

# Phase specific transcriptional regulation of circadian clock and metabolism in mouse liver

THÈSE N° 7188 (2016)

PRÉSENTÉE LE 6 OCTOBRE 2016

À LA FACULTÉ DES SCIENCES DE LA VIE

UNITÉ DU PROF. NAEF

PROGRAMME DOCTORAL EN APPROCHES MOLÉCULAIRES DU VIVANT

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Jonathan Aryeh SOBEL

acceptée sur proposition du jury:

Prof. V. Simanis, président du jury

Prof. F. Naef, directeur de thèse

Prof. N. Hernandez, rapporteuse

Dr P. Westermark, rapporteur

Prof. Ph. Bucher, rapporteur



ÉCOLE POLYTECHNIQUE  
FÉDÉRALE DE LAUSANNE

Suisse  
2016



Why do we do basic research ?  
To learn about ourselves.  
— Walter Gilbert

To my wife Aline and to my parents.





# Acknowledgements

First I would like to thank Prof. Felix Naef for giving me the opportunity to do a Ph.D. in his lab, and for his support and guidance during the past four years. In addition, I would like to acknowledge all the great people that were working in the Naef lab, notably Jerome Mermet, Damien Nicolas, Eric Pacquet, Saeed Omid, Jingkui Wang, Daniel Mauvoisin, Cedric Gobet, Julia Cajan, Laura Symul, Jonathan Bieler, Benjamin Zoller, Nacho Molina and Kyle Gustafson. I had the chance to collaborate with Irina Krier, who taught me a lot about genomics and proper data analysis. I want to thank her for her friendship and the great scientific discussion that we had.

This project was a broad collaboration between experimental labs and computational labs. Therefore, I would like to highlight the enormous amount of work needed to produce the CycliX dataset. I would like to thank particularly Prof. Ueli Schibler and Dr. Teemu Andersin, Prof. Nouria Hernandez, Prof. Bart Deplanke and Dr. Sunil Raghav.

In addition, I would like to thank my dissertation committee for their evaluation of this work, notably Prof. Pal Westermark, Prof. Philipp Bucher, Prof. Viesturs Simanis and Prof. Nouria Hernandez.

Finally, I want to thank my wife Aline for her patience and her support during these intense four years.

*Lausanne, 8 September 2016*

J. S.



# Abstract

The molecular clock has been conserved from cyanobacteria to mammals and is believed to align behavioral and biochemical processes with the diurnal cycle. This cellular mechanism has been an advantage to increase the fitness of organisms through the ability to anticipate food availability or predator presence. It has been recently suggested that the molecular clock is used to optimize the energy consumption of cells and accumulating evidence has revealed that the circadian clock is intimately interconnected with the metabolic cycle. The nature of these interconnections is yet not clear at the transcriptional level, and the contribution of cis-regulatory modules has not been elucidated.

The emergence of novel high-throughput technologies, in the field of next generation sequencing, as ChIP-seq, and DNase I-seq, unveiled chromatin landscape with an unprecedented resolution. These techniques allow a genome-wide investigation of accessible chromatin regions and DNA-binding proteins such as transcription factors, as well as histones modifications or Polymerase II (Pol II) presence. Accessible chromatin regions of the genome are implicated in diverse processes, such as gene regulation through enhancers and promoters, insulation of genomic domains or alternative splicing. For instance, these regions are important in the differentiation process during development. They allow cell-type specific programs that are controlled by tissue-specific transcription factors and chromatin modifiers. The tissue-specific regulation of the circadian clock remains unclear. Therefore, we compared genome-wide BMAL1 binding and chromatin accessibility in the liver and in NIH3T3 fibroblasts.

Moreover, For the first time, our study explores the dynamics of accessible regions every 4 hours over one day in mouse liver. In this study, we show that a substantial fraction of these accessible sites are oscillating during a diurnal cycle with a circadian period. Furthermore, these sites are coordinated in time and space (with respect to their genomic location) using Pol II transcription and some histone modifications, such as histone 3 lysine 27 acetylations (H3K27ac).

We observed that these accessible regions are enriched in the proximity of actively transcribed genes, and that they are dynamically affected by the binding of transcription factors such as BMAL1. We investigated wild-type (WT) and *Bmal1*<sup>-/-</sup> genotypes, in night restricted feeding regimen and in Light-Dark cycle, to study the circadian clock regulatory network underlying diurnal transcription. Our analysis revealed that a certain fraction of genes was still fluctuating with a 24h period in *Bmal1*<sup>-/-</sup> genotype. We observed that these genes were related to fatty acid, or steroid metabolism, suggesting food entrainment. Unexpectedly, a large proportion

## Acknowledgements

---

of genes with a diurnal behavior in the *Bmal1*<sup>-/-</sup> context were not cycling in the WT and were related to insulin signaling, TCA cycle, proteolysis and amino acid or sugar metabolism. In order to investigate the transcriptional control of these biological processes, we applied a penalized generalized linear model to infer the activity of transcription factor binding motifs in oscillating accessible sites using Pol II loadings at the transcription start sites of nearby genes. We were able to recapitulate the known regulatory elements of the circadian clock, notably E-box, D-Box, and ROR-responsive elements (RRE) in the WT genotype. On the other hand, we found that Forkhead box (FOX), glucocorticoids responsive elements (GRE), C-AMP Response Element (CRE) were the main contributors of the regulation of oscillating genes in *Bmal1*<sup>-/-</sup>. Finally, using a mixture model to detect footprints with a base pair resolution, we studied the dynamics of the accessibility overlapping E-box. Our last analysis suggested that BMAL1/CLOCK is binding on double E-boxes with a spacer of 6 or 7 bp in a hetero-tetramer configuration. A 3D structure model further supported this binding mode.

In Summary, we used DNase I-seq, ChIP-seq of H3K27ac and Pol II to study the circadian chromatin landscape in a 4h time-resolved experimental design. We uncovered the underlying circadian transcriptional regulatory network and, we dissected the chromatin accessibility around BMAL1 binding sites at a base pair resolution, which led to an unappreciated mode of binding of BMAL1/CLOCK in a hetero-tetramer conformation on double E-boxes. Lastly, we found tissue-specific factors that might contribute to tissue-specific binding of BMAL1/CLOCK.

Key words: Circadian clock, chromatin accessibility, ChIP-seq, metabolism, nutrient response cycle, mouse liver, Bmal1 knockout, DNase I hypersensitive sites, transcriptional regulation, tissue-specificity, NIH3T3 fibroblasts

## Résumé

L'horloge moléculaire est un mécanisme présent depuis la cyanobactérie jusqu'aux mammifères. Ce mécanisme cellulaire est censé aligner le comportement ainsi que les processus biochimiques au cycle journalier. Il a représenté de formidables avantages en termes d'évolution en permettant d'améliorer la survie d'un organisme grâce à sa capacité à anticiper la présence de prédateurs ainsi que la disponibilité de la nourriture. Il a été récemment suggéré que l'horloge moléculaire permet aussi d'optimiser la consommation d'énergie des cellules. Un nombre croissant d'études permettent d'établir un lien entre cette horloge interne et le métabolisme. En revanche, la nature de cette relation n'est pas claire au niveau transcriptionnel et la contribution de sites régulateurs en cis n'a pas été expliquée.

L'avènement de nouvelles technologies de séquençage à haut débit, comme les techniques de ChIP-seq ou de DNase-seq, ont révélé la complexité de la chromatine à un niveau de résolution sans précédent. Ces techniques permettent d'étudier l'accessibilité de la chromatine et de découvrir les sites de fixation des facteurs de transcription. Elles permettent aussi de révéler les modifications épigénétiques comme les marques d'histones ou la présence de l'ARN polymérase II à l'échelle du génome. Ces régions accessibles du génome sont impliquées dans divers processus biologiques, notamment la régulation de l'expression des gènes au travers des sites promoteurs et amplificateurs (enhancers), la séparation de domaines génomiques grâce aux insulateurs ou encore l'épissage alternatif des ARN messagers. Ces régions jouent un rôle important lors de la différenciation cellulaire pendant le développement. Elles permettent des programmes spécifiques aux tissus, contrôlés par des facteurs de transcription ainsi que des modificateurs de la chromatine.

Malgré son importance, la régulation des processus biologiques spécifiques aux tissus par l'horloge interne reste méconnue. Afin de progresser dans la compréhension de ce mécanisme, nous avons comparé, d'une part, les différents sites de liaison de BMAL1, d'autre part, l'accessibilité de la chromatine à l'échelle du génome dans le foie et dans les fibroblastes NIH3T3.

Notre étude est la première à explorer la dynamique de ces régions accessibles, lors d'un échantillonnage de foie de souris prélevé toutes les 4 heures sur une période d'une journée. Ce travail met en évidence les oscillations d'une fraction non négligeable de ces sites lors du cycle diurne. De plus, l'on voit qu'ils sont coordonnés dans le temps et le long du génome. Leur accessibilité est également en phase avec l'ARN polymérase II et l'acétylation des histones 3 sur la lysine 27.

Nous avons observé qu'il existe davantage de ces régions accessibles à proximité de gènes

## Acknowledgements

---

activement transcrits, et qu'elles sont dynamiquement affectées par la liaison de facteurs de transcription tels que BMAL1. Afin d'étudier la régulation de la transcription circadienne, nous avons comparé les génotypes de type sauvage (WT) ainsi que des souris mutées pour le gène BMAL1 (*Bmal1*<sup>-/-</sup>) en ne leur fournissant de la nourriture que de nuit et en alternant douze heures d'exposition à la lumière et douze heures d'obscurité. L'analyse a révélé que dans le contexte *Bmal1*<sup>-/-</sup>, une fraction des gènes fluctuent toujours avec une périodicité de 24 heures. Ces gènes sont impliqués dans le métabolisme des acides gras ainsi que des stéroïdes, ce qui suggère une synchronisation par la nourriture. De manière inattendue, une large partie des gènes fluctuant dans le contexte *Bmal1*<sup>-/-</sup> n'oscille pas dans le type sauvage. Ces gènes participent au métabolisme des sucres et des acides aminés, au cycle de Krebs, à la signalisation liée à l'insuline ainsi qu'à la protéolyse.

Nous avons appliqué un modèle linéaire généralisé et pénalisé afin d'étudier la régulation transcriptionnelle de ces processus biologiques. Il s'agissait d'inférer l'activité des motifs de liaison des facteurs de transcription sur l'ADN dans les régions accessibles en utilisant le signal de la RNA polymérase II aux sites d'initiation de la transcription des gènes actifs à proximité. Nous avons détecté les principaux motifs connus de l'horloge dans le type sauvage, notamment les boîtes E (E-box), les boîtes D (D-box) ainsi que les éléments répondant à ROR (RRE). De plus, nous avons découvert que dans les souris *Bmal1*<sup>-/-</sup>, les principaux motifs actifs étaient les boîtes F (Forkhead box, FOX), les éléments de réponse aux glucocorticoïdes (GRE) et les éléments de réponse aux AMP cycliques (CREB). Enfin, en utilisant un modèle mixte pour détecter les empreintes des facteurs de transcription sur le signal DNase avec une résolution à la paire de base, nous avons étudié la dynamique de l'accessibilité autour des boîtes E. Notre analyse suggère que BMAL/CLOCK peut se lier à deux boîtes E séparées par six ou sept paires de bases dans une configuration d'hétéro-tétramères. Un modèle de structure en 3D sous-tend ce mode de fixation.

En résumé, nous avons utilisé les techniques de DNase I-seq et de ChIP-seq pour Pol II et H3K27ac afin d'étudier la dynamique de la chromatine lors du cycle circadien avec un échantillonnage toutes les quatre heures. Nous avons découvert le réseau circadien de régulateurs transcriptionnels et avons disséqué le mode de liaison de BMAL1 à la chromatine, ce qui nous a conduit à découvrir un mode de liaison à l'ADN méconnu de BMAL1/CLOCK avec une conformation en hétéro-tétramère. Enfin, nous avons détecté des facteurs propres à un tissu qui participent probablement à la fixation spécifique de BMAL1 à ce dernier le long du génome.

Mots-clefs : horloge circadienne, accessibilité de la chromatine, ChIP-seq, métabolisme, cycle de réponse à la nourriture, foie de souris, sites hypersensibles à la DNase I, régulation transcriptionnelle, spécificité tissulaire, fibroblastes NIH3T3.

## Attribution notice

The CycliX consortium produced most of the data presented in this thesis. The experimental part including mouse handling and bench work was done by several labs at the University of Lausanne (UNIL), University of Geneva (UNIGE) and EPFL. Polymerase II chromatin immunoprecipitation followed by sequencing (Pol II ChIP-seq) were produced by the lab of Nouria Hernandez. Histone 3 lysine 27 acetylation chromatin immunoprecipitation followed by sequencing (H3K27ac ChIP-seq) were produced by Sunil Raghav from Bart Deplanke lab at EPFL. The DNase I-seq data were generated by Teemu Andersin from Ueli Schibler lab at UNIGE. Data analysis was mainly done by myself under the supervision of Felix Naef and in collaboration with Irina Krier from Bart Deplanke's lab. Irina was responsible for the quality control (QC) of Pol II and H3K27ac data. She was involved in the statistical analysis of oscillating accessible sites and data mining of publically available ChIP-seq data.

More precisely, I was responsible for the DNase I-seq data analysis, which included QC, peak calling, quantification and signal normalization. I had the responsibility to find DNase I footprints and to study the dynamics of transcription factor binding in these accessible regions using a mixture model proposed by Felix. I worked as well on the motif scan in DNase I hypersensitive sites (DHS), and I have optimized a penalized linear model to infer the motif activities in these DHSs using Pol II signal. I have performed the GO/Pathway enrichment analysis to characterize differences between WT and *Bmal1*<sup>-/-</sup> mice, and I have worked on the biological interpretation.

I had the opportunity to collaborate with Julia Cajan to study the context specific binding of BMAL1 in liver compared to NIH3T3 cells. For this side project, I re-analyzed DNase I-seq data from ENCODE project (peak calling, footprint detection, sequence scan). I also reanalyzed the ChIP-seq data produced by Julia in NIH3T3 and I compared it with the ChIP-seq data generated by Guillaume Rey.





