gene Expression Omnibus under accession number GSE101423.

**Acknowledgments**

We thank Isabelle Barde at the École Polytechnique Fédérale de Lausanne [EPFL] Transgenic Core Facility for help generating the **Cry1Δe** mouse strain, Jessica Dessimoz at the EPFL Histology Core Facility for smRNA-FISH experiments, Jacques Rougemont and Marion Leuleu for help with analysis of 4C-seq, Keith Harperman and Corinne Peter from the Genomic Technologies Facility at University of Lausanne, and Bastien Mangeat at the EPFL Gene Expression Core Facility for 4C-seq and RNA-seq. This work was supported by Swiss National Science Foundation grant 31003A-1531340 [to F.N.], European Research Council grant ERC-2010-StG-260667 [to F.N.], and École Polytechnique Fédérale de Lausanne. J.Y. benefits from the Natural Sciences and Engineering Research Council of Canada Postgraduate Studies Doctoral scholarship. Computations and analyses were performed at Vital-IT [http://www.vital-it.ch].

**Author contributions:** J.M., J.Y., and F.N. conceived the study; J.Y., K.G., C.G., and F.N. performed the formal analyses; J.M., C.H., D.M., C.J., D.N., and Y.E. performed the investigations; J.M., J.Y., and F.N. wrote the manuscript; J.M., J.Y., P.F., F.G., and F.N. reviewed and edited the manuscript; P.F., F.G., and F.N. supervised the study; and P.F., F.G., and F.N. acquired the funding.

**Note added in proof**

While this manuscript was in review, similar 24-h rhythmic promoter–enhancer chromatin interactions at the Cry1 locus were reported [Kim et al. 2018].

**References**

Aguilar-Arnal L, Haktim O, Patel VB, Baldi P, Hager GL, Sassone-Corsi P. 2013. Cycles in spatial and temporal chromosomal...

Rhythmic DNA loops tune transcription and behavior

 gluco... glucose tolerance and predisposition to the fasted state in liver
dimensional chromatin organization. Genome Biol 16: 264.
Sanyal A, Lajoie BR, Jain G, Dekker J. 2012. The long-range interac-
Schüler U, Gotici I, Saini C, Gos P, Cutie T, Emmenegger Y, Sim-
Takahashi JS. 2017. Transcriptional architecture of the mamma-
Ukai-Tadenuma M, Yamada RG, Xu H, Ripperger JA, Liu AC, Ueda HR. 2011. Delay in feedback repression by crypto-


Clock-dependent chromatin topology modulates circadian transcription and behavior

Jérôme Mermet, Jake Yeung, Clémence Hurni, et al.

*Genes Dev.* 2018, 32: originally published online March 23, 2018
Access the most recent version at doi:10.1101/gad.312397.118

---

**Supplemental Material**
http://genesdev.cshlp.org/content/suppl/2018/03/23/gad.312397.118.DC1

**Related Content**
**A day in the life of chromatin: how enhancer-promoter loops shape daily behavior**
Benjamin J. Weidemann, Kathryn Moynihan Ramsey and Joseph Bass
*Genes Dev.* March, 2018 32: 321-323

**References**
This article cites 44 articles, 17 of which can be accessed free at:
http://genesdev.cshlp.org/content/32/5-6/347.full.html#related-urls

**Creative Commons License**
This article, published in *Genes & Development*, is available under a Creative Commons License (Attribution 4.0 International), as described at http://creativecommons.org/licenses/by/4.0/.

**Email Alerting Service**
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.
Supplemental Figure 1: 4C-seq signal on cis and trans on baits targeting Cry1 and Gys2. (A) Temporal Cry1 and Gys2 mRNA accumulations in mouse liver (Atger et al., 2015) (reads per kilobase per million, RPKM); error bars: standard deviation (SD) of two animals; red and blue dashed lines: 4C-seq time points. (B,C) 4C-seq counts density vs. genomic position from Cry1 TSS (B) and Gys2 TSS (C) for each biological replicate at ZT08 (red) and ZT20 (blue) in the WT mouse liver. Dashed lines: power-law fit. α are decay exponents. Insets: cumulative counts on the cis chromosome from the bait position to chromosome end. Black vertical line shows 1 Mb from bait. (D, E) Cry1 TSS (D) and Gys2 TSS (E) 4C-seq counts density on trans chromosomes.
Supplemental Figure 2: Validation of rhythmic chromatin interactions at Cry1 and Gys2 by targeting their respective enhancers. (A,B) 4C-seq signals from the Cry1 intron1 bait (A) and Gys2 exon8 bait (B) in mouse liver at ZT08 (red) and ZT20 (blue) and the corresponding Z-score (ZT20-ZT08) and signed – log10(p) in a genomic window of 2 Mb. Largest differential signal occurs at the promoter region of the respective genes. Vertical arrows show location of bait relative to TSS of the respective gene.
Supplemental Figure 3: 4C-seq in kidney show rhythmic chromatin interactions at Cry1 TSS but not at Gys2 TSS.

(A) Cry1 mRNA profile in WT mouse liver (black) (Atger et al., 2015) and kidney (red); error bars: SD of two animals. (B) 4C-seq signal from Cry1 TSS bait in WT kidney at ZT08 (red) and ZT20 (blue) in a genomic window of 200 kb. Vertical lines for Cry1 show rhythmic contacts in kidney and liver. Z-score (ZT20 vs 08) and signed –log10(p) for rhythmic contacts in liver (black) and kidney (red). (C) Same as (B) using a 2 MB window. (D) same as (A) for Gys2 mRNA. (E) same as (B) for Gys2 TSS. Vertical lines for Gys2 show contacts that are static in kidney but rhythmic in liver. (F) same as (C) for Gys2 TSS.
**Mermet Supplemental Figure 4**

**A** Cry1 mRNA accumulation

**B** Gys2 mRNA accumulation

**C**

- Tmem263
- Fn4
- Mart2
- Cry1

**D**

- Recql
- Gys2
- Ldhb
- Kcnj8

**E**

- Hoxd4

Position relative to bait [kb]
Supplemental Figure 4: Dynamics of chromatin topology depends on BMAL1. (A,B) Cry1 (A) and Gys2 (B) mRNA profile in WT and Bmal1 KO mouse liver along the diurnal cycle from (Atger et al., 2015); error bars: SD of two animals. (C,D) 4C-seq signals from Cry1 TSS (C) and Gys2 TSS (D) bait in WT and Bmal1 KO livers at ZT08 and ZT20. Vertical lines show BMAL1-dependent rhythmic contacts. (E) 4C-seq signals from negative control region Hoxd4 in WT and Bmal1 KO livers show static chromatin topology. Bottom: Z-score (ZT20 vs 08) and signed –log_{10}(p) for WT and Bmal1 KO for Hoxd4.
Supplemental Figure 5: Deleting Cry1 intronic enhancer in vivo shortens the free-running period and phase advances clock and clock-controlled genes. (A) Genome browser view showing the CRISPR-Cas9 mediated deletion of the Cry1 intronic enhancer in mouse (Cry1Δe). The evolutionary conserved 300bp deletion (red square) covers the DHS containing a RRE described in mouse fibroblasts (Ukai-Tadenuma et al., 2011) within the H3K27ac marked intronic region. Note that this sequence is as conserved as Cry1 exonic sequences. (B) Agarose gel showing the deletion. (C-D) Actograms showing the spontaneous locomotor activity recorded for 14 days in 12 hours/12 hours light/dark (LD) cycles followed by 21 days in constant darkness (DD) in a representative WT (C) and a Cry1Δe animal (D).
A. Liver ○ WT Δ Cry1∆e  
Kidney ○ WT Δ Cry1∆e

B. Liver

Cry1 mRNA [TPM] vs Zeitgeber Time [ZT]

Cry1 mRNA [TPM]

0 4 8 12 16 20

WT

Cry1∆e

0.0

2.0

4.0

6.0

ZT 0 4 8 12 16

D. Mermet Supplemental Figure 6

WT  Cry1∆e

Series 1

Series 2

Series 3

E. CRY1 Western blot

Relative Abundance

0 4 8 12 16 20

ZT

p=0.02

F. Liver Phase Diff WT – Cry1∆e [h]

Kidney Phase Diff WT – Cry1∆e [h]

-2 -1 0 1 2
Supplemental Figure 6: Deleting *Cry1* intronic enhancer *in vivo* shortens phase advances the clock and clock-controlled genes.

(A) PCA analysis of temporal RNA-seq data in the liver (left panel) and the kidney (right panel) of *Cry1Δe* and WT littermates. Right: color code for ZT time.

(B) *Cry1* mRNA profile (in Transcripts Per Million, TPM) in liver (black) and kidney (red) of *Cry1Δe* (dashed line) and WT (solid line) littermates. Error bars: SD of 3 animals. At ZT20, *Cry1* mRNA levels are significantly different between WT versus *Cry1Δe* for liver and kidney (respectively p=0.045 and p=0.0037, t-test). *Cry1* mRNA is phase advanced in *Cry1Δe* versus WT (p=0.011 and p=0.0047 for liver and kidney, respectively, bootstrap test).

(C) *Cry1* intronic reads around the clock for WT and *Cry1Δe* mice. The oscillations in WT versus *Cry1Δe* are significantly different (p=0.0063, F-test test) (D,E) Temporal expression of CRY1 protein in the liver cytoplasm of WT and *Cry1Δe* littermates. Western blot (D) and quantification of CRY1 relative abundance normalized by the Naptho blue black coloration of the membranes (E). CRY1 shows a decreased protein abundance in *Cry1Δe* (p=0.02, F-test). (F) Differential phase between *Cry1Δe* versus WT in liver and kidney. Genes selected for core-clock and clock-controlled transcripts (p<0.01, harmonic regression in both liver and kidney) based on Gene Ontology (GO) annotation indicated in table S5.
Supplemental Figure 7: 4C-seq around the clock at Gys2 TSS and at Gys2 exon8 demonstrates rhythmic chromatin topology.

(A-B) 4C-seq signals (LWMR summarizing n=3 animals per condition) from Gys2 TSS (A) and exon8 (B) baits across time. Right: 4C signal over time at oscillatory chromatin contacts. (C) Statistical significance of rhythmic amplitudes of rhythmic contacts (p < 10^-7 for TSS and exon8, chi-squared test). Fragments with p<0.01 are colored by their time of peak contact frequency according to color legend (right). (D) 4C-seq around the clock on negative control region, Hoxd4, in WT animals shows static chromatin topology.
Supplemental Figure 8: Deleting the Cry1 intronic enhancer disrupts rhythmic chromatin topology. (A) 4C-seq signals (LWMR summarizing n=3 animals per condition) from Cry1 TSS bait across time in WT (top panel) and Cry1Δe livers (middle panel). (B) Statistical significance of amplitudes of rhythmic contacts (chi-squared test, bottom panel) for Cry1 TSS and Cry1Δe baits. Fragments with p<0.01 are colored by their time of peak contact frequency according to color legend (right). (C) Idem as (A) targeting the RRE bait -7 kb upstream Cry1 TSS (Cry1 upstream) (D) Idem as (B) for Cry1 upstream. Vertical lines show rhythmic promoter-enhancer interactions that are disrupted in Cry1Δe liver. Vertical arrows show location of the bait relative to the Cry1 TSS.
Supplemental Figure 9: smRNA-FISH against Cry1 pre-mRNA in WT versus Cry1Δe livers. (A) Size distribution of nuclei for a representative animal. Colored curves show fitted Gaussian mixture model corresponding to populations with 2N, 4N, 8N ploidy. (B) Fraction of nuclei for each animal (n=2 per condition) assigned to different ploidy. (C) Number of active transcription sites (TSs) averaged per animal increases with ploidy. At ZT20, Cry1Δe animals show reduced number of TSs compared to WT: lines show mixed effect linear model with genotype-dependent slopes (p(H0:equal slopes)=0.00014, F-test); at ZT08, the slopes are not different (p=0.84). (D) Active TS intensity averaged per animal shows comparable intensity across ploidy and conditions: lines show mixed effect model with genotype-dependent intercepts, intercept comparisons at both ZT08 (p(H0:equal intercept)=0.53, F-test) and ZT20 (p=0.41) are not significant.
**Supplemental Table 1: Distribution of 4C-seq counts on the genome.** Sheet 1: Distribution of 4C-seq raw counts for each bait and each biological replicate in the genome. Exponent of the power-law fit is indicated. Sheet 2: $\sigma_{\text{min}}$ used to regularize the residual variance for each bait.

**Supplemental Table 2: Sequences of CRISPR-Cas9 RNA guides.** Sequences of CRISPR-Cas9 RNA guides targeting upstream and downstream the Cry1 intronic enhancer element. The sequence of the PCR primers used to screen the deletion is indicated.

**Supplemental Table 3: 4C-seq primers.** Sequence of the PCR primers used for the inverse PCR step during 4C-seq libraries preparation.

**Supplemental Table 4: Motif counts.** List of motif counts for TF motifs from SwissRegulon. TFBS site count probabilities calculated using Motevo. Genomic coordinates from mm9 annotations.

**Supplemental Table 5: GO term annotation.** List of GO term annotation used to analyze temporal RNA-seq data in the liver and kidney of Cry1Δe and WT littermates.
Chapter 3. Clock-dependent chromatin topology modulates circadian transcription and behavior

3.3 Conclusion and perspectives

This study establishes dynamic chromatin interactions as a novel regulatory layer underlying circadian gene expression and behavior. By performing 4C-seq around the 24-hour day, we find that the proper timing of Cry1 gene expression depends on a distal enhancer located in the first intron. We establish the function of this distal enhancer by deleting the Cry1 enhancer. Looking at single-cell transcription with smRNA FISH, we link this deletion with changes in transcriptional bursting, specifically the burst frequency but not burst size. Finally, we find a behavioral phenotype from mice with the deleted enhancer; the period of the circadian locomotor activity was decreased by 15 minutes. This 15 minutes is expected because deleting the entire Cry1 gene shortens locomotor activity by approximately 1.2 hours (Ko and Takahashi, 2006).

1. How are the chromatin conformation data relate to what is happening at the single cells? Does transcription occur immediately after a promoter-enhancer contact?

2. How do we incorporate publicly available Hi-C datasets onto our study to predict more loci with circadian dynamics?
4 Dynamic gene expression and regulation from circadian and sleep-wake processes in the mouse cortex

4.1 Introduction

This collaborative project with the Franken lab at the University of Lausanne investigates the gene expression dynamics in response to acute sleep deprivation. The computational analysis of RNA-seq profiles also use model selection, as in Yeung et al. 2018 but the models go beyond simply periodic functions. We use the EEG data to predict sleep-wake processes. We use a model selection framework to distinguish between sleep-wake and cosine dynamics as well as interactions between the two. Surprisingly, we find that core clock genes are sensitive to sleep deprivation; their amplitudes become damped for at least 48 hours after sleep deprivation treatment. Overall, our results highlight surprising dynamics underlying how the circadian clock embeds into gene regulatory networks regulating sleep/wakefulness.

*This chapter is a manuscript in preparation and should be kept confidential.*

4.2 Contributions

Charlotte Hor conceived the project, performed the RNA-seq, ATAC-seq, and EEG experiments, with help from Yann Emmenegger and Jeffrey Hubbard. I developed the theoretical and computational framework for analyzing the temporal dataset. I analyzed the RNA-seq data, with help from Charlotte Hor and Maxime Jan. Maxime Jan analyzed the ATAC-seq data over time to identify genomic loci with dynamic ATAC-seq signal. I analyzed the ATAC-seq data to identify transcription factor binding motifs linked to gene expression dynamics.
Dynamic gene expression and regulatory outputs of circadian and sleep homeostat processes in the mouse cortex

Charlotte N. Hor*1, Jake Yeung*2,3, Yann Emmenegger1, Maxime Jan1,3, Jeffrey Hubbard1, Felix Naef§2,3, Paul Franken§1

1Centre for Integrative Genomics, University of Lausanne, Lausanne, Switzerland
2Institute of Bioengineering, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
3Swiss Institute of Bioinformatics, Lausanne, Switzerland

* These authors contributed equally to this work.
§ These authors contributed equally to this work.

Abstract

The timing and quality of sleep are thought to be regulated by the interaction of two processes. First, a sleep homeostat accumulates pressure for sleep during wake, and is relieved during sleep; second, the circadian timekeeping mechanism sets preferential times for sleep intervals. Although studies have implicated sleep-wake driven and circadian clock genes in the regulation of sleep, a systematic survey of how sleep and time-of-day govern gene expression dynamics and its underlying transcriptional regulation has not been established. Here, we sleep-deprived mice for 6 hours and tracked sleep-wake history, gene expression, and chromatin accessibility over more than 48 hours. Integrating the electroencephalography (EEG) data with RNA-seq and ATAC-seq, we developed a model selection framework to systematically infer gene expression dynamics that are driven by sleep-wake, time-of-day, or interactions of the two processes. From our model selection, we found that sleep-wake history, as measured by EEG, explained the largest fraction of variance, followed by rhythmic gene expression that was unaffected by sleep deprivation (SD). Remarkably, we found that the majority of circadian clock genes oscillated with damped amplitudes following SD, suggesting that circadian gene expression can adapt to perturbations. This damping was sustained for more than 48 hours after exposure. Sleep-wake driven dynamics in gene expression were accompanied by dynamics in chromatin accessibility sites. Open chromatin regions of immediate early genes were enriched for serum response factor (SRF) binding motifs. Furthermore, the predicted dynamics in SRF motif activity explained the upregulation of expression for many immediate early genes during SD. Our findings indicate that sleep-wake history and time-of-day interact to output diverse transcriptional dynamics in response to sleep deprivation.
Introduction

Sleep regulation is classically viewed as the interaction of sleep homeostatic process and circadian process (Paul Franken 2013). The sleep homeostat increases pressure for sleep during wake and releases pressure during sleep. This sleep need can be monitored by electroencephalogram (EEG)-derived variable delta power (Mongrain et al. 2010). Although the timing at which sleep preferentially occurs is modulated by the circadian process, this time-of-day modulation does not affect the need for sleep. Arrhythmic mice with lesioned suprachiasmatic nuclei (SCN) still maintain the quantitative relationship between sleep-wake distribution and EEG delta power during NREM sleep (Trachsel et al. 1992).

The circadian clock is encoded through negative feedback loops that generate robust oscillations even in constant conditions (Takahashi 2017). This ubiquitous oscillator is involved in a variety of physiological processes such as metabolism, blood pressure, wound healing, and locomotor activity (Lamia et al. 2011; Zuber et al. 2009; Hoyle et al. 2017; Mermet et al. 2018). Surprisingly, although lesioning the SCN of mice did not affect homeostatic sleep need, Cry1/Cry2 double knockout mice, which are also arrhythmic, showed increased time in NREM sleep, suggesting the molecular clock also regulates sleep (Wisor et al. 2002). Thus, both the circadian and sleep-homeostatic processes both contribute to sleep timing and sleep-wake distribution.

Changes in gene expression can be dependent or independent of what time of day the sleep perturbation took place (Maret et al. 2007; T Curie et al. 2013), which is consistent with the two-process model of sleep homeostasis (Paul Franken 2013). Under this model, the homeostatic process (Process S) tracks the time spent awake and thus the sleep need, while the circadian process (also referred to as Process C) modulates the timing of sleep with regards to time of day. Therefore, under normal conditions where environmental changes occur periodically and predictably, the expression of genes displaying a nycthemeral (i.e., rhythm with a period of 24-hours) rhythm can be regulated by either of these processes or both in combination.
Sleep deprivation (SD) experiments allow to uncouple/disentangle the sleep homeostat from circadian processes (P. Franken and Dijk 2009). In fact, SD beginning at the dark-to-light transition will cause mice to stay awake during a time when they normally sleep. Thus, sleep-wake driven genes will respond acutely to SD whereas dynamics that are robust to SD will be comparable to control mice. Studies comparing gene expression levels during SD with controls have identified many differentially expressed genes (Maret et al. 2007; Mongrain et al. 2010; Paul Franken et al. 2007; T Curie et al. 2013; Diessler et al. 2018). However, assessing the many possible dynamics requires comparing beyond time-matched control conditions. Sampling during and after SD and analyzing the entire time course can reveal rich dynamic processes. Thus, identifying the regulatory contributions from the sleep homeostat and circadian clock require considering gene expression dynamics along a sufficiently long time-course.