# Contents

<b>Acknowledgements</b>	<b>i</b>
<b>Abstract (English/Français)</b>	<b>iii</b>
<b>Attribution notice</b>	<b>vii</b>
<b>List of figures</b>	<b>xiii</b>
<b>List of tables</b>	<b>xv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Chronobiology: the study of periodic events in living organisms . . . . .	1
1.1.1 Early circadian pioneers discovered endogenous near 24 h rhythms in plants . . . . .	1
1.1.2 Modern chronobiology: the study of biological rhythms in mammals, insects, fungi and bacteria . . . . .	2
1.1.3 Concepts and definitions in chronobiology . . . . .	3
1.1.4 Human chronobiology . . . . .	8
1.2 Chromatin structure and accessibility . . . . .	13
1.2.1 Chromatin structure and transcriptional regulation . . . . .	13
1.2.2 Measuring chromatin accessibility using DNase I hypersensitivity . . . . .	17
1.3 The organization of the mammalian circadian clock . . . . .	19
1.3.1 The hierarchical organization of the mammalian clock . . . . .	19
1.3.2 The mammalian molecular core clock . . . . .	21
1.4 Circadian transcription and gene regulation . . . . .	22
1.4.1 Transcriptome-wide studies of the circadian clock . . . . .	22
1.4.2 Genome-wide studies of the circadian chromatin landscape . . . . .	25
1.4.3 Chromatin remodelers and topological domains . . . . .	26
1.4.4 Computational analysis of the transcriptional regulation of oscillating genes . . . . .	27
1.5 Interplay between circadian clock and metabolism . . . . .	29
1.5.1 Impaired metabolic phenotypes associated with mutations of core clock genes . . . . .	29
1.5.2 Clock controlled metabolism in mouse liver . . . . .	31

## Contents

---

1.6	Objectives . . . . .	35
1.7	Achievements . . . . .	35
<b>2</b>	<b>Results</b>	<b>37</b>
2.1	Chromatin landscape in mouse liver . . . . .	39
2.1.1	Genome-scale mapping of DNase I hypersensitivity in mouse liver . . . . .	39
2.1.2	DHS dynamics and correlation with Pol II and H3K27ac signals . . . . .	41
2.1.3	DNase I Footprint detection, localization and characteristics . . . . .	48
2.1.4	Genome wide impact of BMAL1 on DNA accessibility . . . . .	52
2.1.5	Digital genomic footprinting and transcription factor binding around the clock . . . . .	54
2.1.6	3D structure of BMAL1/CLOCK hetero-tetramer model bound on DNA . . . . .	58
2.2	System specific chromatin accessibility and BMAL1 co-regulators . . . . .	62
2.2.1	BMAL1 binding and DNase I hypersensitivity differ between fibroblasts and hepatocytes . . . . .	62
2.2.2	BMAL1 co-regulators are enriched in a cell-type specific manner . . . . .	63
2.2.3	E2F1 was identified as a putative fibroblasts-specific regulator enriched in BMAL1 sites . . . . .	67
2.2.4	ZEB1 was identified as a potential co-regulatory factor of BMAL1 . . . . .	68
2.2.5	Digital genomic footprinting confirms predicted liver/NIH3T3 specific factor . . . . .	69
2.3	Phenotypic consequences of whole body <i>Bmal1</i> knockout . . . . .	70
2.3.1	Phase specific regulation by transcription factors in WT and in <i>Bmal1</i> <sup>-/-</sup> mouse liver under night restricted feeding . . . . .	70
2.3.2	CREB phosphorylation is still oscillating in <i>Bmal1</i> <sup>-/-</sup> mouse liver . . . . .	75
2.3.3	Oscillating gene expression and related pathway in WT and <i>Bmal1</i> <sup>-/-</sup> mouse liver . . . . .	76
2.3.4	Glucose, lipids and fatty acids metabolism in WT and <i>Bmal1</i> <sup>-/-</sup> . . . . .	79
<b>3</b>	<b>Discussion</b>	<b>85</b>
3.1	Abundant diurnal DHSs in the mouse liver . . . . .	85
3.2	Limited impact of BMAL1 removal onto the liver chromatin landscape . . . . .	86
3.3	Tissue specific regulation of the circadian clock . . . . .	87
3.4	Regression models uncover factors which maintain cycling patterns in gene expression in the absence of BMAL1 activity . . . . .	88
3.5	Food entrainment and circadian regulation of lipid and sugar metabolism in liver . . . . .	89
3.6	Limitations of our approach . . . . .	90
3.6.1	Enhancer-TSS mapping . . . . .	90
3.6.2	Characterization of transcription factor binding sites . . . . .	91
3.7	Conclusions . . . . .	91
3.8	Perspectives . . . . .	93
3.8.1	Benefits of chrono-nutrition and chrono-therapy . . . . .	93
3.9	More circadian layers of regulation in cells . . . . .	94

3.9.1 Emerging trends in functional -omics and in (chrono)biology . . . . .	95
<b>4 Methods</b>	<b>97</b>
4.1 Experimental design, protocols and data set quality control . . . . .	97
4.2 Peak calling optimization . . . . .	103
4.3 Signal normalization . . . . .	105
4.4 Assessment of oscillatory signal . . . . .	106
4.5 Motif analysis in DNase I hypersensitive sites . . . . .	108
4.6 DNase I-seq footprint detection . . . . .	109
4.6.1 TFBS motif-centric approach . . . . .	109
4.6.2 <i>De novo</i> detection of footprint from DNase I signal . . . . .	112
4.6.3 DNase I cutting bias . . . . .	113
4.7 Linear model to infer transcription factor binding motif temporal activity . . .	115
4.8 Annotation of DHS to active transcripts and enrichment analysis . . . . .	120
4.9 3D structure of the heterotetramer BMAL1/CLOCK and molecular dynamics .	120
<b>A Appendix</b>	<b>123</b>
A.1 Mixture model for DNase I footprint detection . . . . .	123
A.2 <i>Circadian life</i> , a smartphone app	
to monitor our circadian activity . . . . .	126
A.2.1 Introduction . . . . .	126
A.2.2 Goal of the project . . . . .	126
A.2.3 Preliminary results . . . . .	127
A.2.4 Sleep-wake cycle analysis . . . . .	129
A.2.5 Perspectives . . . . .	131
<b>Bibliography</b>	<b>166</b>
<b>Curriculum Vitae</b>	<b>167</b>



# List of Figures

1.1	The flower clock . . . . .	2
1.2	Definition of phase,amplitude,period and mesor . . . . .	4
1.3	Phase resetting and circadian cycle perturbation . . . . .	5
1.4	Transcriptional-translational delayed feedback loop . . . . .	6
1.5	Illustration of the actogram . . . . .	8
1.6	Familial advanced sleep-phase syndrome . . . . .	9
1.7	Chronotype and place of residence . . . . .	10
1.8	The clock coordinate behavioral and metabolic processes according to time of day	12
1.9	Components of transcriptional regulation . . . . .	14
1.10	Illustration of the ChIP-seq principle . . . . .	15
1.11	Histone modification and chromatin organization . . . . .	16
1.12	Principles of a DNase I footprinting experiments . . . . .	18
1.13	The hierarchical structure of the mammalian clock . . . . .	20
1.14	The structure of the mammalian core clock . . . . .	21
1.15	Models for transcriptional regulation of hepatic gluconeogenesis under fasting and feeding conditions . . . . .	32
1.16	Direct and indirect outputs of the core clock mechanism . . . . .	33
2.1	Genome-wide localization of DNase I accessibility . . . . .	40
2.2	DNase I accessibility at <i>Dbp</i> locus . . . . .	42
2.3	DNase I accessibility, Pol II and H3K27ac dynamics at <i>Dbp</i> locus . . . . .	43
2.4	DNase I accessibility, Pol II and H3K27ac dynamics at <i>Npas2</i> locus . . . . .	44
2.5	Rhythmicity assessment, phases and amplitudes of DHS in DNase I, Pol II and H3k27ac signal . . . . .	46
2.6	Phase correlations of DHS in DNase I, Pol II and H3k27ac signal . . . . .	47
2.7	DNase I accessibility and Footprint at <i>Alb</i> promoter . . . . .	48
2.8	Location-dependent footprint characteristics of DHSs . . . . .	50
2.9	<i>Reverba</i> enhancer footprint characteristics . . . . .	51
2.10	Genome wide impact of BMAL1 on DNA accessibility . . . . .	53
2.11	Time-dependent footprint on double E-box motif with a spacer of 6bp over- lapped by BMAL1 binding sites . . . . .	54
2.12	Time-dependent footprint on single E-box of USF1 and BMAL1, and double E-box footprint with a spacer of 7 bp . . . . .	55

## List of Figures

---

2.13	Time-dependent footprint of ROR, HSE, SREBP and CREB . . . . .	57
2.14	Molecular dynamic simulation BMAL1/CLOCK hetero-tetramer model bound on DNA . . . . .	59
2.15	3D structure of BMAL1/CLOCK hetero-tetramer model bound on DNA . . . . .	60
2.16	BMAL1/CLOCK hetero-tetramer model bound on DNA: side view with interacting residuals . . . . .	61
2.17	Context specific binding of BMAL1/CLOCK . . . . .	62
2.18	Context-specific chromatin accessibility and BMAL1 co-regulators . . . . .	64
2.19	Enrichment of liver/NIH3T3 specific regulators in DHS and in BMAL1 binding sites . . . . .	65
2.20	Predicted liver/NIH3T3 specific regulators . . . . .	66
2.21	Promoter state of <i>E2f1</i> in fibroblasts and liver . . . . .	67
2.22	ZEB1 was identified as a potential co-regulatory factor of BMAL1 . . . . .	68
2.23	Context-specific chromatin accessibility of E-box, NRF1 and HNF4 . . . . .	69
2.24	Computational method for discovering phase specific activity of cis-regulatory elements . . . . .	71
2.25	Phase-specific activity of cis-regulatory elements in WT and <i>Bmal1</i> <sup>-/-</sup> mouse liver . . . . .	72
2.26	CREB validation by western blot in WT an in <i>Bmal1</i> <sup>-/-</sup> . . . . .	75
2.27	Rhythmicity analysis of mRNA from microarray in WT and <i>Bmal1</i> <sup>-/-</sup> . . . . .	76
2.28	Kegg and Reactome Pathway enrichment and amplitudes analysis . . . . .	77
2.29	Glycolysis and gluconeogenesis pathways . . . . .	80
2.30	Effect of Bmal1 knockout on key steps of glycolysis/gluconeogenesis . . . . .	81
2.31	<i>Gck</i> and <i>G6pc</i> promoter: example of transcriptional control of glucose metabolism	83
2.32	Free Fatty acids, triglycerides and ketone bodies in WT and in <i>Bmal1</i> <sup>-/-</sup> . . . . .	84
4.1	Peak calling strategy . . . . .	104
4.2	Motif stratification . . . . .	109
4.3	Digital genomic footprinting motif-centric method . . . . .	111
4.4	Digital genomic footprinting signal-centric method . . . . .	112
4.5	DNase I cutting bias . . . . .	114
4.6	Linear model for motif activity inference . . . . .	115
4.7	Penalized linear model with four different motif databases . . . . .	118
4.8	Optimisation of $\alpha$ and $\lambda$ in the elastic-net . . . . .	119
A.1	<i>Circadian life</i> App structure . . . . .	127
A.2	Humans actograms sample of two users from our data acquisition using our smartphone app . . . . .	128
A.3	Average activity during weeks and weekends for a single individual using our smartphone app . . . . .	128
A.4	HMM for smartphone app data analysis . . . . .	130
A.5	User-friendly App interface . . . . .	131

# List of Tables

1.1	Circadian transcriptome studies . . . . .	22
1.2	Core clock Mutations and Metabolic Phenotypes . . . . .	30
2.1	Residuals implicated in the interaction of the two hetero-dimers BMAL1:CLOCK	58
2.2	Top 40 inferred motif activity in WT mouse liver . . . . .	73
2.3	Top 40 inferred motif activity in <i>Bmal1</i> <sup>-/-</sup> mouse liver . . . . .	74
2.4	Reactome and Kegg pathway analysis of oscillating genes in WT and <i>Bmal1</i> <sup>-/-</sup> mouse liver . . . . .	78
4.1	mapping Quality control . . . . .	101
4.2	PTIH . . . . .	102
4.3	DNase I cleavage bias in our DNase I libraries . . . . .	114
4.4	Selected non-redundant PWMs from different sources . . . . .	117





# 1 Introduction

## 1.1 Chronobiology: the study of periodic events in living organisms

### 1.1.1 Early circadian pioneers discovered endogenous near 24 h rhythms in plants

Jean-Jacques d'Ortous, Lord of De Mairan (1678–1771) wrote the first publication, in the field of chronobiology [74] almost three centuries ago, and is still cited in scientific articles. He studied a sensitive plant which was given an evocative name of *Mimosa pudica*. The branches and leaves of *Mimosa pudica* always turn in the direction of where there is most light. He observed that the leaves and their peduncles fold themselves away and contract around sunset, in a similar manner as they do when the plant is touched or shaken. More importantly, this phenomenon takes place even when the plant is not in the sun or outdoors. This phenomenon is less pronounced when the plant is always kept in a dark place. The plant still opens up during the day and closes for the night at the same time every evening. De Mairan performed his experiment towards the end of summer and repeated it several times. The leaves rhythm persisted despite their lack of perception of night and day. This persistence is the signature of what would be known as circadian rhythms from the Latin *circa* (around) and *diem* (day) based on the terminology initiated by Franz Halberg in 1959. De Mairan suggested that diurnal variations in temperature, likewise in light, could synchronise circadian rhythms. Indeed, these two oscillating environmental factors are considered as “time givers” or “zeitgebers” of circadian rhythms. In the same epoch as De Mairan, Carl von Linné (1707–1778) constructed a “floral clock” noting the predictability of petal opening and closing times of various species of flowers (figure 1.1). Afterward, The Swiss botanist Augustin-Pyramus de Candolle (1778–1841) was the first to show that when plants are deprived of natural light, their “plant sleep” follows a circadian rhythm. Moreover, The German botanist and physiologist Wilhelm Pfeffer (1845–1920) performed similar experiments on many plant species, thereby confirming that leaf movements are not simply determined by the earth's rotation. Indeed, the leaf movements are not performed in an exactly 24-hour period and vary from species to species [67, 84, 168, 180].



Figure 1.1 – The flower clock was designed by Carl von Linne in 1745. The left part of the figure (6 AM–12 PM) depict when the petals of different species are opening; the right part (12 PM–6 PM) presents when the petals are closing. Reproduced from [306].

### 1.1.2 Modern chronobiology: the study of biological rhythms in mammals, insects, fungi and bacteria

Erwin Bünning provided the first evidence for the genetic basis of circadian rhythms generation by showing that period length is heritable in bean plants [45]. He also delivered a prominent hypothesis that circadian oscillators can be used to monitor seasonal changes in addition to measuring daily cycles and pointed out the adaptive importance of tracking seasonal changes. Thus, the field of circadian rhythms began from intense observation of plants. At the same epoch, the first observations of endogenously driven rhythms in bacteria [228], single-cell eukaryotes [127], insects [30], birds [179], rodents [276], primates [303], and humans [14] were discovered [180].

A major breakthrough in characterizing the genetic basis for rhythms generation was made by Ronald Konopka and Seymour Benzer in 1971 using a mutant screen in *Drosophila melanogaster* [172]. Mutagenized flies were investigated for the persistence of two circadian behaviors: locomotor activity and pupal eclosion. Flies presented one of three categorical mutant phenotypes: a shortening of the circadian period, a lengthening of the circadian period, or arrhythmia. All phenotypes were associated with a single locus, now referred to as the *Period* gene. Two years

## 1.1. Chronobiology: the study of periodic events in living organisms

---

after this discovery in fruit flies, the *Frequency* gene was shown to be essential for rhythms to persist in the filamentous fungus *Neurospora crassa* [97]. These unexpected results revealed that single-gene mutations could perturb a complex behavior and, together with the discovery of a heritable timing mutation in hamsters by Martin Ralph and Michael Menaker [262], provided the motivation for conducting a large-scale mutant screen in mice [225]. Subsequently, using a collection of strategies, dozens of clock genes have been identified in both prokaryotic and eukaryotic systems, including cyanobacteria, fungi, plants, insects, and mammals. Even human rhythms are significantly impacted by clock gene mutations [260]. An impressive common principle appears from inspecting these various clock gene systems. All organisms seem to have evolved transcriptional/posttranslational feedback loops to establish high-amplitude, near 24-hour, rhythms production. The development of real-time bioluminescent and fluorescent reporters, in combination with the identification of specific clock genes, has allowed the spatial resolution required to track rhythmicity at the single cell [129]. Such resolution is essential to study how individual oscillators are coupled within a population of rhythmic cells. However, there was a critical lesson to be learned from a decisive experiment performed using clock components of cyanobacteria. In 2005, Nakajima *et al.* [235] reconstructed a circadian oscillator in a test tube using only ATP and cyanobacterial proteins. This experiment showed that it is possible to construct a near 24-hour oscillator in the absence of gene transcription. Moreover, a non-transcriptional oscillator have been recently observed in human red blood cells, where peroxiredoxins, highly conserved antioxidant proteins, undergo 24-hour redox cycles, which persist for many days under constant conditions [244, 91].

### 1.1.3 Concepts and definitions in chronobiology

*This section will put forward several important concepts and definitions in chronobiology. These definitions are essential to the understanding of the circadian field and the variety of related research area.*

#### **Chronobiology**

Chronobiology is a domain of biology that studies periodic (cyclic) event in living organisms and their adaptation to solar- and lunar-related rhythms. These cycles are known as biological rhythms. Chronobiology derives from the ancient Greek "chronos", meaning "time", and biology, which refers to the science of life. Chronobiological studies consist non-exhaustively of physiology, genetics, molecular biology and behavior of organisms within biological rhythms mechanics. Other aspects of chronobiology cover development, reproduction, evolution, and ecology [73].

#### **Circadian rhythm**

A circadian rhythm is any biological process that displays an endogenous, entrainable oscillation with a period of about 24 hours.

#### **Infradian rhythm**

## Chapter 1. Introduction

---

An infradian rhythm is a biological rhythm with a period of oscillations longer than 24 hours. For instance, seasonal cycle (or circannual rhythm), or monthly variations as circalunar rhythm (one lunar cycle 29.5 days) or menstrual cycle.

### Ultradian rhythm

An ultradian rhythm is a biological rhythm with a period of oscillations shorter than 24 hours. For example, the developmental oscillator in somite segmentation in mouse embryo has a period of 2 hours. This process is called the segmentation clock [160].

### Period

The period is the length of time separating two consecutive maxima in a periodic signal. (figure 1.2).

### Phase

The phase is the instantaneous state of an oscillation within a period. It corresponds to the fraction of the wave cycle (measured in angular units) that has elapsed relative to the origin. The acrophase can be described as the time at which the peak of a rhythm occurs (figure 1.2). In general in this work we use the term phase to refer to the acrophase.

### Amplitude

We defined the amplitude as the ratio between the peak and the trough of a waveform cycle using the  $\log_2$  scale. Though in general, the amplitude is defined as half the value of the range of oscillation (the distance between the mesor and the peak, figure 1.2).

### Mesor

The mesor is defined as the average value around which the variable oscillates. The mesor is a circadian rhythm-adjusted mean based on the parameters of a cosine function fitted to the raw data (figure 1.2).

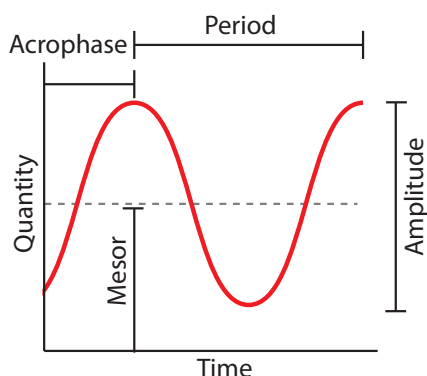


Figure 1.2 – Illustration of phase, amplitude, period and mesor definitions

## 1.1. Chronobiology: the study of periodic events in living organisms

### Phase shift

A phase shift is any change that occurs in the phase of a rhythm (e.g. in a gene, protein, metabolite), typically following a perturbation.

### Phase resetting curve (PRC)

Phase resetting is a behavior observed in different biological oscillators and plays a role in creating synchronization (or re-synchronization with the environment in the example of jet-lag). Phase resetting curves (PRCs) are measured by administering a precisely timed perturbation (light cues, food cues, drugs) to an oscillator and measuring the ensuing phase shift (figure 1.3). In other words, the PRC illustrates a change in the cycle phase of an oscillation induced by a perturbation as a function of the phase at which it is received. PRCs can be divided into two groups based. Type 1, or weak resetting, indicates perturbations in which the final phases cover all phases. Type 0, or strong resetting, indicates perturbation in which the final phases only cover a subset (typically a subinterval) of all phases [28].

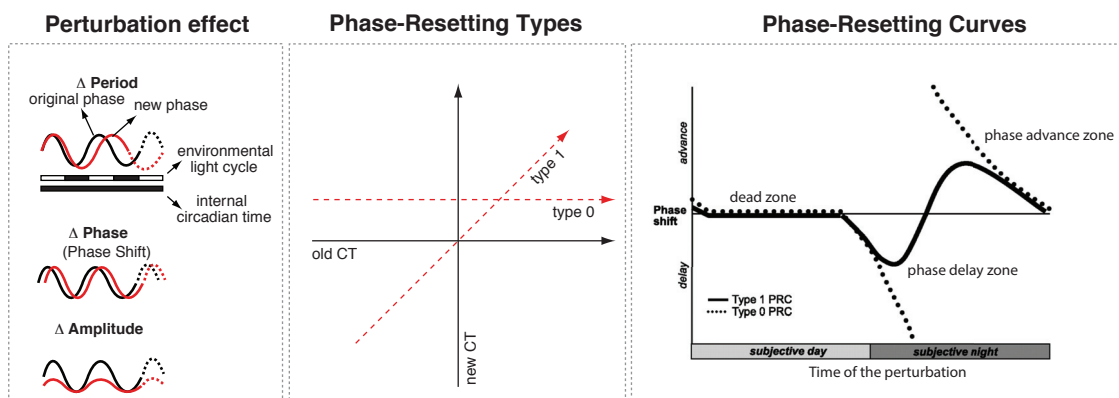


Figure 1.3 – Illustration of circadian disruption and phase resetting types. The phase response curve can be represented by plotting the time of the phase shift vs the perturbation. A PRC may include a "dead zone" where the perturbation has no effect, a "phase delay zone" or "phase advance zone". Modified from [28, 113].

### Transcriptional-translational delayed feedback loop (TTFL)

The transcriptional-translational delayed feedback loop is a gene regulatory network motif. In its minimal form an oscillation in gene expression can be produced and sustained. In a minimal TTFL, a protein *A* activates a gene *b*. The protein *B* is produced and represses the gene *a*. After a moment as gene *a* is not transcribed and the protein *A* is degraded, the gene *b* is not activated anymore. When the protein *B* is degraded the gene *a* is not repressed and can be transcribed again (figure 1.4). Delays are crucial in order to obtain an oscillatory behavior, and includes transcription to mRNA, splicing and post-transcriptional modification, translation to protein and post-translational modifications followed by trans-location to the nucleus. Mathematical models of TTFL represent an important area of the circadian research. TTFL are used to perform *in silico* knockout studies in order to understand core clock perturbation effects [305, 29, 194].

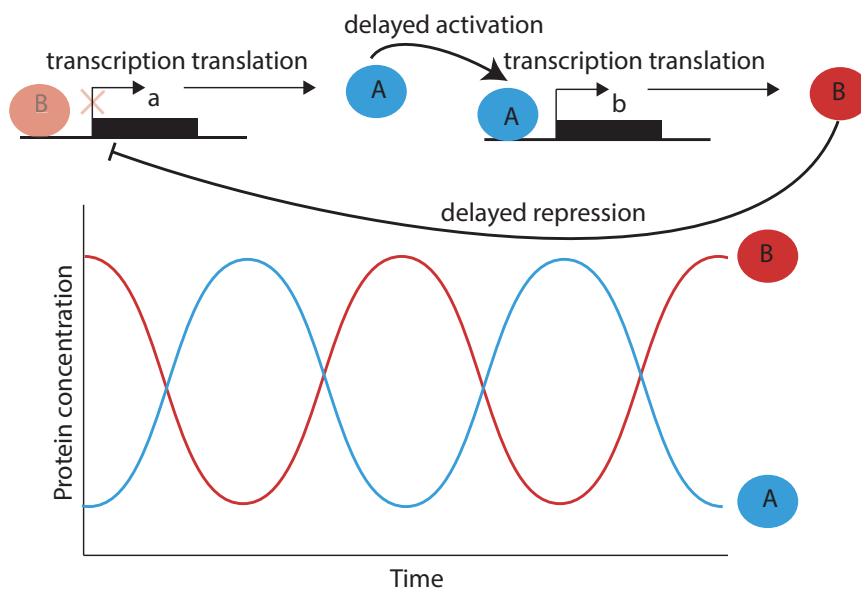


Figure 1.4 – Illustration of the transcriptional-translational delayed feedback loop principle.

### Entrainment

The entrainment is defined as the coupling of an endogenous rhythm to an environmental oscillator with the result that both oscillations have the same periods. In addition, the phase of the endogenous rhythm is affected by entrainment. For instance in Light-Dark (LD) conditions the circadian clock is entrained by light cues and consequently the period is exactly 24 hours. In contrast to the Dark-Dark (DD) condition, where the period is slightly longer.

### Zeitgeber

Zeitgeber is a german word meaning “time giver”. A cue (such as light or food) that entrains the circadian clock [89]. ZT refers to “zeitgeber time” and ZT0 typically corresponds to the switch

## 1.1. Chronobiology: the study of periodic events in living organisms

---

between dark and light, whereas CT refers to "Circadian Time". CT is used in free running conditions (no entrainment cues).

**Temperature compensation** A defining property of the circadian rhythm is that the period of the circadian rhythm is kept relatively constant within a physiological range of temperatures, which imply that the oscillator is temperature compensated. The mechanisms behind temperature compensation and temperature entrainment are not fully understood, neither biochemically nor mathematically [33].

**Clock controlled genes (CCGs)** Clock controlled genes are downstream core clock genes in the regulatory networks. These CCGs contain transcription factor binding sites (TFBS) related to the clock in their promoter or nearby enhancers.

**Chronotype** The chronotype is defined as the mid-sleep time. It allows to compare an individual to the population. A morning person has an early chronotype (larks profile) and an evening person has a late chronotype (owl profile). The chronotype can be assessed using standard questionnaires as the Munich Chronotype questionnaire (MCTQ) [282].

### **Actogram**

Actograms are used to represent the rhythm of an organism over the 24-h cycle. An actogram consists of digitized activity values that are displayed as a double plot, where a line contains data for that day as well as the preceeding day (figure 1.5). This visualization is used to monitor wheel-running activity of mice, or sleep/wake cycle in humans [89].



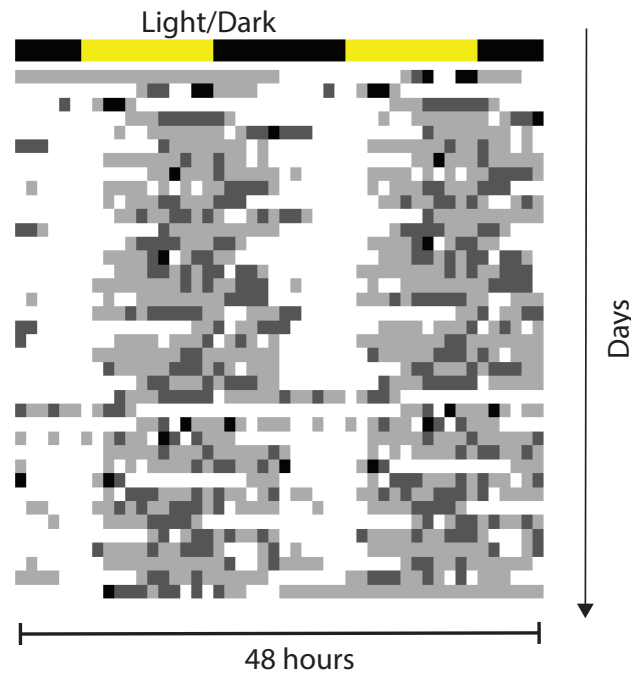


Figure 1.5 – Actogram of human activity measured using a smartphone over one month. The grey/black squares represent the intensity of the activity at a certain time (see appendix A.2).