Here, we measured gene expression and chromatin accessibility in adult C57BL6/J mice before, during, and over 48 hours following one 6-hour session of total sleep deprivation (SD), as well as 7 days after the intervention. This time course can then be compared with a baseline control day. We also report a long-term electroencephalographic (EEG) recording of vigilance states and correlates of sleep need over the same time interval.

Our analysis integrates EEG, transcriptome, and chromatin accessibility data over 78 hours sampled at least every 6 hours. Analyzing the entire time series using both unsupervised and supervised methods, we find that the largest group of gene expression dynamics come from a sleep-wake driven process, which could be modeled from the EEG data. We also find unexpected diversity of gene expression dynamics. Notably, the (peak to trough) amplitudes of most core clock gene expressions, such as *Arntl*, *Nr1d1*, and *Nr1d2*, were prolongedly damped after SD, suggesting that SD can have long-term effects on gene expression dynamics in the cortex beyond the initial treatment. Our time course data also reveal interactions between the sleep homeostat and time-of-day processes. Combining both the sleep homeostat and time-of-day allows genes to respond with larger fold changes relative to baseline compared to genes that are only sleep-wake driven. These gene expression dynamics are accompanied by dynamics in chromatin accessibility. Finally, we find sleep-wake driven activity of serum response factor (SRF) underlies the gene expression dynamics of many
immediate early genes previously implicated in sleep response, such as Arc, Egr2, Fos, and Junb.

**Materials and Methods**

**Animals**
C57BL/6J male mice were purchased from Charles River France (Lyon, France) and allowed to habituate to our sleep study facility for 2-4 weeks prior to habituation to the experimental setting, and the experimentation. Animals were kept in accordance to the Swiss Animal Protection Act, and all experimental procedures were approved by the local veterinary authorities (Authorization nb. VD3037).

**Surgery and EEG recording**
The EEG cohort consisted of 6 male C57BL/6J mice 10-12 weeks at the time of SD. Surgical implantation of electrodes, EEG recording and data collection were performed according to our standard procedure (Mang and Franken 2012). EEG was recorded from 2 days prior to SD (which were averaged to constitute a 24-hour baseline) until 7 days after SD. Electrophysiological signals were captured at 2000Hz, transformed from analog to digital, and downsampled and stored at 200Hz (EMBLA A10 and Somnologica-3; Medcare Flaga; Thornton). Sleep and wake states were annotated according to established criteria based on the properties of the EEG and EMG signals (Mang and Franken 2012). To determine spectral composition, EEG signals (0 to 90 Hz) underwent a discrete Fourier transformation, using a window of 4-seconds (Hamming function), to determine power spectral density. Delta power (1-4Hz) was extracted for NREM epochs and averaged across the experiment. To counteract differences in absolute EEG power between individuals, power spectral density for each 0.25 Hz bin was expressed as a percentage of the 2 baseline days from ZT8-12, when NREM delta power is lowest. SD and recovery time points were compared to baseline by means of a t-test.

**Sleep deprivation and tissue collection**
Mice for tissue collection were divided into two experimental cohorts, sleep deprived (SD) and non-sleep deprived (controls, or Ctr). After a one-week habituation to the experimental setting, at the age of 11-12 weeks, the SD mice were sleep-deprived by gentle handling for 6 hours starting at light onset (Zeitgeber time ZT0-ZT6) as
described in (Mang and Franken 2012), and allowed to recover according to the tissue collection schedule. Mice were anesthetised with isoflurane prior to decapitation. Cortex was rapidly dissected and flash frozen in liquid nitrogen. NSD mice were sacrificed at ZT0, ZT3, ZT6, ZT12, ZT18 of the first day of experimentation (samples T0-T18), serving as a baseline day (BL). SD mice were sacrificed at the same time of day on Recovery Day 1 (R1, samples T24-T42), at ZT0, ZT6, ZT12, ZT18 on Recovery Day 2 (R2, samples T48-66), as well as ZT0 and ZT6 on the 3rd day (samples T72-78), see Figure 1A. Finally, two groups of mice were allowed to recover for 7 days after SD, before being sacrificed at ZT0 and ZT6 (samples T192-198). We refer to each time point in hours from the start of the baseline day (T0), with sleep deprivation occurring from T24 to T30. In the analysis, samples at T0 and T24 were assigned evenly across two different batch runs of ZT0 NSD. We collected 3-4 replicates per time point and condition, and 8 replicates of ZT0 NSD.

Tissue processing and sequencing library preparation

Frozen cortex of each individual was ground in liquid nitrogen and stored at -80°C until further use. Total RNA was extracted using the miRNeasy kit (Qiagen; Hilden, Germany) following the manufacturer's instructions. RNA-seq libraries were prepared using 1000 ng of total RNA and the Illumina TruSeq Stranded mRNA reagents (Illumina; San Diego, CA, USA) on a Sciclone liquid handling robot (PerkinElmer; Waltham, MA, USA) using a PerkinElmer-developed automated script. Libraries were sequenced on the Illumina HiSeq 2500 sequencer, producing >36 million (median 55 million) single-end 100 bp reads. ATAC-seq was performed with minor modifications from (Jason D Buenrostro et al. 2015). 100'000 nuclei were treated with 2.5 µl Tagment DNA enzyme (Nextera DNA Sample Preparation Kit, Illumina) in transposition buffer (10mM Tris Base, 5mM MgCl₂, 10% DMSO, pH 7.6, adapted from (Wang et al. 2013)) at 37°C for 30 minutes, followed by cleanup on a Qiagen Minelute column. Fragments >1kb in size were removed using AmpureXP beads (Beckman Coulter Life Sciences; Indianapolis, IN, USA) using 0.6X and 1X volumes. DNA fragments were subjected to 11 cycles of PCR amplification with Nextera index primers (Illumina) and NEBNext High Fidelity 2X PCR Master Mix (New England Biolabs; Ipswich, MA, USA). PCR reactions were cleaned up with one volume AmpureXP beads, quantified by Qubit (ThermoFisher Scientific; Waltham, MA, USA) and quality controlled by Fragment Analyzer.
Advancing Analytical Technologies; Ankeny, IA, USA). Libraries were sequenced on the Illumina HiSeq 2500 sequencer, producing >25 million (median 41 million) 50 bp paired-end reads per sample after removal of duplicate and mitochondrial sequences. Due to sequencing failure, the ATAC-seq data of two out of three replicates of T66 were excluded from the analysis.

**Sequencing data analysis**

Transcript abundance was quantified by *kallisto* version 0.43.0 (Bray et al. 2016) using the GRCm38 reference transcriptome (mm10) and the parameters `--single -l 100 -s 20 -b 100`. The abundances were further processed using *sleuth* version 0.29.0 (Pimentel et al. 2017), starting with merging transcript abundance into gene counts. We applied a detection cutoff of 5.5 on the mean gene counts across samples in the time series, yielding a set of 13'842 expressed genes which were used for further analysis. Batch effects were corrected by ComBat (R package *sva_* version 3.25.4 (Leek et al. 2012)).

ATAC-seq reads were aligned to the mouse genome (mm10) using bowtie2 (Langmead and Salzberg 2012) in paired-end mode, with the parameters recommended for open chromatin (`--very-sensitive --maxins 2000 --no-mixed --no-discordant`). Duplicate sequences were removed using *samtools* rmdup (Li et al. 2009).

**Differential gene expression**

To identify genes displaying a statistically significant effect over time, we used a likelihood ratio test implemented by *sleuth* version 0.29.0 (Pimentel et al. 2017), comparing a full model of time plus batch effects with a null model (no time effect) plus batch effects. We used FDR-adjusted p-value of 0.001 to identify 3461 statistically significant genes, which were used in the clustering and modeling analysis.

**mRNA profile clusters**

To uncover temporal patterns of mRNA abundance, we performed k-means clustering on statistically significant temporal gene expression (3461 genes). For a range of number of clusters, *k*, we calculated the within cluster variation as the sum of the Euclidean distance between data points and their assigned cluster centroids and
empirically chose $k=10$ as a balance between variance explained and generalizability of each cluster. Mean $-\log_{10}(p\text{-value})$ and effect sizes across genes in each cluster are represented as shaded rectangles above each plotted cluster, $p\text{-values}$ are calculated from likelihood ratio test.

**mRNA time course analysis**

We used a model selection approach to classify temporal log mRNA abundance $m(t)$ of expressed genes into the scenarios described in Results. The models can be expressed as stated below. For models 3, 5 and 6, sleep-wake history was used to model the synthesis rate of mRNA as the S process (P Franken, Chollet, and Tafti 2001) using EEG data from $n=15$ C57BL/6J mice (this study and (Diessler et al. 2018)).

1: Flat model with constant $\mu$ and noise $\epsilon$ (flat)

$$m(t) = \mu + \epsilon$$

2: Sinusoidal, oscillatory model with 24 h periodic rhythmic parameters $a$ and $b$

$$m(t) = \mu + a \cos(\omega t) + b \sin(\omega t) + \epsilon \text{ (Cosine)}$$

Angular frequency $\omega = 2\pi/24 \text{ h}^{-1}$

3: Sleep-wake driven model (S)

$$\frac{dm(t)}{dt} = S(t, U, L, \tau_w, \tau_s) - \gamma_{mRNA}(m(t)),$$

where $S$ is defined recursively:

$$S(t, U, L, \tau_w, \tau_s) = \begin{cases} U - (U - S(t - 1) \exp(-\Delta t/\tau_w)) & \text{if awake} \\ L + (S(t - 1) - L) \exp(-\Delta t/\tau_s) & \text{if sleep} \end{cases}$$

$S(t_0) = S_0$ : initial value at time $t=0$

$\Delta t$: mean period of continuous wake or sleep, defined by EEG data from 15 mice

$U$: asymptotic value for long periods of wake

$L$: asymptotic value for long periods of sleep

$\gamma_{mRNA}$: inferred degradation rate of mRNA. Slow degradation rates damp fluctuations from EEG data, while fast degradation rates follow the fluctuations.