#### 1.1.4 Human chronobiology

Our innate inclination for mornings or evenings is determined by the phase of our circadian rhythms. A recent study performed a genome-wide association study (GWAS) of self-reported morningness on 90,000 individuals, followed by analyses of biological pathways and related phenotypes. *Hu et al.* identified 15 significantly associated loci, including seven established circadian genes. Circadian and phototransduction pathways were enriched in their results. In addition, they reported that morningness was associated with insomnia, sleep phenotypes, body mass index and depression [140]. Several GWAS have also suggested interrelationships between clock gene variations and metabolism. For instance, *Cry2* was shown to be associated with a propensity to type 2 diabetes [86], while *Clock* haplotypes were associated with metabolic syndrome and obesity [297]. Single nucleotide polymorphisms (SNP) in *Clock* are associated with high plasma ghrelin concentration, obesity, altered eating behaviors, evening preference, short sleep [104, 105]. Intriguingly, recent GWAS studies also indicated that melatonin, a circulating hormone whose diurnal expression is tightly controlled by the circadian clock, and its G-protein-coupled receptors *MTNR1A* and *MTNR1B* are associated with the development of type 2 diabetes and impaired insulin secretion [86, 208, 288]. Melatonin, whose receptors are present in both the pancreatic islets and the SCN, has been shown to affect a collection of physiological processes, including glucose metabolism, insulin secretion and sleep [254].



## 1.1. Chronobiology: the study of periodic events in living organisms

Furthermore, a familial advanced sleep-phase syndrome was reported in 1999 as a short-period (FASPS) circadian rhythm variant in humans (figure 1.6) [260]. In this study, the authors have investigated three kindreds with a profound phase advance of the sleep-wake, melatonin and temperature rhythms associated with a very short period. The mutation segregates as an autosomal dominant with high penetrance. These kindreds served as a well-characterized familial circadian rhythm variant in humans and contributed to the genetic analysis of human circadian physiology.

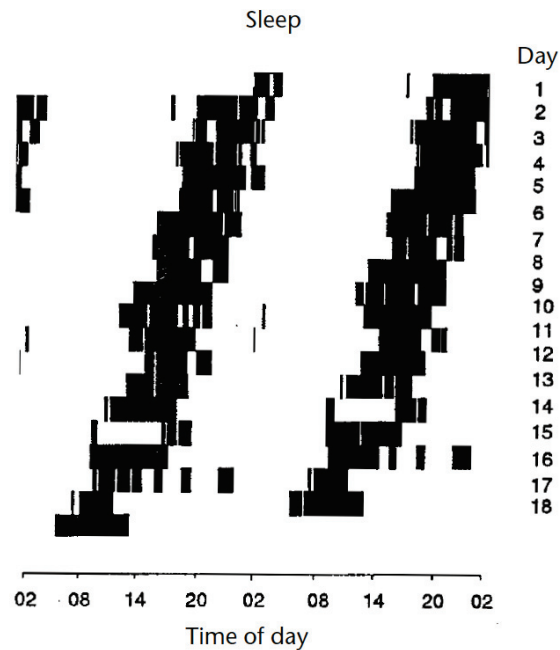


Figure 1.6 – Free-running period of the circadian sleep-wake and temperature rhythms in one FASPS subject. Sleep-wake rhythms of a 69-year-old female, studied in time isolation for 18 d. Data are double-plotted. Sleep data (filled bars) are derived from polygraphically-recorded sleep. Chi-squared periodogram analysis showed a free-running period of 23.3 h during the 18-day recording interval. Modified from [260].

To better understand the genetic basis of temporal organization in humans, a questionnaire was developed to document individual sleep times, self-reported light exposure, and self-assessed chronotype, considering work and free days separately [284]. This questionnaire is known as the Munich Chronotype Questionnaire (MCTQ). The authors of the study summarized the results of 500 surveys completed in a pilot experiment in 2003. Individual sleep times depicted large differences between work and free days, except for extreme early types. During the workweek, late chronotypes accumulated a substantial sleep debt, for which they counterbalanced on free days by prolonging their sleep by several hours. For all chronotypes, the quantity of time spent outdoors in broad daylight significantly impacts the timing of sleep: Increased self-reported light exposure advances sleep [284].

In a follow-up study using the MCTQ done by the same research group, they wanted to explore

## Chapter 1. Introduction

what zeitgebers entrain the human clock in real life by exploiting the common discrepancy between social time and sun time. At the time of the study, their MCTQ database comprised more than 40,000 individuals (including 22,000 Germans). The results of the study showed that the human clock entrains to sun time (figure 1.7). Within a given time zone, people live with respect to a common social time — which tells them, for example, when to go to work. Dawn and dusk, however, progress from East to West, generating a continuum in sun time. This continuum creates differences between, for instance, the actual mid-dark phase and midnight according to local clock time. In general, exposure to natural light decreased statistically in the bigger the city thereby weakening the strength of this zeitgeber. The fact that inhabitants of even larger cities are less coupled to sun time could be due to less exposure to outdoor light, resulting in a weaker zeitgeber strength of the natural light–dark cycle. Light and darkness even play a role when humans are entirely entrained by social cues: when we sleep, we close our eyes and in most cases, avoid light. That social cues alone cannot entrain the human circadian clock — without concurrent (behavioral) light changes [283].

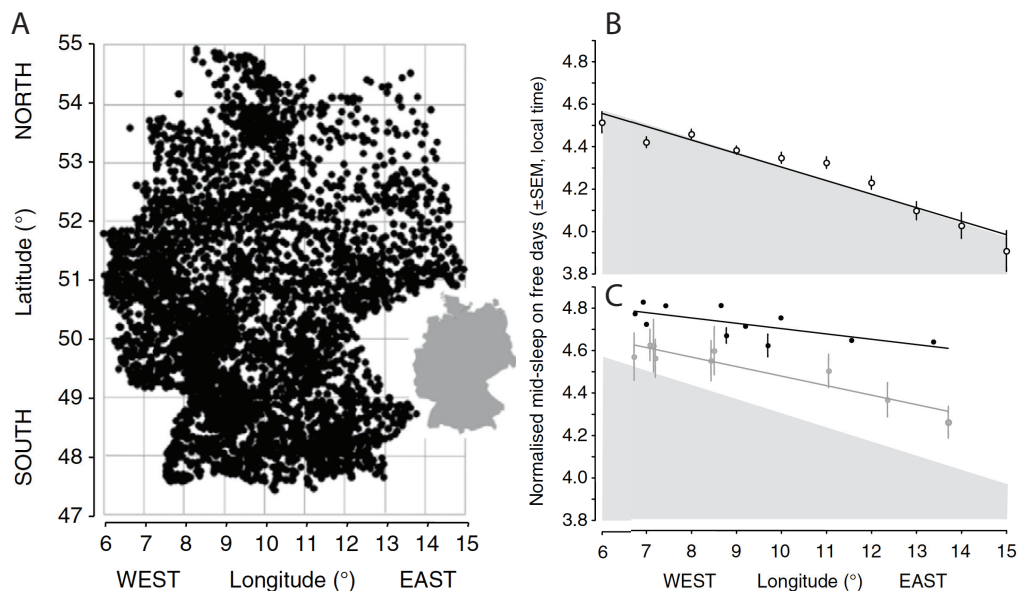


Figure 1.7 – **A**) The geographical distribution of the German places of residence of the study participants ( $N = 21,600$ ). **B**) The dependence on the longitude of normalized chronotype of people living in areas with up to 300,000 inhabitants, representing 82% of the German population. **C**) The dependencies of average chronotype on longitude for nine towns with populations between 300,001 and 500,000 (gray symbols) and 11 cities with more than 500,000 inhabitants (black symbols). All correlations are also highly significant for the respective raw, un-binned datasets. For reference, the slopes of the shaded areas in (B,C) reflect approximate discrepancies in dawn times at the different longitudes on June 21st. Standard error of the mean is present for all data points but often is smaller than the size of the dots. Reproduced from [283].

The effects of social cues on sleep remain largely unquantified. A recent study showed how a smartphone app can accurately collect data on sleep habits around the world [336]. The

## 1.1. Chronobiology: the study of periodic events in living organisms

---

authors observed that social pressures increase and/or cover biological signals in the evening, leading participants to delay their bedtime and shorten their sleep. They observed that a country's average bedtime, but not average wake time, predicted sleep duration. Moreover, they confirmed that women schedule more sleep than men and that users reporting that they are more exposed to outdoor light go to sleep earlier and sleep more than those reporting indoor light. Finally, they confirmed that age is the primary determinant of sleep timing and that age plays a critical role in the variability of population-level sleep habits.

A second smartphone app-based study revealed that erratic diurnal eating patterns in humans can be modulated for health benefits [112]. A diurnal rhythm of eating-fasting improves health, but the eating pattern of humans is rarely measured. Thanks to their app, they were able to monitor ingestion events in healthy adults with no shift-work for several days. Most subjects ate frequently and erratically throughout wakeful hours, and overnight fasting duration paralleled time in bed. They reported a bias toward eating late, with an estimated <25% of calories intake before noon and >35% after 6 p.m. "Metabolic jet lag" arising from weekday/weekend changes in eating pattern similar to travel across time zones was frequent. The daily consumption duration (95% interval) exceeded 14.75 hr for half of the cohort. When overweight individuals with >14 hr consumption duration ate for only 10–11 hr daily for 16 weeks supported by a data visualization (raster plot of dietary intake pattern called "feedogram") that they developed, study participants decreased their body weight, reported being energetic, and improved sleep.

The link between social jet lag and obesity has been extensively studied [88]. Social jet lag quantifies the differences that often arises between circadian and social clocks, which induces chronic sleep loss. Besides, the circadian clock regulates energy homeostasis, and its perturbation such as social jet lag, may participate to weight-related pathologies. A large-scale epidemiological study showed that beyond sleep duration, social jet lag is associated with increased BMI [281]. They demonstrated that living "against the clock" may be a factor leading to the epidemic of obesity. This is crucial in pending considerations on the implementation of Daylight Saving Time and on work or school schedules, which all contribute to the amount of social jet lag accrued by an individual. They suggested that improving the correspondence between biological and social clocks will help the management of obesity.

Circadian clocks affect nearly all facets of our physiology and behavior, including rest-wake cycle, cardiovascular activity, hormone secretion, body temperature and metabolism (figure 1.8). In summary, circadian desynchrony, a characteristic of shift work and sleep disruption in humans, also leads to metabolic pathologies [28]. Together, these epidemiologic and association studies highlight the clinical importance of interrelationships between melatonin, glucose homeostasis, circadian rhythms, and sleep.

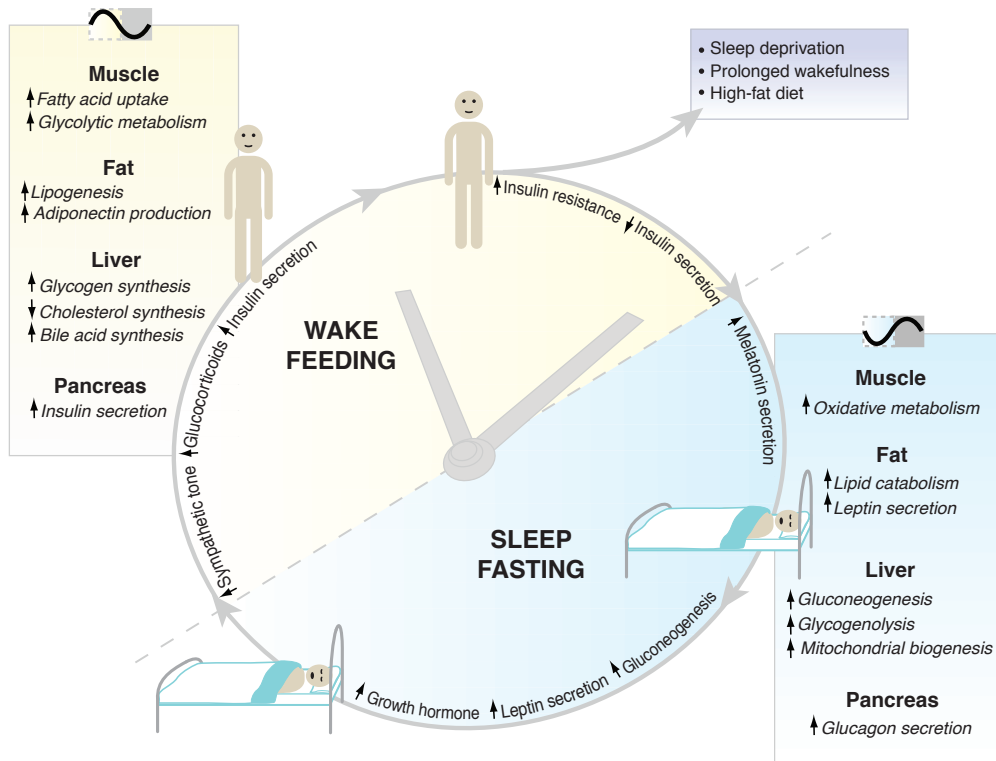


Figure 1.8 – The circadian clock controls suitable metabolic responses within peripheral tissues over the diurnal cycle. For instance, the liver clock promotes gluconeogenesis and glycogenolysis during the sleep/fasting period, while it promotes glycogen and cholesterol synthesis during the wake/feeding period. Correct functioning of peripheral clocks keeps metabolic processes in sync with the environment, which is necessary for maintaining the health of the organism. Different tissues depict distinct clock-controlled functions; In addition, clock defect in certain tissues will provoke opposing effects on the metabolic function as revealed by changes in nutrient conditions. Aging, diet, and environmental perturbations such as shift work may also impact the integration of circadian and metabolic systems. Reproduced from [28].

## 1.2 Chromatin structure and accessibility

*Since we aimed at investigating the transcriptional regulation of the clock, I have to present some general concepts about chromatin structure and transcriptional regulation, before describing the molecular organization of the circadian clock. In addition, technologies to investigate the chromatin landscape such as ChIP-seq and DNase I-seq will be introduced in this section.*

### 1.2.1 Chromatin structure and transcriptional regulation

Transcription is a highly regulated process (figure 1.9). It starts with the recognition of a piece of genomic DNA, defined as the promoter of a gene, by a set of transcription factors, which modify the local chromatin structure and recruit RNA polymerase II to produce mRNA [299]. Transcription factors (TF) bind to specific transcription factor binding sites (TFBS) that are either close or distant to a transcription start site (proximal/distal regulation). Sets of TFs can cooperate on cis-regulatory modules (CRM), stretches of DNA of about 100-1000 bp, to achieve specific regulatory events. One cis-regulatory element can regulate several genes, and one gene can have several cis-regulatory modules. Interactions between bound transcription factors and cofactors stabilize the transcription initiation process to allow gene expression [170]. The regulation conferred by sequence-specific binding is highly restrained by the three-dimensional structure of chromatin [50].

The basic units of chromatin are nucleosomes, which are formed by octamers of core histone proteins. Eukaryotic genomes are tightly wrapped around nucleosomes to achieve the 10,000–20,000-fold compaction necessary to fit a genome into the small volume of the nucleus but that must also allow proteins involved in transcription, replication, and repair to access DNA [342]. A thermodynamic equilibrium model has been used to describe experimentally observed genome-wide in vivo nucleosome occupancy arrangements. In this model, both nucleosomes and transcription factors have a constitutive affinity for DNA sequence, which is dependent on sequence composition [264]. Transcription factors compete with nucleosomes for DNA, and the thermodynamic equilibrium defines the configuration of nucleosomes and transcription factors. In addition, nucleosome presence is shaped by kinetic elements; in particular by chromatin-remodeling factors using the energy driven from ATP hydrolysis to actively influence DNA-histone interactions [62]. Several mechanisms of chromatin remodeling have been put forward, including nucleosome sliding, nucleosome eviction, and looping of DNA away from the histone core [207, 304]. Important aspects of the chromatin environment may include post-translational histone modifications, the composition of the nucleosomes themselves, and the binding of other protein complexes. Histone post-transcriptional modifications may affect transcription factor binding by enhancing the affinity of transcription factor related protein complexes for the modified histones or by diminishing the affinity of the histone for DNA.

The improvements of methods such as Chromatin Immuno-Precipitation followed by sequencing (ChIP-seq, figure 1.10), for the genome-wide mapping of individual histone modifications

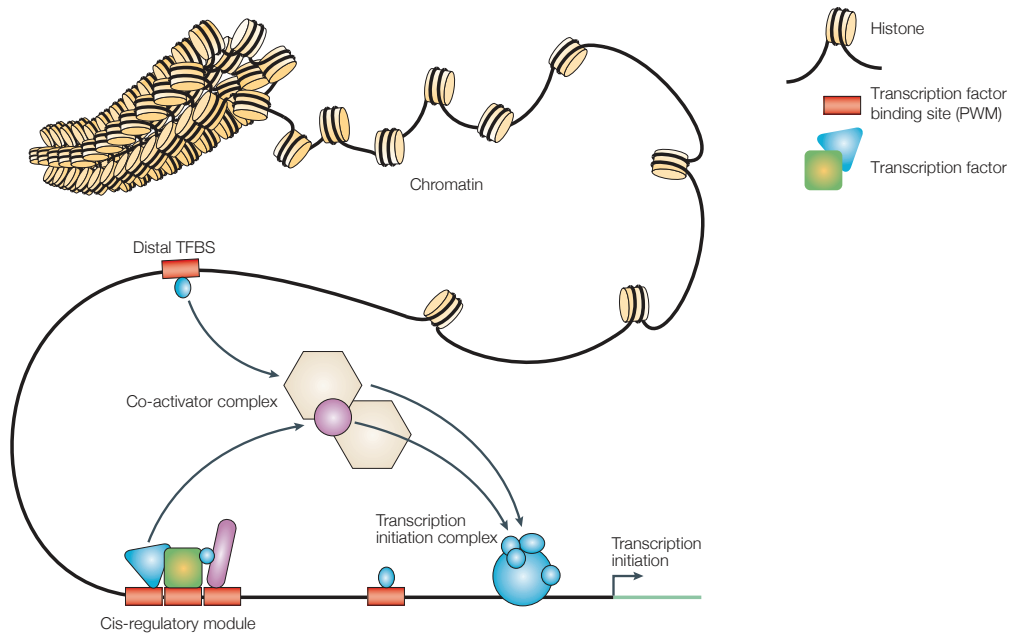


Figure 1.9 – Transcription factors (TFs) bind to specific locations (transcription-factor binding sites; TFBS) that are either proximal or distal to a transcription start site. Sets of TFs can act in functional cis-regulatory modules (CRMs) to achieve specific regulatory events. Interactions between bound TFs and cofactors stabilize the transcription-initiation complex to enable gene expression. The regulation that is allowed by sequence-specific binding TFs is highly dependent on the three-dimensional structure of chromatin. Reproduced from [340]

or DNA binding proteins, has enabled the detection of correlations between histone modification patterns or transcription factor binding and distinct states of gene activity [71, 358, 27]. Undoubtedly histone modifications can serve as a marker for a particular gene activity (or inactivity), but the question of causality remains (figure 1.11).

## 1.2. Chromatin structure and accessibility

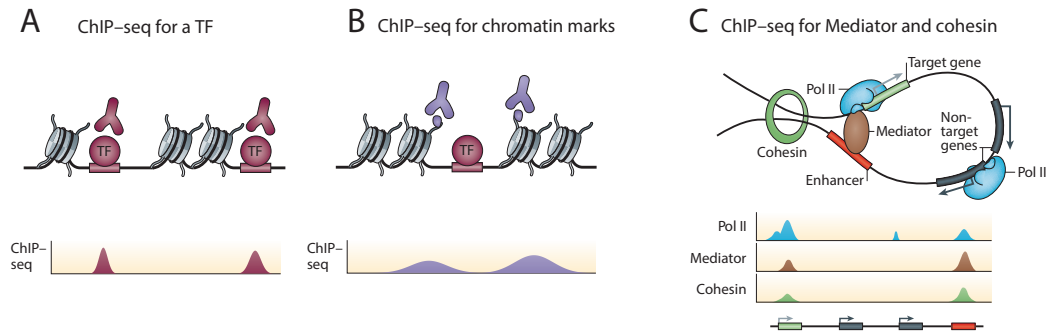


Figure 1.10 – Illustration of the ChIP-seq principle. The schema of the different methods (top panel of each part) and the matching data signature (such as read density) that is used for regulatory element investigation (bottom panel of each part) are presented. **A)** Chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) uses antibodies to find the location of transcription factor binding sites genome-wide. **B)** Nucleosomes that flank active enhancers carry characteristic histone modifications that can be identified by ChIP-seq using specific antibodies. **C)** Enhancers are brought into close proximity of their respective target promoters through the formation of chromatin loops, which are established by cohesin and Mediator complexes. ChIP-seq can detect the contact points of cohesin and Mediator from promoters and enhancers, and has been used to predict enhancers. Modified form [71]

The role of distal cis-regulatory elements, such as enhancers, and their interactions with promoters, is not well characterized in the circadian context. The identification of a core regulatory network of cis-regulatory modules, promoters, and corresponding transcription factors will provide a comprehensive picture of the regulation and the interplay between metabolism and internal clock.

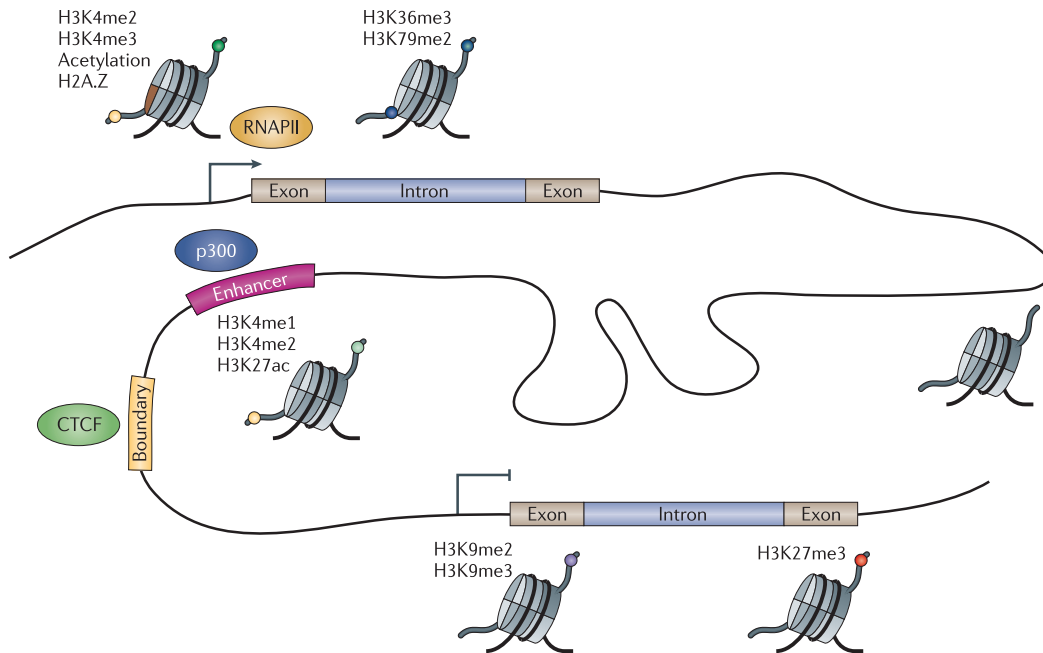


Figure 1.11 – Histone modification and chromatin organization. Promoters, gene bodies, enhancers and boundary elements are illustrated on a schematic genomic region. Active promoters are marked by histone H3 lysine 4 dimethylation (H3K4me2), H3K4me3, acetylation (ac), and histone variant H2A.Z. The structure of the nucleosome core may also determine nucleosomes as being more or less permissive to transcription factor binding. Histones that constitute nucleosomes cores variants, such as H2A.Z, have been reported to alter nucleosome properties [153]. Transcribed regions are enriched for H3K36me3 and H3K79me2. Repressed genes may be placed in large regions of H3K9me2 and/or H3K9me3 or H3K27me3. Enhancers are relatively enriched for H3K4me1, H3K4me2, H3K27ac and for the histone acetyltransferase p300. The transcriptional repressor CTCF binds many sites that may act as insulators, boundary elements or structural scaffolds. These various features of chromatin help to organize the DNA and determine functional elements. Reproduced from [358]



### 1.2.2 Measuring chromatin accessibility using DNase I hypersensitivity

DNase I hypersensitivity is a measure of chromatin accessibility and its dynamics. It provides a general method for predicting cell-type specific cisomes. DNase I hypersensitive sites (DHS), small regions of chromatin of 200-1000 bp, which are extremely sensitive to cleavage by DNase I, typically occur in nucleosome-depleted regions and often arise as the result of transcription factor binding. DNase I digestion followed by high throughput sequencing has developed into a powerful technique to identify genome-wide DHS [199, 124, 302]. Because transcription factor binding sites favor DNase I hypersensitivity yet DNase I-seq does not require a factor specific antibody, DNA sequence motif analysis on DHS data has been highlighted as a method to identify the binding sites of multiple transcription factors in a single experiment [257, 37]. While DNA sequence motifs may not be entirely predictive of transcription factor binding across the entire genome, they are good predictors within regions of open chromatin [131]. Thus, DNase I hypersensitivity patterns are likely to be dependent on the structure of TF-DNA interactions and the chromatin environment at binding sites. Furthermore, at a high sequencing depth, it is possible to identify depleted narrow regions in the DHS core, corresponding to protein footprints of 8-30 bp [138, 57] (figure 1.12). Differential DNase I hypersensitivity analysis proved the utility of quantitative analysis of chromatin accessibility differences between condition to predict transcription factor activity [131].

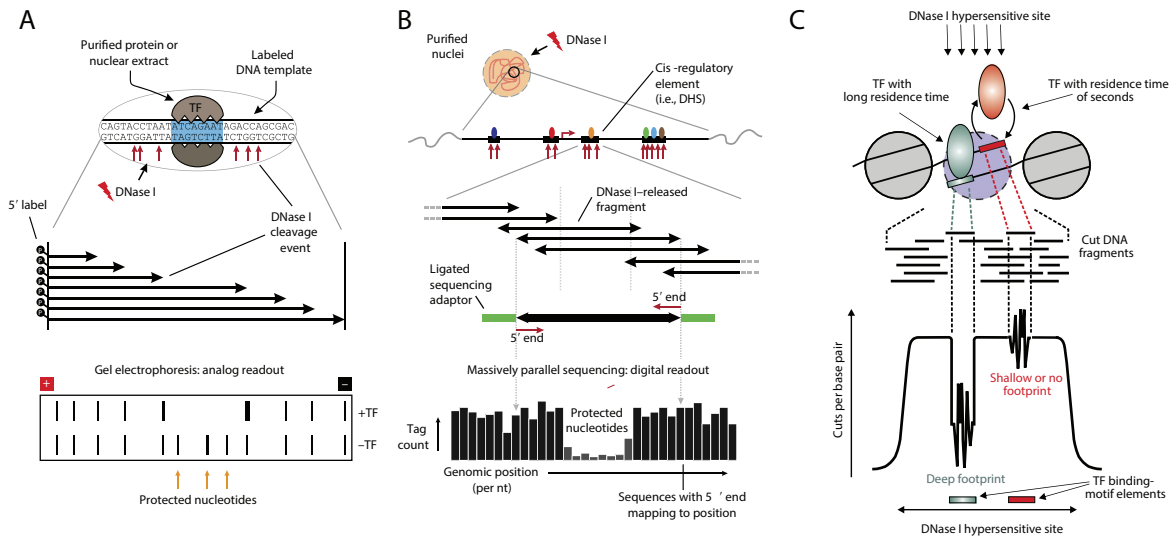


Figure 1.12 – Principles of a DNase I footprinting experiments. **A)** The classical DNase I footprinting method was executed *in vitro* and combined purified protein or nuclear extract with a radiolabeled DNA probe. A shallow DNase I digestion resulted in a series of fragments that were separated using gel electrophoresis. **B)** DNase I-seq combines exposure of nuclei to DNase I, purification of small DNase I-released fragments, and massively parallel sequencing of fragment ends (DNase I cleavage sites) to generate a digital signal of per-nucleotide cleavages genome-wide. **C)** DHSs versus TF footprints. An accessible site is detected as a DHS enriched for sequencing reads in DNase I-seq signal. In the DHS, one or more narrow sites may be identified as putative TF footprints thanks to the local protection from DNase I cleavage. The identity of TFs is inferred from the sequence underlying the corresponding protected regions. Represented here are two example TFs with different DNA-binding dynamics that influence the degree of protection from DNase I cleavage at the binding sites. The DNase I cleavage pattern can be present over motif elements with deep, very shallow or no footprints. Modified from [331, 224]

### 1.3 The organization of the mammalian circadian clock

*This section will introduce the structure of the mammalian molecular clock, Moreover, a review of recent transcriptomics and functional genomics studies as well as computational research (modeling and promoter sequence analysis) of the transcriptional regulation of circadian genes will be presented. Finally, molecular mechanisms that are implicated in the cross-talk between the circadian clock and the metabolism will be described.*

#### 1.3.1 The hierarchical organization of the mammalian clock

In mammals, the main zeitgeber is the light signal. Daylight is detected by photoreceptors in the retina and transmitted to the suprachiasmatic nucleus (SCN) in the anterior hypothalamus by the retinohypothalamic tract (RHT). Reception of the light signal rapidly induces the expression of the *mPer1* or *mPer2* genes in a time-dependent manner. Thus, light input at the onset of the night activates both *mPer1* and *mPer2*, whereas *mPer1* only is activated if a stimulus is provided at the end of the night. This activation is mediated through pathways implying cAMP, Ca<sup>+2</sup> and phosphorylated CREB which activates the *Per* genes [272].