We solved the differential equation for $m(t)$ using the Euler method with a time step of 0.1 h.
We will call the solution of this differential equation $D(t, \tilde{\theta}_{\text{sleep}})$ where $\tilde{\theta}_{\text{sleep}}$ are the sleep parameters, $U, L, \tau_w, \tau_s$.

4: Sinusoidal model with change in amplitude (A).

$$m(t) = \mu + C(t)(a \cos(\omega t) + b \sin(\omega t)) + \epsilon, \quad C(t) = \begin{cases} 1 & \text{for } t \leq 33h \\ c & \text{for } t > 33h \end{cases}$$

where $t=33$ h corresponds to 3 h after the end of sleep deprivation. Thus, in this model the amplitude is changed by a factor $c$ after $t = 33$ h.

5: Sleep-wake and oscillatory model (S+A)

$$m(t) = D(t, \tilde{\theta}_{\text{sleep}}) + a \cos(\omega t) + b \sin(\omega t) + \epsilon$$

where $D(t, \tilde{\theta}_{\text{sleep}})$ is the solution to the differential equation in the sleep model.

6: Combined with change in amplitude model (S+A)

$$m(t) = D(t, \tilde{\theta}_{\text{sleep}}) + C(t)(a \cos(\omega t) + b \sin(\omega t)) + \epsilon$$

where D and C are defined as above.

7: Model with one parameter for each sampled time point (this is the most complex model, termed generic).

$$m(t) = \beta(t) + \epsilon; \quad t \in \{0, 3, 6, 12, 18, 24, 27, 30, 36, 42, 48, 54, 60, 66, 72, 78\}$$

For models that are nonlinear (models 3-6) with respect to the parameters, we fitted the model with optim in R using the L-BFGS-B method. Linear models (models 1, 2, and 7) were solved using the lm() function in R. The gene expression was fit in the log scale.

For each gene, we estimated the posterior probability of each model by first calculating the Bayesian Information Criterion (BIC) scores:

$$B_i = -2 * L_i + k_i \log(n)$$

where $L$ is the log likelihood. A better fit will improve (decrease) the BIC, while a more complex model will penalize (increase) the BIC. Intuitively, an optimal model will fit the data while not using an excessive number of parameters. We assume the model
errors are independent and identically distributed following a Gaussian distribution with variance estimated from the fits:

\[ \sigma^2 = \frac{1}{n} \sum_i (m_i - \bar{m}_i)^2 \]

Exponentiating the BIC scores yields Schwarz weights \( w_i \):

\[ P(M_i|D) \approx w_i = \frac{\exp(-B_i/2)}{\sum_j \exp(-B_j/2)} \]

and we then assigned each gene to the model \( i \) corresponding to the largest \( w_i \).

\( w_i \) assigns a probability to each model, and this probability measurement takes into account the number of parameters \( k \) in the model through the BIC score (i.e. complex models with large \( k \) are penalized by having a larger \( B \), which would have smaller \( w \)).

For comparison with previously published sleep-related gene lists, we opted to remove close calls between models by selecting the genes with a weight \( w_i \) in any model of at least 0.7, which yielded a set of 9620 confidently assigned genes.

**ATAC-seq peak detection:**
ATAC-seq data files were processed before peak calling as follows. Alignment files were converted into bed files and tags were extracted using bedtools version 2.26.0. Each tag position was shifted from +4 base pairs on the positive strand and -5 base pairs on the negative strand to center tags on transposase binding events as suggested by (J D Buenrostro et al. 2013). The peak calling was performed on pooled tags for replicates using Macs2 version 2.1.1 (Zhang et al. 2008) [--nomodel --shift -75 --extsize 150], and peaks were filtered using a 0.05 FDR cutoff. Peak boundaries were merged between time points and conditions in order to build a common peak mapping reference among all samples, encompassing a total of 215'045 peaks. Finally, peak coverage was quantified using HTSeq version 0.6.1 for each sample using the
common mapping reference. A region was considered positive if it passed a $q=0.05$ FDR threshold vs. random noise. As a quality control, we probed whether genes within accessible regions were enriched in cortex/brain tissue. To this end, we used the Bgee database and topAnat (Komljenovic et al. 2016) to look for significant enrichment, and found that the top 20 enriched tissues were all nervous system structures (FDR p-value < 10e-8). The proximity in the PCA of the two technical replicates at T24 attests the reproducibility of ATAC-seq over different batches of sequencing.

**ATAC-seq differential accessibility analysis:**
To identify peaks with differential accessibility, we first normalized count data using a TMM normalization, applied a 10 read count threshold, and used a likelihood ratio test implemented in edgeR. We compared chromatin accessibility of sleep deprivation samples (T24-198) with the corresponding ZT during baseline (T0-18, see Figure 1A). Thus, for differential accessibility at ZT3, we compare T27 with T3, at ZT6, T30 and T6, etc. P-values were adjusted using the Benjamini & Hochberg (FDR) method (Benjamini and Hochberg 1995). K-means clustering was performed using $k=10$ clusters.

**Genomic distribution of ATAC-seq peaks:**
The annotation of the detected ATAC-seq peaks was performed using PAVIS with the `Ensembl_GRCm38/mm10 all genes` reference annotation (Huang et al. 2013).

**Peak-to-gene expression association:**
To associate gene expression dynamics with chromatin accessibility dynamics, we used a pearson correlation coefficient across the samples and limited the possible association test within previously defined topological interaction domains (TADs), which were computed from cortex tissue by (Dixon et al. 2012). TAD boundaries position were originally detected using the mm9 reference genome, so we converted them to mm10 using CrossMap 0.2.6 (Zhao et al. 2014). For association statistics, we used a strategy similar to that implemented within FastQTL (Ongen et al. 2016). Each pair consisting of a peak and a gene within the same TAD were associated using the pearson correlation coefficient. For each gene, only the top correlation to a peak was kept. To control for multiple associations within a TAD and adjust nominal p-values, we used 1000 permutations per gene and modelled the null distribution fitting a beta distribution.
parameters were estimated using a maximum likelihood approach (R/MASS::fitdistr). Finally, a genome-wide p-value adjustment was computed using a qvalue procedure (R/qvalue). Of the 10894 genes mapping within a TAD, 3386 were associated to an ATAC-seq peak within the same TAD.

Predictions of transcription factor binding site (TFBS) activity in promoters
To predict the activity of TFBSs, we used position weight matrices of 179 mouse transcription factors (TFs) defined by SwissRegulon on mm9 (http://swissregulon.unibas.ch/fcgi/sr/downloads). For each of the 179 position weight matrices, we scanned genomic regions (500 bp windows) 15 kb from promoters using MotEvo (Arnold et al. 2012). We filtered regions containing ATAC-seq counts greater than 0.1 RPM (reads per million mapped reads). The sitecount matrix of each motif was scaled across genes so that ranges in sitecounts were comparable across motifs. We inferred TF activity using the TF binding site predictions and the temporal mRNA abundance. To infer the TF activity, we applied a penalized regression model as previously described (Balwierz et al. 2014; Yeung et al. 2017) using an L2 norm penalty for regularization (ridge regression). Prior to the regression, we mean-centered the input matrix of temporal mRNA abundances, standardized the columns of the sitecount matrix (each motif across genes), and excluded genes that were assigned to the flat model.

Motif search
We used MEME-ChIP of the MEME suite (www.meme-suite.org) (Bailey et al. 2009) to search for motifs within differentially active regions.

Results
Study design
We subjected mice to 6 hours of total sleep deprivation (SD), starting at light onset of the 12-hour light-dark cycle and collected cerebral cortex during SD and over the following 48 hours. We used Zeitgeber time (ZT) to indicate time since light onset. ZT0 indicates light on while ZT12 indicates lights on. Mice are nocturnal and tend to be active during dark (ZT12 to ZT0/ZT24) and sleep during light (ZT0 to ZT12). A second group of mice was allowed to sleep ad libitum, undisturbed in another room, and
sacrificed at the same times of day as SD mice, serving as non-sleep deprived controls or baseline day (Figure 1A). We profiled gene expression by RNA sequencing of polyadenylated transcripts (RNAseq) and accessible portions of the genome by ATAC-seq (J D Buenrostro et al. 2013). To characterize the behavioural response and recovery after SD, we recorded EEG data in a separate group of 6 mice over 9 days, including 2 days of baseline prior to SD, the day of SD and 6 days of recovery.

**Behavioral response and recovery after SD**

To assess the long-term dynamics of the sleep phenotype, we quantified non-rapid eye movement sleep (NREMS) in 6-hour intervals, as well as hourly EEG delta power, a well-known correlate of sleep need (P Franken, Chollet, and Tafti 2001). We observed the typical distribution of sleep over 24 hours in baseline, with mice spending most of the time of the light period asleep, while being predominantly awake during the dark period (Figure 1B, bottom). Delta power in NREMS (Figure 1B, top) followed the amount of sleep pressure, high after spontaneous waking in the dark phase and low during the light phase. We also observed the well-known effects of sleep deprivation. NREMS rebounded during the 12 hours following SD and an increase in delta power immediately after the end of SD. We found that values stopped differing from baseline already during the second half of the dark phase after SD for NREMS, and as of Recovery day 2 for delta power (Figure 1B). REM sleep was affected in the same manner as NREMS (not shown).

**Principal component analysis reveals sleep-wake state as main driver of transcriptome dynamics**

We asked whether the fast recovery of the phenotype in the EEG data could be observed at the gene expression level, and whether novel dynamic patterns beyond the EEG dynamics could be observed. We therefore analyzed the temporal dynamics of transcriptomes.

We first examined the detected fraction of our transcriptome dataset (13'842 genes) using principal component analysis (PCA, Figure 1C). We observed that samples formed three groups along the first principal component (PC1) axis. The right-most group clustered time points during the light phase where mice generally spend more time asleep, while the middle group represents time points taken during the dark phase of the LD cycle where mice are predominantly awake. Surprisingly, ZT3 and ZT6
during SD (T27 and T30) separate far from their time-matched baseline (T3 and T6) towards the awake group, suggesting that PC1 dynamics follow sleep-wake history rather than Zeitgeber time.

To illustrate the sleep-wake-driven dynamics underlying PC1, we overlaid PC1 with the average EEG signal over time (Figure 1D). This integration showed that sleep-wake history explains PC1. During periods of wake, PC1 increases, while during periods of sleep, PC1 decreases. Importantly, the six-hour sleep deprivation period, PC1 reaches its maximum. PC1 thus reflects the amount of sleep prior to sample collection, and illustrates the impact of sleep-wake distribution on mRNA expression.

**No significant differences in gene expression 7 days after exposure**

We next examined the temporal dynamics of gene expression. We noted that expression levels on Day 7 (T192 and T198) were not significantly different from baseline at T0 and T6 according to our differential gene expression analysis (FDR adjusted pval > 0.05, Supplemental Figure 1A). We therefore focused our analysis and modeling to the baseline day 0 and recovery days 1 and 2 (T0 to T78).

**Clustering analysis reveals dynamics that are driven by sleep-wake history and also robust to sleep-wake history**

To uncover general temporal patterns in our data, we performed an exploratory analysis using k-means clustering. We clustered temporal expression of 3461 genes identified as displaying statistically significant temporal variation from T0 to T78 (FDR-adjusted p-value < 0.001, Methods).

We observed distinct dynamics over 78 hours (Figure 2). Genes in clusters 1-6 displayed an immediate response to SD, many of which showed statistically significant differences at T27 and T30 (p-values across genes in each cluster are summarized for in Figure 2). In cluster 7, the response is longer, many genes showing differences up to T36. Cluster 8 showed a delayed response; the largest differences occur 18 hours after the end of SD.

Clusters 9 and 10 showed prominent 24-hour rhythm and, on average, the genes are not perturbed by SD (mean pvalues across genes > 0.24). However, comparing individual
genes at days 1 and 2 against baseline day 0 found damped rhythms after SD such as Arntl, Fabp7, Obscn, and Lfng. (Supplemental Figure 2).

Concerning the recovery properties, clusters 1-4 showed a fast recovery, at T36 the expression on average has returned to baseline. By contrast, genes in clusters 5 and 6 reverted more slowly on average, returning to baseline at T42. Genes in cluster 7 also displayed slow recovery, returning back to baseline levels 12 hours after peak response. Finally, in cluster 8 we observed a pattern compatible with a "rebound", in the form of an exacerbated increase at T48 following the initial downregulation at T36. By T54, cluster 8 has returned to baseline.

Generally, we observed that SD can have three effects on dynamics relative to baseline. First, SD can up or downregulate expression of genes that are normally down or upregulated during ZT0-6 in baseline, respectively. This effect explains a significant variance (20%) of the transcriptome, as shown in our PCA analysis (cluster 1-6). Second, this down or upregulation can extend to ZT12 (cluster 7 and 8), suggesting slower dynamics or downstream effects from the first effect. Third, SD can dampen 24 h periodic oscillations in mRNA abundance.

Overview of identified models
Explicitly modeling the temporal dynamics of mRNA profiles can offer advantages over unsupervised methods. From our downstream analysis of clusters, we devised 6 models to explain the log gene expression dynamics of 13842 detected genes (Supplemental Figure 3): (1) constant flat model; (2) sleep-wake history modeled from the EEG data (S); (3) cosine dynamics with 24-hour rhythm (C); (3) cosine with amplitude change after SD (A); (5) sleep-wake + cosine (S+C); (6) sleep-wake + cosine with amplitude change (S+A). The parameters of the model can give useful insights to the underlying dynamics, competing models can be systematically compared, and explicit hypotheses can be tested. For example, using a sleep-wake model will unify dynamics that appear in separate clusters (e.g., Cluster 1 and 4 both may have genes that are driven by sleep-wake). The C and A models can separate dynamics that appear in the same cluster (e.g., Cluster 9 contains both SD-robust and SD-sensitive dynamics).

To select competing models, we used the Bayesian Information Criterion (BIC) to balance model fit and model complexity (Methods). Of note, we also included a generic
model where the gene expression is modeled as the mean expression across replicates at each time point to assess the possibility of more complex dynamics not explained by any of the models. We found that the BIC weight $w$ was always lower for the generic model than the other 6 models, meaning no genes were selected in the generic model, and therefore not included in subsequent analyses.

For each gene, we fit temporal gene expression to 6 models (examples shown in Figure 3A). We summarized the genes assigned to each model genome-wide (Figure 3B). We found that, out of all temporal models (i.e., excluding flat model, Figure 3C), sleep-wake driven model had the most genes assigned to it (Model S Figure 3D). This large number of genes is consistent with our PCA and clustering analysis, where sleep-wake driven was also observed in the first component in PCA (Figure 1C) and in clusters 1-6 (Figure 2) in clustering analysis.

The cosine model contained the second highest number of genes (Model C, Figure 3E), which is consistent with the clustering analysis (cluster 9 Figure 2). We found Hif3a, hypoxia-inducible factor 3, as a prominent example of robust oscillations despite SD. This robust model suggests a significant fraction of dynamic genes is unaffected by SD.

Interestingly, the cosine with amplitude change model had the third most genes, suggesting SD can affect dynamics beyond the 3-hour SD (Model A Figure 3B). For Nr1d1 (Figure 3A), the best model based on BIC (BIC weight $w$) was cosine with amplitude change, suggesting SD damped its amplitude (Figure 3A bold line). SD also damped the amplitude of Arntl, another core clock gene (Takahashi 2017). Surprisingly, we found that SD damped the amplitudes of many clock genes, suggesting that the circadian clock is adaptive to sleep-wake history.

Finally, our approach identified more complex models involving both a sleep-wake plus cosine effect (Model S+C Figure 3G) or a sleep-wake plus cosine with amplitude change (Model S+A Figure 3H). These novel dynamics suggest a significant number of genes have expression dynamics that integrate sleep-wake history with time-of-day.

**Model parameters predict novel transcriptional dynamics**
We found the overwhelming majority of known sleep-wake driven genes to be correctly assigned by our model selection method. 58 out of 61 genes previously described as sleep-wake driven (Mongrain et al. 2010) were fully or partially explained by the EEG data (Model S, S+C, S+A, Supplemental Figure 4A). Similarly, we looked at a list of 92 genes previously described as affected by SD (Maret et al. 2007), and found that 77/92 of the genes were also inferred to be affected by SD in our model select (Supplemental Figure 4B). These corroborations from independent datasets suggest that our model selection recapitulated known SD-affected genes.

Analyzing sleep-wake driven genes, we found that SD can upregulate or downregulate genes (Figure 4A) in the sleep-wake model. The largest fold changes (log2 FC > 2.5, or nearly six fold) were exclusively upregulated genes. The sleep model also predicted an mRNA half-life which buffers fluctuations from the EEG data and calculates the mRNA abundance (Figure 4A) (Methods). We found that large fold changes corresponded to short half-lives, consistent with fast dynamics (REF here? E.g. Wang 2018 or/and Zeisel MSB?). Finally, the fitted time constants describing the S process corresponding to wake tended to be longer than the one for sleep, as expected from mice spending more time awake than sleep (Figure 4B). Overall, the inferred parameters from sleep-wake driven genes were consistent with fast dynamics and relatively large fold changes, most of them upregulated during wake.

In contrast to the sleep-wake model, the cosine dynamics are robust to SD over 78 hours. Analysis of the fold change and peak times from the model found that large amplitude oscillations tended to occur near the dark-light or light-dark transition (Figure 4C). Notable genes such as Sgk1, glucocorticoid regulated kinase, and Cldn5, principal tight junction protein in blood-brain barrier, showed 24-hour oscillations in gene expression. We found that SD can alter amplitudes in rhythms, suggesting effects that last beyond the initial 3 hour SD. Notable examples with damped amplitudes include circadian clock genes Nr1d1, Nr1d2, Arntl, and Per3 (Figure 4D, Figure 4E). This model contained also genes with increased amplitudes after sleep deprivation such as Erbb3, Eva1b, Zfp473, and Akr1cl. In sum, SD can have long-term effects lasting at least 48 hours after sleep deprivation.
Finally, we found interactions between the cosine and sleep model (additive in the log scale). These combined models (S+C and S+A) in general allow genes to be upregulated or downregulated with significant fold changes compared to sleep model (Figure 4F left, Supplemental Figure 5A,B). This dynamic process can be coupled with decreased fold changes after SD (S+A model). Examples here include Per2, a gene shown to have complex interactions between the two processes (Thomas Curie et al. 2013), as well as clock output genes Dbp, Nfil3, and Bhlhe41 (Supplemental Figure 5C-F). The large proportion of core clock and clock output genes found to show interactions between sleep-wake history and time-of-day highlight how these two fundamental processes interact to generate complex temporal dynamics. On day 2, aggregate analysis for models S, S+C, and S+A, showed that the amplitudes are comparable between day 0 (Figure 4E right).

**Genome-wide analysis of ATAC-seq shows sleep-wake driven dynamics in chromatin accessibility**

We next asked which regulatory elements are involved in the response to SD, a compartment that has hitherto not been explored in this context. We identified a total of 215'045 ATAC-positive regions (read counts > 10 in at least one time point). Principal component analysis of accessibility levels in these regions separated samples by sleep-wake history (Figure 5A), consistent with the RNA-seq analysis (Figure 1C).

Accessibility peaks from all time points and conditions were mainly located in intronic or intergenic regions (Supplemental Figure 6A). When considering only peaks that were differentially active at specific time points (Supplemental Figure 6B-D), the proportion of intergenic regions was increased, suggesting that SD influences accessibility of distal rather than genic or proximal elements. Genes associated with DAS were enriched among models involving sleep-wake dynamics (Supplemental Figure 6E-J) compared to all peaks at all ZT (p-values < 2e-10, chi-square test) except ZT12 (p-value = 0.48).

To probe the general dynamics of chromatin accessibility, we performed a K-means clustering analysis (Figure 5B) and found sleep-wake driven dynamics. The strongest differential signal relative to baseline occurred during the 6h SD (T27 vs T3 and T30 vs T6). We examined the differential accessibility at each time point after SD compared
to its corresponding baseline day time. 1542 peaks were differentially accessible in SD at ZT3 (T27, after 3h SD) and 1906 at ZT6 (T30, end of 6h SD) and 678 at ZT12 (T36, after 6h of recovery, (Figure 5C). Chromatin accessibility signal tended to increase during sleep deprivation (ZT3 and ZT6). Overall, dynamics in chromatin accessibility was most pronounced during sleep deprivation, and returned to baseline by 12 hours after SD.