The SCN is the master circadian oscillator, where the rhythm is generated, and then converted to systemic signals such as hormones like the glucocorticoids, which affect the metabolic processes and behavior of the entire animal [280] (figure 1.13). Ablation of the SCN leads to a loss of phase coordination between the different tissues [351]. Therefore, SCN is crucial for the proper function of the circadian clock in the whole animal.

Indirect signals such as temperature cycles can also induce synchronization. The peripheral oscillators are sensitive to temperature entrainment via the heat-shock response pathway. Several studies showed that heat-shock protein (HSP) and the transcription factor heat-shock factor 1 (HSF1) are involved in the mechanism of clock resetting by temperature cycles [291, 175].

Another indirect signal such as feeding rhythm can entrain peripheral oscillators such as in the liver. The intracellular redox state (assessed by the NADH/NAD<sup>+</sup> ratio) is a direct measure of the nutritional state of a cell. The redox state affects the activity of the core clock transcription factor BMAL1/CLOCK through the interaction with the NAD<sup>+</sup> dependent deacetylase SIRT1 [15].

In addition, the presence of mammalian peripheral clocks has been demonstrated by recording gene expression in cultured fibroblasts or tissue explants [25]. The circadian clock is involved in the regulation of several biological processes. For instance sleep-wake cycles, heartbeat, blood pressure, renal plasma flow, body temperature, sensorial perception, and the secretion of many hormones fluctuate during the day in a coordinated manner [128]. These physiological changes persist under constant conditions, which suggests their regulation by one or more circadian regulators.

The hierarchical synchronization of the peripheral clocks supports the coordination of the different tissues. However, due to tissue-specific functions, the group of clock-controlled genes are different between tissues. The molecular mechanisms underlying tissue specificity are not determined yet but, several studies showed significant differences in the nature, the number and the phase of the rhythmic transcripts in various tissues [312, 248, 5, 355]

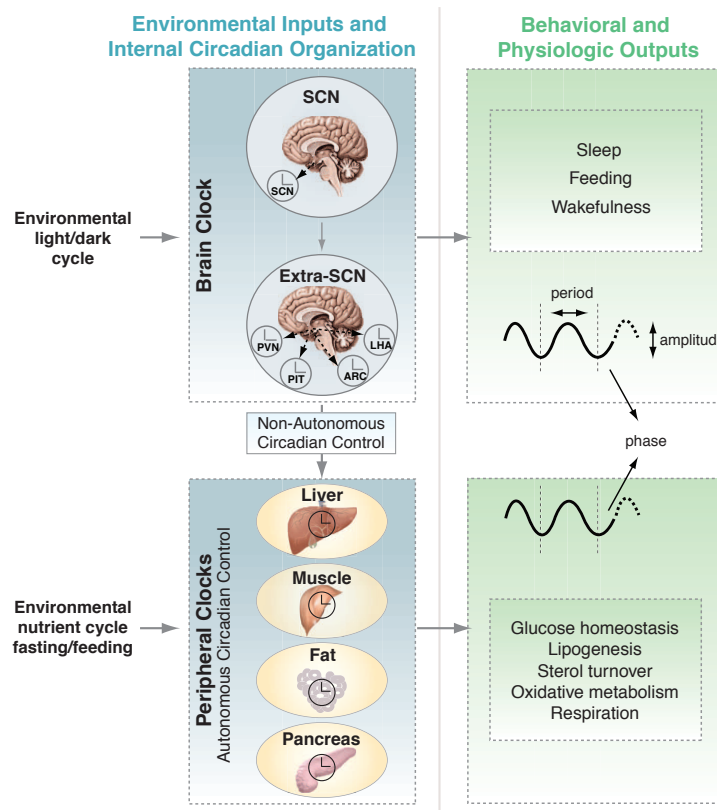


Figure 1.13 – The hierarchical structure of the mammalian clock. Central and peripheral clocks coordinate external cues with behavior and metabolic outputs. Light entrains the master pacemaker in the SCN, which in subsequently synchronizes extra-SCN and peripheral clocks. Brain clock outputs involve behavioral rhythms such as sleep and feeding. In the other hand, peripheral clock outputs consist of metabolic rhythms such as glucose and lipid homeostasis. The hierarchical structure of the mammalian clock is described by “non-autonomous” regulation of peripheral tissue clocks meaning the regulation of peripheral tissue oscillators through direct neural and humoral signals, and “autonomous” control indicating the intrinsic regulation of local cellular oscillators independently of the brain clock. Modified from [28]

#### 1.3.2 The mammalian molecular core clock

At a molecular level, the circadian oscillator consists of two interacting transcriptional-translational feedback loops (figure 1.14). The PAS domain helix-loop-helix transcription factors BMAL1 and CLOCK are the principal constituents of a positive feedback loop [41, 167, 225, 139]. They form a heterodimer, which binds to the E-box with a consensus sequence 5'-CACGTG-3'. This E-box is a DNA binding site, known as a cis-regulatory element of target genes such as Period genes (*Per1*, *Per2* and *Per3*) and Cryptochromes (*Cry1* and *Cry2*). A negative feedback loop is formed when CRYs and PERs form heterocomplexes, which are translocated back to the nucleus and inhibit their own expression [182, 54, 191]. A secondary feedback loop is mediated by the orphan nuclear receptors REVERB $\alpha$  and ROR $\alpha$ , which are also controlled by the CLOCK/BMAL1 heterodimer. There is a competition in the nucleus between REVERB alpha and ROR alpha for binding to the ROR-responsive element (RRE) with a consensus sequence 5'-AGGTCA-3' preceded by a 5-bp A/T-rich sequence, in the *Bmal1* promoter. ROR $\alpha$  activates the transcription of *Bmal1*, whereas REVERB $\alpha$  represses it [259, 4].

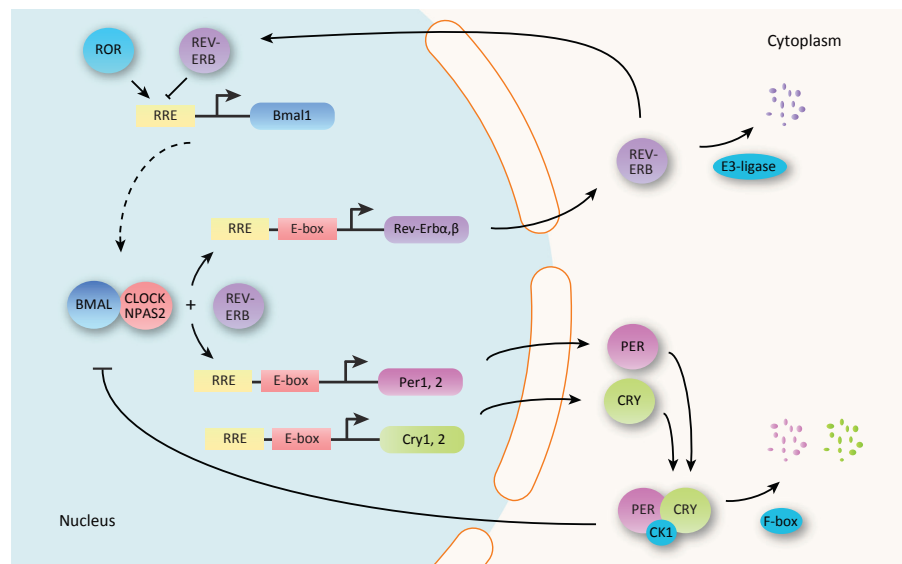


Figure 1.14 – The key components of the mammalian transcription/translation delayed feedback loop (TTFL) model. Both BMAL/CLOCK(NPAS2) and REVERBs contribute to the production of rhythmic transcription at the *Per*, *Cry*, and *Reverb* promoters. PER and CRY proteins are degraded via F-box proteins, whereas REVERB depends on E3 ligases. Reproduced from [274]

In addition to these TTFL, the clock requires numerous post-translational modifications to achieve a sufficient delay to set a 24h period. Protein kinases, notably CK1 $\epsilon$  and CK1 $\gamma$ , modulate the activity and/or stability of positive and negative components of the feedback loop [203, 204, 332]. The general mechanism of the circadian clock and the genes required for its maintenance are mostly conserved between different tissues. Nevertheless, circadian output genes are tissue-specific, as one would expect given the diverse physiologies regulated by the clock [146]. Metabolically active tissues, such as liver, have their own clock that regulates tissue-specific functions [48].

## 1.4 Circadian transcription and gene regulation

### 1.4.1 Transcriptome-wide studies of the circadian clock

To study circadian regulation of gene expression, several groups have performed transcriptome-wide studies to detect circadian transcription, using mRNA microarrays, RNA-seq or nascent RNA-seq time-series with two to four hours resolution over several cycles (or replicates) of 24 hours in mouse liver.

Roughly 2-20 percent of transcripts undergo circadian oscillation in the liver (table 1.1), including enzymes and regulators of major metabolic processes [248, 312]. Though this percentage depends mainly on the applied criteria, such as the set of genes (all the genes of the organism or only the expressed genes in a given condition), cutoffs considered, experimental conditions, such as feeding regimen and Light-Dark (LD) or Dark-Dark (DD), temporal resolution and the number of replicates. However, even in the liver, the number of cyclic mRNAs with large amplitudes (larger than two-fold) is relatively low.

Table 1.1 – Non-exhaustive list of circadian transcriptome studies in mouse liver based on a collection of techniques. About 2-20 percent of transcripts undergo circadian oscillation in the liver.

Method(s)	#rhythmic genes	Reference
Microarray	335	Panda <i>et al.</i> (2002) [248]
Microarray	575	Storch <i>et al.</i> (2002) [312]
Microarray	716	Miller <i>et al.</i> (2007) [227]
Microarray	3667	Hughes <i>et al.</i> (2009) [144]
Microarray	2997	[333]
RNA-seq	1160	Vollmers <i>et al.</i> (2009) [334]
Microarray	2828	Eckel-Mahan <i>et al.</i> (2012) [90]
RNA-Seq	3408	Koike <i>et al.</i> (2012) [171]
PoI II ChIP-Seq	892	Le Martelot <i>et al.</i> (2012) [188]
Microarray	885	Le Martelot <i>et al.</i> (2012) [188]
Nascent-Seq	822	Menet <i>et al.</i> (2012) [223]
RNA-Seq	1204	Menet <i>et al.</i> (2012) [223]
RNA-Seq	1630	Du <i>et al.</i> (2014) [83]
RNA-Seq and Microarray	3186	Zhang <i>et al.</i> (2014) [355]
GRO-Seq	1261	Fang <i>et al.</i> (2014) [95]
RNA-Seq	4544	Atger <i>et al.</i> (2015) [17]

In a landmark publication, Panda *et al.* compared the circadian oscillation in the liver and the suprachiasmatic nuclei (SCN) of the hypothalamus. They observed that about 650 genes were oscillating in both tissues, while the overlap between the two set of genes was quite low with only 27 genes. Nutrient metabolism, in liver, appeared as an important target of circadian regulation. Furthermore, genes related to metabolism of sugar were expressed

#### 1.4. Circadian transcription and gene regulation

---

during feeding in the early evening, and enzymes for cholesterol biosynthesis peaking at night, when nutrients are absorbed. In particular, so-called "rate-limiting" steps were shown to be subject to a precise circadian control [248]. Panda *et al.* suggested that both core clock mechanisms and feeding cues regulate these circadian patterns of gene expression.

Storch *et al.* compared the heart and the liver and transcriptome using mRNA microarray. They observed that about 10% of all expressed genes in those tissues depict a circadian oscillation. They found that the overlap between the oscillating transcripts in both tissues was low, notably a core set of 37 genes with similar circadian regulation in both tissues. The authors observed that the distributions of circadian phases in the two tissues are considerably different as well as the downstream biological processes enriched in their gene ontology analysis.

In addition, Miller *et al.* compared liver and skeletal muscle from WT and *Clock* mutant mice [227]. In WT tissue, they observed that a large percentage of expressed genes were transcription factors that were rhythmic in either muscle or liver, but not in both, suggesting that tissue-specific output of the pacemaker is regulated at a transcriptional level. The comparison between WT and *Clock* mutant mice tissues revealed that the *Clock* mutation affects the expression of many genes that are rhythmic in WT tissue, but also deeply affects many non-rhythmic genes. In both liver and skeletal muscle, a significant number of CLOCK-regulated genes were associated with the cell cycle and cell proliferation.

Another study by Hughes *et al.* compared oscillating transcription from mouse liver, NIH3T3, and U2OS cells [144]. The authors reported a 100-fold difference in the number of cycling transcripts in autonomous cellular models of the oscillator versus tissues harvested from intact mice. Moreover, they observed two clusters of genes that cycle with a period of 12 hours and 8 hours in liver, but not cultured cells. They confirmed that 12-hour oscillatory transcripts occur in several other peripheral tissues as well including heart, kidney, and lungs and the reported that these harmonics are lost in cell cultures, as well as under restricted feeding conditions.

Vollmers *et al.* were interested in understanding the main drivers of oscillations in hepatic transcripts in dark-dark conditions [333]. They wanted to study the contribution of the endogenous oscillator and the food intake. To address this question, they used distinct feeding and fasting experimental conditions on wild-type (WT) and circadian clock-deficient mice. They monitored temporal patterns of feeding and hepatic transcription. They reported that in the absence of feeding, only a small subset of transcripts continued to exert circadian patterns. Conversely, temporally restricted feeding restored rhythmic transcription of hundreds of genes in oscillator deficient mouse liver. They concluded that both temporal pattern of food intake and the circadian clock drive rhythmic transcription.