**Chromatin accessibility correlate with gene expression dynamics**

We correlated ATAC-seq peaks to genes by calculating the Pearson correlation across samples (Methods). We restricted the possible peaks associated with each gene within topologically associated domains (TADs) defined from Hi-C data generated from the mouse cortex (Dixon et al. 2012). This analysis identified chromatin interactions that correlated with gene expression dynamics (Figure 6A-F). For example, we found Hif3a, a gene relatively robust to SD, with a correlated ATAC-seq peak 2 kb from the promoter (Figure 6C). The promoter of Ciart, a gene that has been reported to be directly regulated by the clock (Anafi et al. 2014), had peak with correlated dynamics, which appeared to have damped amplitudes on the first day after SD and a partial amplitude recovery in the second day. We found that ATAC-seq dynamics in mouse cortex can show about a fold change of about 1.5, comparable with changes in the RNA-seq data. In sum, gene expression dynamics can be accompanied by dynamics in chromatin accessibility.

**Activity of Serum Response Factor underlies immediate early gene expression**

To infer which transcription factors drive the gene expression dynamics, we performed a motif activity analysis using the MARA (Balwierz et al. 2014). Briefly, we used the SwissRegulon database (179 mouse TF motifs) to search for TF motifs within ATAC-positive 500 bp bins within a distance of 15 kb of promoter of genes. We included genes that were assigned to any dynamic model (i.e., S, C, A, S+C, S+A). Using the sitecount matrix as covariates and the log gene expression as observed data, MARA uses a ridge regression framework to infer the temporal activity of each motif that could explain the gene expression data.

We found the SRF motif to be the most statistically significant candidate underlying the temporal gene expression dataset (Figure 7A). The inferred temporal activity of
SRF increased during sleep deprivation and rapidly returned to baseline, suggesting that its activity could be predicted by the sleep-wake history of the mouse (Figure 7B). Indeed, our model selection method on the SRF motif activity confirmed a sleep-wake-driven response. Furthermore, analysis of TF motifs near (<5 kb) promoters of sleep-wake-driven genes, such as Egr2, Fos, Egr1, and Arc, predicted SRF binding sites (Figure 7C). These SRF binding sites corresponded to open chromatin regions in the cortex. SRF ChIP-seq data in fibroblasts (C Esnault et al. 2014) showed binding of SRF to the SRF motifs present in those genes (Figure 7C, Supplemental Figure 7A). The ATAC-seq signal overlapping the SRF motif and ChIP-seq data is flat over time (Figure 7D, Supplemental Figure 7B), consistent with the regulation of many SRF target genes requiring cofactors (Cyril Esnault et al. 2017). SRF has previously been reported to regulate wake-dependent immediate early gene expression (Ramanan et al. 2005), consistent with SRF activity increasing during SD. In sum, our analysis suggests that SRF may underlie the immediate early response of many sleep-wake-driven transcripts.

Discussion

We characterized the dynamics of transcriptome and regulatory elements over time before, during and after a 6 hours of sleep deprivation. By integrating EEG data into the analysis, we developed a model selection approach to systematically identify modules of gene expression with distinct dynamics. This framework separated genes that could be explained by the EEG data (S), cosine dynamics (C), cosine with amplitude change (A), as well as interactions between sleep-wake and time-of-day (S+C and S+A). We attributed the largest gene expression variance to S dynamics, followed by C, A, S+C, and S+A.

The EEG data explained sleep-wake gene expression dynamics. These dynamics often show large fold changes in baseline. During SD, gene expression responds divergently relative to baseline and rapidly recover to baseline within 12 hours after SD. These dynamics reflect the sleep-wake behavior as determined by EEG. Remarkably, the first principal component could be explained by the EEG data, suggesting that sleep-wake behavior explains the largest variance in our dataset. We inferred SRF as a potential transcriptional regulator underlying dynamics of immediate early genes, which are upregulated during SD.
We also found that many SRF target genes had accessibility that was open but static, consistent with the role of cofactors in SRF target gene regulation (Cyril Esnault et al. 2017). Overall, we found that the largest fold changes (> 2.5) due to SD tended to come from SRF target genes (e.g., Arc, Egr2, Fos, Egr1, Junb, Nr4a1, Homer1), suggesting that SRF may play a role in sleep regulation. Our integrated analysis also found widespread chromatin accessibility dynamics that coupled with transcriptome, suggesting that sleep-wake history can alter transcription factor binding and mediate gene expression dynamics.

In a significant fraction of genes (734 genes), SD had a prolonged effect (>48 hours after SD) on mRNA levels. Remarkably, clock genes These long-term effects may eventually recover, but our simple cosine model with amplitude change after SD adequately explained expression of 734 genes, suggesting that these genes did not fully recover back to the baseline within 48 hours. Among the 734 genes are many clock genes, such as Nr1d1, Nr1d2, Cry2, and Arntl. Our results demonstrate that SD can disrupt oscillations in circadian clock expression by damping amplitudes, and this damping can be sustained long after exposure.

Finally, the S+C and S+A models combined EEG data with time-of-day highlight the interaction between sleep-wake history, sleep deprivation, and time-of-day. S+C allowed baseline fold changes to be modest and increase substantially during sleep deprivation.

The S+A model incorporated an amplitude change, often damping, following sleep deprivation. Many clock output genes such as Dbp, Tef, Nfil3, Bhlhe41 showed damped oscillations after SD. This model also included complex interactions between the two processes, exemplified by Per2 (Thomas Curie et al. 2013). Overall, we found the dynamics of most clock and clock output genes to be affected by sleep deprivation, many of them through damping of amplitudes. These damped amplitudes persist long after SD. Further studies will determine exactly how this disruption effect on other parts of physiology, such as metabolism, or alter response in subsequent environmental exposures.
Acknowledgments

We thank Shanaz Diessler, Marieke Hoekstra, Konstantinos Kompotis, Dessislava Petrova and the Lausanne Genomics Technologies Facility for technical assistance. J.Y. benefits from the Natural Sciences and Engineering Research Council of Canada Postgraduate Studies Doctoral scholarship. Work in the F.N. laboratory was supported by Swiss National Science Foundation Grant 310030_173079 and the École Polytechnique Fédérale de Lausanne.

Bibliography


Benjamini, Y, and Y Hochberg. 1995. "Controlling the False Discovery Rate - a


Ongen, H, A Buil, A A Brown, E T Dermitzakis, and O Delaneau. 2016. “Fast and
Efficient QTL Mapper for Thousands of Molecular Phenotypes.” 
*Bioinformatics* 32 (10): 1479–85. 


https://doi.org/10.1038/nn1462.

https://doi.org/10.1038/nrg.2016.150.


https://doi.org/10.1038/nprot.2013.118.


Zhang, Yong, Tao Liu, Clifford A Meyer, Jérôme Eeckhoute, David S Johnson, Bradley E Bernstein, Chad Nusbaum, Richard M Myers, Myles Brown, Wei Li, and X Shirley Liu. 2008. “Model-Based Analysis of ChIP-Seq (MACS).”

Figure 1

A

Baseline Day (BL/NSD)  Recovery Day 1 (R1)  Recovery Day 2 (R2)  Recovery Day 7 (R7)

Experiment time (hours) 0 3 6 12 18 24 27 30 36 42 48 54 72 78
Zettgeber time (hours) 0 3 6 12 18 0 6 12 18 0 6 0 6

B

NREM EEG delta power [% of ZT8-12]

Time [h] 0 24 48 72 96 120 144 168 192

Time spent in NREM [h/6h]

C

PC1 (0.20)

PC2 (0.0)

D

Wake

Condition

PC1

Time [h] 0 6 12 18 24 30 36 42 48 54 60 66 72 78

Technical replicates

Condition

EP

C1

Technical replicates

Time [ZT]

0 24 48 72 96 120 144 168 192
Figure 1
Overview of experimental design, sleep-wake activity, and gene expression data.
A. Tissue collection schedule with time from beginning of the experiment and corresponding ZT. White and grey bars below the timeline represent the 12h:12h light/dark cycle. Red bar: sleep deprivation. B. Long-term effects of sleep deprivation on NREMS delta power and quantity. Following a 6-hour SD (red shade), mice display a significant increase in delta power (1-4Hz; top), which returns to baseline levels within 6 hours. Subsequent recovery days (starting at T48) show no differences from baseline and persists until the end of the recording period, 6 days after SD (T192). Quantification of NREM sleep showed similar patterns across all recording days (bottom). Mean delta power values (± SEM) are expressed as the percentage of intraindividual deviations from the period of baseline with the lowest overall power (average across 2 days, ZT8-12). Blue bars represent significant differences from corresponding baseline values (t-test, p<0.05, n=6). C. Principal component analysis of expression of 13842 detected genes in RNA-seq data. Parenthesis of axis label denotes fraction of variance explained by the component. Colors denote Zeitgeber time (ZT0 to ZT12: light period, ZT12 to ZT18: dark period). Text labels denote time of experiment, in hours. D. First principal component plotted over time (top). Temporal EEG data over time averaged across 15 mice (bottom). EEG y-axis denotes number of minutes awake over last 5 minutes. PC1 increases during wake and decreases during sleep.
Figure 2

Immediate response, fast recovery

Log2 expression level (centered, scaled)

0 6 12 18 24 30 36 42 48 54 60 66 72 78

-log10(pval)
0.4
0.8
1.2
1.6

Immediate response, slow recovery

Prolonged/delayed response

Log2 expression level (centered, scaled)

0 6 12 18 24 30 36 42 48 54 60 66 72 78

-log10(pval)
0.4
0.8
1.2
1.6

No/weak response
Figure 2
Clustering analysis of gene expression data reveals diverse responses to sleep deprivation.
K-means clusters of 3461 genes (filtered with FDR adjusted pval < 0.001, likelihood ratio test).
Blue dashed line: average of the cluster under baseline, repeated over the three days of the experiment. Light grey thick line: cluster average. Red shaded box: time of sleep deprivation. Grey shaded bar: Mean statistical significance across genes using t-test between sleep-deprived and baseline conditions at the same Zeitgeber time.
Figure 3

A

Nr1d1

B

Glpcl1 (Flat)

C

Hif3a (C)

D

Egr2 (S)

E

Gkn3 (S + C)

F

Arnt1 (A)

G

Dpp (S + A)
Figure 3  
Model selection identifies sleep-wake dynamics and amplitude changes.  
A: Example of model fitting on *Nr1d1*. Dots represent RNA level data points, the red box the SD.  
Data points and all models, listed on bottom right. Best fitting model (cosine with amp change) is  
highlighted in dark bold.. Lower bar show EEG data for minutes spent awake in 5-minute bins. B.  
Number of genes per model. C-H. Examples of 6 genes with their best fit model shown in solid  
line.
Figure 4

A. Model S

B. Model S

C. Model C

D. Model A

E. Model A: damped amplitudes

F. Relative Amp Difference (baseline mean to peak)
Figure 4
Summary of gene modules by model parameters.
A. Summary of log2 abundance (positive values denote increased during sleep deprivation) as a function of predicted mRNA half-life in the sleep-wake driven model (S). B. Time constants of S process show a trend where the rate of change during wake is slower than during sleep.
C. Polar plot of log2 fold change (radial distance) and time of peak expression (clockwise angle) for genes in cosine model (C).
D. Log2 fold change before (x-axis) and after (y-axis) sleep deprivation treatment for genes in cosine with amp change (A).
E. Polar plot of log2 fold change before and after of selected genes with damped amplitudes.
F. Relative change in amplitude relative to day 0 (baseline) in sleep-wake driven models. Left panel show relative changes for day 1, right panel for day 2.
Figure 5

A  
ATAC-Seq

B  
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
Cluster 10

C  
Condition
- SD
- C4
- Technical replicates

Time [ZT]
- ZT0
- ZT3
- ZT6
- ZT12
- ZT18

log2 FC

-log10(p-value)
-2 0 2
-2 0 2
-2 0 2
0 6 12
0 6 12
0 6 12

Figure 5
Figure 5
Dynamics in chromatin accessibility and response to sleep deprivation.
A. PCA of heatmap of accessibility levels in SD and NSD samples at ZT6 and ZT12 (215’045 ATAC-seq loci). B. K-means clustering analysis of ATAC-seq signal (K=10) showing sleep-wake driven dynamics. C. Volcano plots comparing ATAC-seq differences at T3 vs. T27 (3 hours of SD), T6 vs. T30 (6 hours of SD), and T12 vs T36 (6 hours after SD).
Figure 6

A. Wiap1 (Model S)
chr15 66893881 66894552
3 kb from promoter

B. Dusp6 (Model S)
chr10 99250564 99251579
12 kb from promoter

C. Hif3a (Model C)
chr7 17058086 17058738
2 kb from promoter

D. Prkg2 (Model S+C)
chr5 98943551 98944236
14 kb from promoter

E. Ciart (Model S+A)
chr3 95881808 95883363
0.5 kb from promoter

F. Metn1 (Model S+A)
chr11 121679503 121680155
22 kb from promoter
Figure 6
Chromatin accessibility correlate with gene expression dynamics
(A-F) Examples of ATAC-seq peaks with dynamics that correlate with gene expression. Plot titles highlight gene name and its respect model, peak location (mm10), and distance from promoter. The log RNA-seq (red) and ATAC-seq (blue) counts are plotted on the same y-axis. Solid lines represent mean across biological replicates. Dashed lines represent the baseline signal. Open circles are signal from individual mice.
Figure 7

A

Zero from Mean Activity

Index

B

TF activity [A.U.]

Time (ZT)

C

RNA-seq or ATAC-seq levels

(log2 counts)

Time [h]

D

RNA-seq or ATAC-seq levels

(log2 counts)

Time [h]

RNAseq

ATACseq
Figure 7
Serum response factor underlies immediate early gene expression dynamics.

A. 179 TF motifs ranked by z-score that explains the temporal dynamics in the RNA-seq dataset. 

B. Inferred temporal activity of SRF. Error bars are standard deviations of the activity estimates. Lower bar denotes time spent awake in 5-minute bins.

C. Candidate SRF target genes. RNA-seq (top) and ATAC-seq (bottom) signal near SRF target genes. Additional tracks: SRF motif: predicted SRF motif instances. SRF ChIP-seq: ChIP-seq targeting SRF from (Esnault et al. 2014).

D. ATAC-seq (blue) and RNA-seq (red) levels over time at candidate SRF target genes. Solid lines represent mean across biological replicates. Open circles represent individual mice. Dashed line shows the baseline signal.
Supplemental Figure 1

A  T6 vs. T198 (ZT6)

B  T0 vs. T192 (ZT0)
Supplemental Figure 1
MA plot showing almost no differential expression 7 days after the beginning and end of sleep deprivation.
Supplemental Figure 2

No/weak response

Fabp7

Arntl

Obscn

Lfng
Supplemental Figure 2
Examples of individual genes in cluster 9 and 10 with damped amplitudes after sleep deprivation treatment.
Supplemental Figure 3

Flat (F)

Sleep-wake driven (S)

Cosine (C)

Cosine with amplitude change (C)

S + C

S + A

mRNA Abundance

Time
Supplemental Figure 3
Schematic of the 6 models used for model selection.
Supplemental Figure 4

A

Sleep-wake driven genes from Mongrain et al. 2010

B

Genes affected by SD from Maret et al. 2007
Supplemental Figure 4
A. Distribution of 61 sleep-wake driven genes previously identified from Mongrain et al. 2010 to the 6 models used in this study.
B. Distribution of 92 previously identified to be affected by sleep deprivation from Maret et al. 2007 to the 6 models used in this study.
Supplemental Figure 5

A-B. Examples of genes in S+C model showing upregulation (A) and downregulation (B) during sleep deprivation.

C-F. Examples of genes in S+A model. Clock and clock output genes show immediate response to sleep deprivation and damped amplitudes to after sleep deprivation.
Supplemental Figure 6

A. All peaks
B. Differentially active peaks at ZT3
C. Differentially active peaks at ZT6
D. Differentially active peaks at ZT12

E. Genes associated with ZT3 DAS
F. Genes associated with ZT6 DAS
G. Genes associated with ZT12 DAS
H. Genes associated with ZT3−ZT6 DAS
I. Genes associated with ZT6−ZT12 DAS
Supplemental Figure 6
A-D. Genomic features of ATAC-seq peaks. A. All peaks. B-D Peaks differentially active at T27 vs T3 (B), at T30 vs T6 (C), and at T36 vs T12 (D).
Supplemental Figure 7
Serum response factor underlies immediate early gene expression dynamics.
A. Candidate SRF target genes. RNA-seq (top) and ATAC-seq (bottom) signal near SRF target genes. Additional tracks: SRF motif: predicted SRF motif instances. SRF ChIP-seq: ChIP-seq targeting SRF from (Esnault et al. 2014). B. ATAC-seq (blue) and RNA-seq (red) levels over time at candidate SRF target genes. Solid lines represent mean across biological replicates. Open circles represent individual mice. Dashed line shows the baseline signal.
Chapter 4. Dynamic gene expression and regulation from circadian and sleep-wake processes in the mouse cortex

4.3 Conclusion and perspectives

This study systematically analyzes how dynamics in the transcriptome are regulated by the sleep homeostat (Process S), time-of-day (Process C), as well as interactions between the two. Sleep deprivation can have complex effects on circadian gene expression. Robust oscillations in clock and clock outputs genes become damped after sleep deprivation, and this damping can last more than 48 hours. Thus, our results highlight that observing clock genes during short exposures can have surprising responses. We also introduce a statistical framework to infer the types of dynamics that occur in gene expression. Our work also motivates future questions:

1. How do these non-periodic temporal functions manifest in other systems such as liver and feeding?
2. What are the single-cell dynamics and are there cell-types with specific dynamics within the cortex?
3. What underlying processes make dynamics robust or sensitive to sleep deprivation? And is it possible to perturb or reinforce rhythms using drugs?
This thesis addressed three specific questions in chronobiology:

1. How does the circadian clock regulate gene expression in a tissue-specific manner?
2. What is the role of chromatin interactions in regulating circadian gene expression and circadian rhythms?
3. How does diurnal gene expression respond to acute perturbations such as sleep deprivation?

I explored the first question in Chapter 2 by studying transcriptional regulatory modes that enable gene expression to oscillate in one tissue but not others. I developed new methods to analyze high-dimensional data with periodic structure, namely complex-valued singular value decomposition (cvSVD) in combination with model selection. This combination of unsupervised methods (cvSVD) and supervised methods (model selection) allowed me to explore circadian gene expression data at different levels of granularity: from whole transcriptome to gene modules and to individual genes.

Future extensions from this study will deepen our understanding of dynamic gene regulation across tissues. First, we can extend to posttranscriptional regulation to ask how the proportions of transcriptional and posttranscriptional regulation vary across tissues. I participated in the development of methods to systematically compare proportion of genes regulated by transcriptional and posttranscriptional regulation in mouse liver (Wang et al., 2018). Applying these methods to multiple tissues may reveal insights on dynamic gene expression of different regulatory layers across tissues. Second, the question of how gene products from one tissue influence gene expression in others have not been systematically explored. Inferring dynamics of inter-tissue communication requires the analysis of temporal gene expression data across tissues. Tissue-specific activation of ligand-receptor pathways can be predicted by analyzing RNA-seq data of tissues, using computational frameworks such as MARA. These predictions could then be validated by using agonists to activate candidate pathways. Moving
Chapter 5. Conclusions and Perspectives

beyond mRNA abundance towards protein abundance and activity (e.g., protein phosphorylation) (Wang et al., 2017) will provide a more direct readout of pathway activation. Finally, the 4C-seq analysis across tissues found that enhancers can contact a rhythmic promoter while looping out nearby nonrhythmic alternative promoters, confining rhythmic enhancer activity to specific promoters. How this specificity arises poses an intriguing open question for future work. Scaling up this analysis to many promoters using techniques such as promoter capture Hi-C (Mifsud et al., 2015) will allow different promoter structures (e.g., CpG islands, CG content, presence of core promoter motifs) to be analyzed to infer links between enhancer-promoter contacts and promoter sequences.