In recent past, Eckel-Mahan *et al.* compared the hepatic circadian transcriptome with a comprehensive dataset of over 500 metabolites identified by mass spectrometry [90]. They observed that many metabolites depict clock-controlled oscillation, including those within the amino acid and carbohydrate metabolic pathways as well as the lipid, nucleotide, and



xenobiotic metabolic pathways.

Menet *et al.* have investigated rhythmic transcription by quantifying genome-wide nascent mRNA produced around the clock in mouse liver [223]. They compared nascent mRNA with the amount mRNA expressed. Although the authors found that many genes exhibit rhythmic mRNA expression in the mouse liver, about 70% of them did not show comparable transcriptional rhythms. As in Le Martelot *et al.* They conclude that the post-transcriptional regulation must have a critical role in the circadian system. In addition, Menet *et al.* observed that the influence of CLOCK:BMAL1 differed from what was expected, which suggests that it collaborates with several other transcription factors to control the transcription of target genes.

As post-transcriptional mechanisms seems to contribute to mRNA abundance rhythms. Du *et al.* have investigated how microRNAs (miRNAs) affect the core clock and clock-controlled gene expression using mice with an inducible defect in miRNA biogenesis [83]. The authors observed that cyclic transcription paired with miRNA-mediated regulation was a frequent phenomenon that affected up to 30% of the rhythmic transcriptome and served to post-transcriptionally adjust the phases and amplitudes of rhythmic mRNA accumulation. Nevertheless, only few mRNA rhythms were actually generated by miRNAs. They hypothesized that miRNAs function to adapt clock-driven gene expression to tissue-specific requirements.

Using a similar technique to nascent mRNA sequencing called Global Run-On sequencing (GRO-seq) Fang *et al.* studied the underlying mechanisms of multiple phases of gene expression in the liver [95]. Their examination of enhancer RNAs (eRNAs) that cluster in specific circadian phases lead to the identification of functional circadian enhancers driven by distinct transcription factors (TFs). In addition, on a global scale some components of the TF cistromes, such as D-box, E-box, and RRE, that function to orchestrate circadian gene expression were identified. Integrated genomic analyses also revealed putative mechanisms by which a single circadian factor controls opposing transcriptional phases.

Another impressive study was performed by Zhang *et al.* to elucidate the role of the circadian clock in mouse physiology and behavior [355]. The authors performed RNA-seq and mRNA microarray to quantify the transcriptomes of 12 mouse organs around the clock. They reported that 43% of all protein coding genes showed circadian rhythms in transcription one or more tissue, but mainly in an organ-specific fashion. In most organs, the authors reported the characteristic bi-modal phase distribution with many oscillating genes peaking during transcriptional "rush hours" preceding dawn and dusk (as in Vollmers *et al.* and Le Martelot *et al.*). Moreover, by looking at the genomic landscape, the authors reported that rhythmic genes clustered together, were longer, and had more spliceforms than non-rhythmic genes. Systems-level analysis depicted a complex rhythmic orchestration of gene pathways throughout the body. In addition, they observed rhythmic expression of more than 1,000 known and novel noncoding RNAs (ncRNAs). Importantly, the authors claim that the majority of drugs and World Health Organization essential medicines directly target the products of rhythmic genes,



## 1.4. Circadian transcription and gene regulation

---

which may represent a turning-point for advancement in chronotherapy.

More recently, Atger *et al.* measured transcription, accumulation, and translation, of mouse liver mRNAs under light–dark conditions and ad libitum or night-restricted feeding in WT and Bmal1 knockout animals [17]. The authors reported that rhythmic transcription mostly leads to rhythmic mRNA accumulation and translation. Comparison of wild-type and Bmal1 KO mice confirmed that circadian clock and feeding rhythms have large effect on rhythmic gene expression, Bmal1 deletion affecting both transcriptional and post-transcriptional levels unexpectedly.

### 1.4.2 Genome-wide studies of the circadian chromatin landscape

Thanks to the development of recent sequencing-based technologies, as Chromatin Immunoprecipitation followed by sequencing [27], it is possible to study the chromatin landscape at a genome-wide scale, including transcription factor binding sites and histones modifications. Several studies in the circadian field used this approach in order to investigate the transcriptional regulation of the core clock as well as transcription factor related to feeding/fasting cycle.

For instance, Koike *et al.* [171] performed an impressive large-scale study using ChIP-seq for genome-wide temporal profiling of core clock factors such as BMAL1, CLOCK, CRY1, CRY2, PER1, PER2 and NPAS2. Moreover, they measured PolII loadings (RNAPII-8WG16) and elongating PolII (RNAPII-Ser5P), as well as coactivators p300 and CBP, and histone variants, such as H3K27ac, H3K9ac, H3K4me3, H3K4me1, H3K36me3, and H3K79me2. In addition, they performed a whole transcriptome RNA-seq with the same temporal resolution. The phase of activity of the transcriptional regulators displayed expected specificity, with CLOCK, BMAL1 and NPAS2 binding at ZT6-ZT8, PER1, while PER2 and CRY2 were reported at ZT15-ZT17, and CRY1 binding at ZT0. For the circadian transcriptional control, Koike *et al.* suggest that an activation phase from the early inducer BMAL1/CLOCK dimer recruits p300 and promote histone acetylation at regulatory elements, including enhancers marked by H3K4me1. Subsequently, transcription then occurs with nascent transcription starting around ZT13 to peak around ZT15, followed by a repression phase by PER1/2 and CRY2. In addition, they showed that only 22% of cycling mRNA are driven by *de novo* transcription, suggesting that both transcriptional and posttranscriptional mechanisms underlie the mammalian circadian clock. Moreover, they found that circadian modulation of RNAPII recruitment and chromatin remodeling occurs on a genome-wide scale.

At the same time, Vollmers *et al.* used genome-wide temporal profiles of H3K4me1, H3K4me3, H3K36me3, H3K27ac, H3K9ac, DNA methylation and strand-specific RNA sequencing to study epigenetic modifications associated with transcript oscillations [334]. They observed that 9% of expressed transcripts undergo circadian oscillation, notably protein-coding genes as well as lincRNAs, miRNAs and antisense transcripts such as *asPer2*. Transcripts phase distribution was bimodal with a cluster of genes expressed in the morning and a second cluster expressed in the

evening. Robust transcript rhythms were correlated with rhythmic histone modifications in promoters, gene bodies, or enhancers, while promoter DNA methylation levels were constant. Coupling of cycling histone modifications with nearby circadian transcripts consequently established a temporal relationship between enhancers, genes, and transcripts on a genome-wide scale in mouse liver.

Le Martelot *et al.* performed a genome-wide analysis of PolII, H3K4me3 and H3K36me3, combined with mRNA microarray [188]. PolII oscillations in the promoter were correlated with Pol II loading in the transcribed region. In addition, H3K4me3 in the promoter depicted a phase delay of 1.3 hours, while H3K36me3 signal oscillated at the end of a transcript with a phase delay of 3 hours. Moreover, Pol II occupancy preceded mRNA accumulation by 3 hours, consistent with mRNA half-lives. They observed, as in Vollmers *et al.* that transcripts phase distribution was bimodal with a group of genes expressed in the morning and a second group expressed in the evening. Finally, they were able to distinguish three classes of genes using modeling profiles of Pol II occupancy and mRNA accumulation. The first one showing rhythmicity both in transcriptional and mRNA accumulation, the second class with rhythmic transcription but flat mRNA levels, and a third with constant transcription but rhythmic mRNAs. The last class suggested post-transcriptional regulation in the mouse liver.

Several studies have focused on one or more transcription factors or DNA binding proteins to investigate the transcriptional regulation in the liver. Current models of time-specific transcription include the core clock transcription factors BMAL1/CLOCK that activate transcription maximally at ZT8 [273, 171], as well as the nuclear receptors RORs and REVERBs, whose targets are maximally transcribed around ZT20 [60, 43, 316]. Rhythmically active TFs also include clock-controlled outputs, notably the PAR-bZIP proteins (DBP, TEF, HLF), maximally active near ZT12 [277, 103]. Furthermore, diurnally fluctuating systemic signals may drive rhythmic TF activities, for example, HSF1 shuttles to the nucleus and activates transcription at ZT14 [270, 291], and similarly, SRF shows activity at the night-day transition [109]. Moreover, regulators controlled by feeding-fasting cycles include FOXO TFs that are active during the day, CREB/ATF family members at the light-dark transition, and SREBP during the night [333, 111]. Finally, the glucocorticoid receptor (GR) signals the onset of behavioral activity (light-dark transition) [266].

### 1.4.3 Chromatin remodelers and topological domains

Chromatin remodelers including histone-modifying enzymes, such as MLL1 [163], MLL3 [327] and JARID1A [79] have been lately associated with the circadian rhythm. In addition, CLOCK has been shown to possess a histone acetyltransferase activity [80]. Constitutive binding of BMAL1 in the presence of HDAC inhibitors was impaired which imply that proper chromatin environment is needed for correct BMAL1 binding [313]. While Menet and colleagues reported that CLOCK:BMAL1 can exhibit a pioneer-like function and thus open the chromatin [222]. These chromatin remodelers or some pioneer factor are potentially responsi-

ble for the mechanism of tissue-specific gene regulation in cooperation with core clock factor, but it is still an open question. Finally, the 3D structure of the chromatin or more precisely topological associated domains (TADs) have been shown to be dynamic along the circadian cycle [3] although the precise mechanisms and the role of these TADs remain unclear.

### 1.4.4 Computational analysis of the transcriptional regulation of oscillating genes

Bozek *et al.* addressed the question of tissue and phase specificity using a large scale promoter analysis of clock-controlled genes (CCGs) [38]. They performed a meta-analysis of microarray data from mouse tissues, notably heart, liver, SCN, skeletal muscle, by looking at promoter regions of 2065 CCGs for highly enriched transcription factor binding sites (TFBS). Many of the enriched transcription factors displayed themselves circadian oscillations. As expected TFBS associated with the clock were overrepresented such as CLOCK/BMAL1, DBP, HLF, E4BP4, CREB, ROR $\alpha$  and the lately described regulators notably HSF1, STAT3, SP1 and HNF4 $\alpha$ . Moreover, additional putative circadian transcriptional regulators were enriched such as PAX-4, C/EBP, EVI1, IRE, E2F, AP1, HIF1 and NFY. In addition, GC-rich motifs (SP1, EGR, ZF5, AP2, WT1, NRF1) and AT-rich motifs (MEF2, HMG1Y, HNF1, OCT1) were significantly overrepresented in promoter regions of CCGs. The authors suggested putative tissue-specific TFBS such as HNF3 (FOXA) for liver, NKX2.5 for heart or Myogenin for skeletal muscle.

In a subsequent study from the same authors, they used a similar approach but focused on mouse liver [39]. Again motifs such as E-boxes (CLOCK/BMAL), D-boxes (DBP, HLF, E4BP4) and cAMP responsive elements (CREB) appeared as enriched TFBS in the promoter of circadian genes. In addition GC-rich motifs (SP1, E2F, NRF1), AT-rich motifs (TBP, FOXO4, MEF2), Y-box motifs (NF-Y, C/EBP) and cell cycle regulators (E2F, ELK1) were among the enriched factors. In a subset of genes regulated by systemic signals such as hormones and body temperature, they were able to detect motifs of the serum response factor (SRF) and the estrogen receptor (ER). Finally, they confirmed their predictions using published ChIP-seq datasets by demonstrating that some of their predicted circadian transcriptional regulators (C/EBP, E2F, HNF1, MYC, MEF2) targeted clock-controlled genes.

Two studies on core clock motifs, notably D-box, E-box and RRE, proved their phase-setting ability [325, 326]. In the first study, using computational approach and synthetic constructs, the authors were able to produce high amplitude circadian oscillations *in vitro* and *in silico*. Ukai-Tadenuma *et al.* performed an *in vitro* assay with mammalian cells to test the role of motifs in promoter regions driving the expression of a reporter gene at a particular phase. They hypothesized that the control of downstream output by the circadian cycle was determined by the phase of expression of activators and repressors, targeting regulatory elements. The activator and the repressor were placed under the control of three core-clock elements, notably D-box, E-box or RRE. A morning phase for the activator (under E-box control) and a night-time phase for the repressor (under RRE control) resulted in reporter oscillation at phase around ZT8. Thus, they reproduced the endogenous gene regulation where a morning activator

DBP (controlled by E-box) targets D-boxes under competition by the night-time repressor E4BP4 (controlled by RRE), where the phase of D-box controlled expression was found to be around ZT8 in this artificial system. Moreover, Authors could also reproduce a similar output around ZT16 by expressing the repressor under E-box control in the morning (similar to REVERB promoter) and the activator in the evening under D-box control (as ROR $\alpha$ ). This experiment demonstrated how phase setting was the product of transcription factor binding at promoters of target genes. Furthermore, the group was able to synthesize promoters leading to unusual output phases by combining different phases for activators and repressors, concluding that combinations of transcription factors with the three main phases in the core-clock can generate a phase-specific transcriptional output at various times of the day.

In the second study, by investigating the *Cry1* locus, and assessing the significance of an intronic RRE for the phase control of expression of *Cry1* transcript, Udai-Tadenuma *et al.* [326], showed that phase-specific expression could also occur from combinations of several clock controlled elements. Adding an RRE to the luciferase reporter gene under control of D-boxes or E-boxes in its promoter resulted in an output phase similar to a prediction from a vector sum of the contributions of both regulatory elements. Such a vector additive model could explain how phase-specific circadian control of output genes results in the variety of phases observed in the mouse liver. This simple additive vector model is called the phase vector model.

In addition, Rey *et al.* studied oscillations of BMAL1 binding in mouse liver [273]. The authors applied a linear model approach to infer transcription factor activities for several factors which were shown to be targeted by BMAL1 binding. This method used the mRNA expression of BMAL1 target genes combined with transcription factor binding sites, determined by position weight matrix scans (PWMs from SwissRegulon) in promoter regions. The results of this analysis showed a circadian activity for DBP/HLF/TEF/E4BP4, REVERB/ROR, HIF1A, PPAR $\alpha$ , and BACH1 motifs.