For the second question, discussed in Chapter 3, Jerome and I established chromatin interactions as a fundamental layer that enables proper timing of the circadian clock as well as robustness of circadian locomotor activity. Deleting a regulatory enhancer in Cry1 showed not only disrupted 24-hour dynamics in chromatin looping, but also shortened period of locomotor activity. The period decreased by 15 minutes, which is expected considering that deleting the entire Cry1 gene shortens the locomotor activity by approximately 1.2 hours (Ko and Takahashi, 2006). This study incorporated analysis at the level of single cells, whole tissue, and behavior. Nevertheless, how robust oscillations at the tissue level result from from noisy gene expression in single cells, which involve promoter-enhancer interactions, remain fascinating open questions. Single-cell technologies that combine single-molecule RNA FISH with DNA labeling are beginning to uncover the temporal relationships between chromatin interactions and gene expression (Chen et al., 2018). Extending these techniques to investigate gene expression dynamics will elucidate how intrinsic and extrinsic noise components influence dynamics.

Third, Chapter 4 looks beyond periodic gene expression by investigating how the plasticity of the circadian clock responds to sleep deprivation over multiple days. This unique study design allowed us to incorporate sleep-wake history of mice to model long-term dynamics in gene expression. Our findings that many core clock and clock output genes have damped amplitudes and perturbed dynamics in response to sleep deprivation suggest that clock genes are highly adaptive to physiological states and acute environmental changes. We have gained quantitative insights in the role of sleep homeostat and the circadian clock on gene expression dynamics, and further investigations that include other perturbations such as feeding and temperature would reveal the different pathways in which the circadian clock responds to environmental signals. These other perturbations would look into other tissues such as liver and link with how different tissues may coordinate physiology to adapt to environmental changes. The tissue heterogeneity in the mouse cortex warrants further investigation into the single-cell dynamics and how different cell-types may have different dynamics within the cortex. Our findings also open up questions to whether robustness or sensitivity of sleep deprivation could be modulated pharmacologically.

We expect advances in experimental techniques as well as in computational methods to uncover general principles of how organisms have embraced environmental periodicity
across all scales of biological. For example, theoretical frameworks of how rhythms in enzyme activities, metabolic flux, and metabolites coordinate to propagate rhythms may uncover novel insight into the regulation of temporal metabolism (Thurley et al., 2017). Assaying enzyme activities in a high-throughput manner, which is yet not available, would enable systematic analysis of the relationships between gene expression, protein abundance, and enzyme activities.

While most molecular chronobiology findings have relied on studying model organisms, there are currently exciting opportunities for human chronobiology. Genome-wide association and candidate gene sequencing studies have reported genetic variations associated with circadian clock-related and sleep phenotypes (Hu et al., 2016; Allebrandt et al., 2010; Shi et al., 2017). Beyond self-reported questionnaires, temporal activity patterns or other behaviors, such as food intake, can be directly measured through smartphone apps (Gill and Panda, 2015; Roenneberg, 2017; Aledavood et al., 2015). Time stamping assays such as the BodyTime assay combines both computational methods and, critically, a robust experimental assay to accurately determine internal circadian time from a blood sample (Wittenbrink et al., 2018).

Analysis of large consortia of gene expression across human tissues (Ardlie and Guigó, 2017) can be used to discover circadian gene expression in human by applying machine learning methods discussed above (Anafi et al., 2017; Ruben et al., 2018). Taking into account natural variation may reveal coding and noncoding variants that affect circadian gene expression. Noncoding variants may be associated with disruption of TF binding (Deplancke et al., 2016), which can reveal links between genome variation and gene regulatory mechanisms underlying behavioral phenotypes. A first indication that non-coding DNA may have phenotypic consequence on the mammalian circadian clock was reported in mouse (Mermet et al., 2018), and it will be interesting to assess how this generalizes to human.


147


Bibliography


Johan H Gibcus, Kumiko Samejima, Anton Goloborodko, Itaru Samejima, Natalia Nau-
mova, Johannes Nuebler, Masato T Kanemaki, Linfeng Xie, James R Paulson, William C Earnshaw, Leonid A Mirny, and Job Dekker. A pathway for mitotic chromosome for-


Shubhroz Gill and Satchidananda Panda. A Smartphone App Reveals Erratic Diurnal Eating
Patterns in Humans that Can Be Modulated for Health Benefits. *Cell Metabolism*, 22(5):

Xiaoping Han, Renying Wang, Yincong Zhou, Lijiang Fei, Huiyu Sun, Shujing Lai, Assieh
Saadatpour, Ziming Zhou, Haide Chen, Fang Ye, Daosheng Huang, Yang Xu, Wentao Huang,
Mengmeng Jiang, Xinyi Jiang, Jie Mao, Yao Chen, Chenyu Lu, Jin Xie, Qun Fang, Yibin Wang,
Rui Yue, Tiefeng Li, He Huang, Stuart H. Orkin, Guo-Cheng Yuan, Ming Chen, and Guoji
science/article/pii/S0092867418301168.

Michael H. Hastings, Elizabeth S. Maywood, and Marco Brancaccio. Generation of circadian
ISSN 1471-003X. doi: 10.1038/s41583-018-0026-z. URL http://www.nature.com/articles/
s41583-018-0026-z.

Nathaniel P Hoyle, Estere Seinkmane, Marrit Putker, Kevin A Feeney, Toke P Krogager, Jo-
hanna E Chesham, Liam K Bray, Justyn M Thomas, Ken Dunn, John Blaikley, and John S
O’Neill. Circadian actin dynamics drive rhythmic fibroblast mobilization during wound
healing. *Science translational medicine*, 9(415):eaal2774, nov 2017. ISSN 1946-6242. doi:

Youna Hu, Alena Shmygelska, David Tran, Nicholas Eriksson, Joyce Y. Tung, and David A.
Hinds. GWAS of 89,283 individuals identifies genetic variants associated with self-reporting
of being a morning person. *Nature Communications*, 7:10448, feb 2016. ISSN 2041-1723. doi:
10.1038/ncomms10448. URL http://www.nature.com/doifinder/10.1038/ncomms10448.

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 Bio-


Anna Kriebs, Sabine D Jordan, Erin Soto, Emma Henriksson, Colby R Sandate, Megan E Vaughan, Alanna B Chan, Drew Duglan, Stephanie J Papp, Anne-Laure Huber, Megan E


Bibliography


Yichi Xu, Weimin Guo, Ping Li, Yan Zhang, Meng Zhao, Zenghua Fan, Zhihu Zhao, and Jun Yan. Long-Range Chromosome Interactions Mediated by Cohesin Shape Circadian Gene Expression. *PLOS Genetics*, 12(5):e1005992, may 2016. ISSN 1553-7404. doi: 10.1371/journal.pgen.1005992. URL http://dx.plos.org/10.1371/journal.pgen.1005992.


Jake Yeung

Education

Ph.D. in Biotechnology & Bioengineering, 2018 (Expected)
École Polytechnique Fédérale de Lausanne (EPFL), Building AAB 040, Station 15, Lausanne, Switzerland

M.Sc. in Genome Science and Technology, 2014
University of British Columbia
GPA: 91.5/100,

B.A.Sc. in Chemical Engineering, 2012
University of Waterloo
GPA: 90.3/100,
Graduated with Distinction and Dean’s Honours,
Thesis: Mathematical modeling of a renewable energy hub.

Skills

Bioinformatics: Temporal analysis of RNA-Seq, 4C-Seq, DNase-Seq, ATAC-Seq, and smRNA FISH
Engineering: Machine learning, mathematical modeling, dynamical systems
Computing: R, Python, MATLAB, UNIX-like OS, shell scripts

Manuscripts in the Pipeline


Manuscripts


A. Wyatt*, F. Mo*, ..., J. Yeung et. al., The inter-tumor transcriptome heterogeneity of high-risk primary prostate cancer, Genome Biology.

163
Talks at Conferences


Silver award for outstanding research talk

Papers at Conferences


Awards

2017: Francophone Society of Chronobiology Travel Award (European Biological Rhythms Society in Amsterdam)
2015-2018: Natural Resources and Engineering Research Council of Canada Doctoral Fellowship
2012-2014: Natural Resources and Engineering Research Council of Canada Graduate Award
2012-2014: Genome Science and Technology Graduate Award
2013: The Sidhoo Charities Fund Travel Grant (Genome Informatics conference at Cold Spring Harbor)
2012: First Prize Walmart Green Student Challenge Competition
2011: University of Waterloo Undergraduate Research Assistantship
2010: University of Waterloo Dr. F. Hecker Foreign Exchange Scholarship
2007-2012: Dean’s Honours List
2007: University of Waterloo President’s Scholarship

Teaching and Mentoring Experience

**Teaching Assistant**, *Computational and Mathematical Modeling in Biology*, 2014-2018

Prof. Naef, EPFL.
**Guest Lecturer**, *Gene Set Analysis*, 2013

Genome Science and Technology Laboratory Course: Genome Analysis Module, UBC.

**Languages**

English (native), French (conversational), Chinese (conversational, Cantonese and Mandarin), German (basic)

**References**

**Felix Naef**,
*École Polytechnique Fédérale de Lausanne,*
Email: felix.naef@epfl.ch.
Main PhD advisor

**Paul Franken**,  
*University of Lausanne,*  
Email: paul.franken@unil.ch.
Experimental collaborator and expert in sleep homeostasis. Together we work on temporal dynamics in gene expression during sleep-wake cycles and in response to acute sleep-deprivation.

**Frederic Gachon**,  
*Nestlé Institute of Health Sciences S.A.,*  
Email: frederic.gachon@epfl.ch.
Experimental collaborator and expert in circadian biology. His lab is involved in performing assays on most of the mouse experiments (RNA-Seq, ribosome profiling, nuclear proteomics, phosphoproteomics).

**Hobbies**

Eating homemade baked goods, playing Bach on the piano, and snowboarding in the Swiss alps.
Ce document a été imprimé au Centre d'impression EPFL, imprimerie climatiquement neutre, certifiée myClimate.