A last interesting study from Westermarck and Herzel focused on the mechanism producing rhythms with a period of 12 hours [341]. They hypothesized that previously reported 12 hours rhythms in gene expression are the results of an interplay between components of the circadian clock. Therefore, they constructed a theoretical model involving pairs of circadian transcription factors. Using this model, they observed that the conditions required for these 12 hours oscillations include a certain circadian phase relationship between the two component of the clock as well as a non-competitive binding to the promoters of regulated genes. Finally, the authors demonstrated that binding sites for transcription factor pairs with phase relationships predicted by theory were overrepresented in promoters of genes displaying 12 hr rhythms using published mRNA microarray and position-weight matrices scan.

### 1.5 Interplay between circadian clock and metabolism

#### 1.5.1 Impaired metabolic phenotypes associated with mutations of core clock genes

Liver-specific deletion of *Bmal1* disrupts the rhythm of glucose metabolism (table 1.2) and causes a low fasting glucose level [185]. Shimba and colleagues reported that BMAL1 is an essential regulator of adipogenesis and lipid metabolism in matured adipocytes [300]. Embryonic fibroblast cells deficient in BMAL1 gene fail to differentiate into adipocytes except with the introduction of an exogenous BMAL1 copy. *Bmal1* knockout mice show increased respiratory quotient, reduced fat storage, increased circulating fatty acid, and increased ectopic fat formation in liver and muscles. Most of the genes implicated in adipocyte function (e.g., PPAR $\gamma$ , C/EBP $\alpha$ , SREBP1c, and lipin1) are expressed at a low level in BMAL1 deficient mice, whereas the preadipocyte marker PREF-1 is expressed at a high level. The impaired adipogenesis in BMAL1 knockout mice prevent the increase of the adipose tissue upon nutrient excess and thus lead to ectopic accumulation of fat in liver and muscles. Intriguingly, the differences in adipose tissue size appear in adults but not juveniles of BMAL1 knockout mice, linking a functional clock to aging.

In addition, hepatic overexpression of *Cry1* inhibits gluconeogenesis and reduces blood glucose levels [354], whereas a whole body double knockout of *Cry1* and *Cry2* leads to Glucose intolerance and constitutively high levels of circulating corticosterone [158]. Several other clock mutants depict perturbed metabolic phenotypes. More intriguingly the whole-body knockout of PPAR $\gamma$  coactivator-1 $\alpha$  (PGC-1 $\alpha$ ), a metabolic regulator, exhibits an abnormal circadian pattern of gene expression, suggesting a feedback mechanism of the metabolism to the clock through PGC-1 $\alpha$ .

In summary, genetic disruption of clock genes in mice perturbs metabolic functions of specific tissues at distinct phases of the diurnal cycle.

## Chapter 1. Introduction

Table 1.2 – Metabolic phenotypes associated with mutations of key component of the circadian clock in mice. Reproduce from [210]

<i>Protein</i>	<i>Mutation</i>	<i>Metabolic Phenotype</i>	<i>Reference</i>
CLOCK	whole-body loss-of-function	Attenuated feeding rhythm, obesity, hyperphagy, hyperlipidemia, hyperglycemia, hepatic steatosis, hypoinsulinemia	Turek <i>et al.</i> (2005), Marcheva <i>et al.</i> (2010) [324, 214]
BMAL1	whole-body knockout	Glucose intolerance, hypoinsulinemia, increased respiratory quotient, reduced fat storage, increased circulating fatty acid, increased ectopic fat formation in liver and muscles, hypoinsulinemia	Lamia <i>et al.</i> (2008), Shimba <i>et al.</i> (2011), Andrews <i>et al.</i> (2010), Marcheva <i>et al.</i> (2010) [185, 300, 12, 214]
BMAL1	Liver-specific knockout	Hypoglycemia in the rest phase	Lamia <i>et al.</i> (2008) [185]
CRY1 CRY2	whole-body double knockout	Glucose intolerance and constitutively high levels of circulating corticosterone	Lamia <i>et al.</i> (2011) [158]
REVERB $\alpha$ REVERB $\beta$	whole-body double knockout	Hepatic steatosis, hyperglycemia, hyperlipidemia	Bugge <i>et al.</i> (2012), Cho <i>et al.</i> (2012) [43, 60]
HDAC3	Liver-specific knockout	Hepatic steatosis	Feng <i>et al.</i> (2011) [98]
PGC-1 $\alpha$	whole-body knockout	Abnormal diurnal rhythms of activity, body temperature and metabolic rate	Liu <i>et al.</i> (2007) [200]
AMPK	whole-body knockout	body temperature and metabolic rate	He <i>et al.</i> (2003) [133]

### 1.5.2 Clock controlled metabolism in mouse liver

The liver is a central metabolic organ where various processes occur during a diurnal cycle, such as carbohydrate metabolism, protein metabolism, and lipid metabolism [47, 151]. Several other liver functions controlled by the internal clock implies blood detoxification [103] through drug metabolism, blood homeostasis, through the production of albumin, a major osmolar component of blood serum, immune function, through Kupffer cells (macrophages located in the liver), and blood pressure regulation, through production and secretion of angiotensinogen [136].

Opposing metabolic processes, such as glycolysis/gluconeogenesis, and lipogenesis/fatty acid oxidation, necessitates temporal separation; therefore circadian clock plays a significant role in the regulation of hepatic function. For example, the rate-limiting enzymes of glycolysis and gluconeogenesis oscillate and peak in the early morning and early evening, respectively. Metabolic phenotyping studies of clock mutant mice illustrate the importance of a functional clock on liver metabolism [286].

Studies on the "rate-limiting" gluconeogenic enzyme, phosphoenolpyruvate carboxylase (PEPCK), have revealed important links between metabolism and circadian cycle [286]. The PEPCK activity is diurnal in mouse liver, implying rhythmic glucose production. The oscillation of enzymatic activity is mainly due to a cyclic cellular accumulation of PEPCK since this enzyme has a short half-life and the major control mechanism is mediated by transcriptional regulation, through CREB and FOXO1 [116] (figure 1.15). Feeding/fasting cycles generate daily oscillation of metabolic hormones, such as glucagon, glucocorticoid hormone, thyroid hormone, and insulin. In addition, Glucagon activates CREB through the cAMP/PKA axis [8].

Glucocorticoid hormone activates the transcription of *Pepck*. Recently, Evans and colleagues have characterized the glucocorticoid-dependent interaction between CRYs and glucocorticoid receptor [158]. CRYs inhibit glucagon and glucocorticoid hormone signaling, which constitute a negative feedback loop to the transcriptional control of *Pepck*. The REVERB/ROR pair directly regulates *Pepck* expression. PPAR $\gamma$  coactivator-1 $\alpha$  (PGC1 $\alpha$ ) exhibits circadian gene expression and contributes to *Pepck* gene regulation by protein-protein interaction with FOXO1 and ROR [46]. PGC1 $\alpha$ , which is a transcriptional coactivator for the regulation of energy metabolism, is also involved in circadian regulation and promotes the expression of *Bmal1* and *RevErba* through the RRE site [200]. Thus, the core metabolic genes are tightly related to the clock systems, and their activities undergo circadian changes.



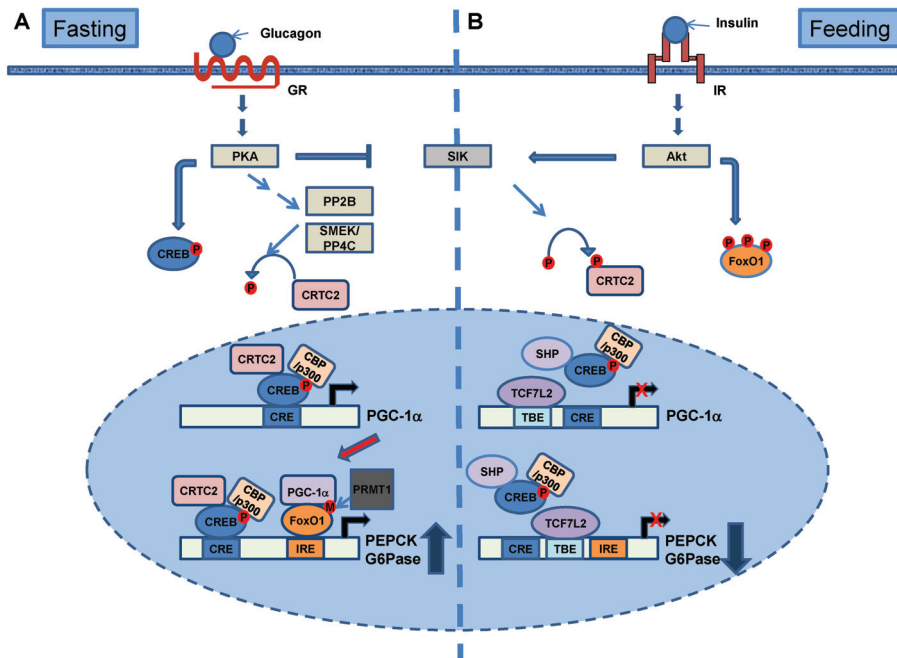


Figure 1.15 – **A)** Transcriptional control of hepatic gluconeogenesis under fasting conditions. Under fasting conditions, elevated secretion of pancreatic glucagon initiate activation of PKA, which phosphorylates CREB at the serine 133 residue, driving the increased association of this factor with co-activator CBP/p300 onto the chromatin. Moreover, PKA dephosphorylates CRTC2 at the serine 171 residue by inactivating SIK kinases and activating serine/threonine phosphatases SMEK/PP4C and PP2B, inducing the nuclear localization and increased association of CRTC2 with chromatin-bound CREB. These events induces the expression of gluconeogenic genes such as PEPCK and G6Pase, during the early phase of fasting. Simultaneously, CREB/CRTC2 can increase the expression of PGC-1 $\alpha$  and ERR $\gamma$ , which are controlling hepatic gluconeogenesis during the later phase of fasting. PRMP1 is implicated in the regulation of gluconeogenesis by modifying the arginine residues (Arg 248 and 250) of FoxO1 during this process. **B)** Transcriptional repression of hepatic gluconeogenesis under feeding conditions. On the other hand, feeding induces the reduced plasma concentration of glucagon and increases secretion of pancreatic insulin, which activates insulin signaling pathways in the liver. Akt activation control subsequent SIK kinases activation, consequently increasing phosphorylation of CRTC2. Simultaneously, Akt phosphorylates critical residues of FoxO1, switching off the transcription of hepatic gluconeogenesis. In addition, transcriptional repressors implicated in hepatic gluconeogenesis such as SHP, DAX-1, and TCF7L2 are expressed under this condition, which allow the inactivation gluconeogenic genes as well as PGC-1 $\alpha$ . Reproduced from [240]



## 1.5. Interplay between circadian clock and metabolism

Key metabolic factors like AMPK, SIRT1, PPAR $\alpha$ , and PGC1 $\alpha$  act as important regulators for core circadian mechanisms (figure 1.16). AMPK, which is a nutrient sensor, play a role in the destabilization of the CRY protein in the core of the circadian system [159]. *Sirt1* is related to anti-aging and is regulated by NAD<sup>+</sup>. Furthermore, *Sirt1* regulates HAT activity of the CLOCK protein [346] and promotes deacetylation and degradation of PER2 [15]. PPAR $\alpha$ , which is a nuclear receptor for lipid metabolism in the liver, binds with PER2 [294] and increases *Bmal1* expression through a peroxisome proliferator-activated receptor (PPAR) response element in the promoter of *Bmal1* [149].

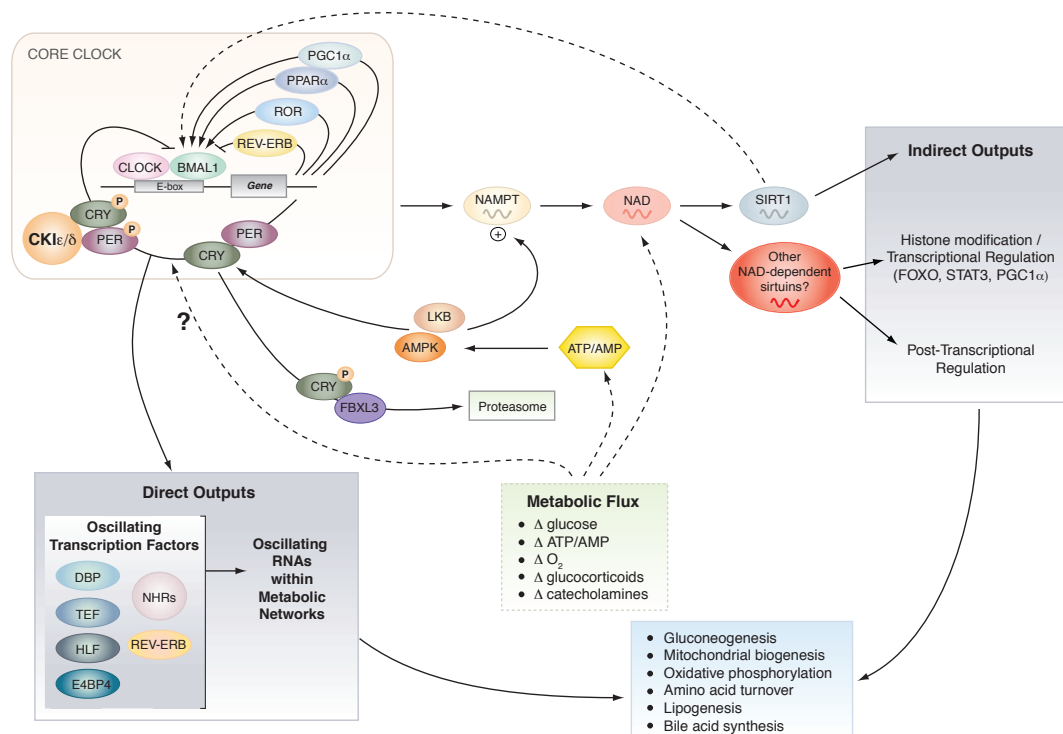


Figure 1.16 – Direct and indirect outputs of the core clock mechanism. The core clock consists of a series of transcription/translation feedback loops that synchronize diverse metabolic processes through both direct and indirect outputs, including gluconeogenesis and oxidative metabolism. The clock also receives reciprocal input from nutrient signaling pathways (including SIRT1 and AMPK), which function as rheostats to couple circadian cycles to metabolic flux, especially in peripheral tissues. Reproduced from [28]

Perturbation of lipid homeostasis in the liver is involved in the development of obesity-related metabolic diseases, such as hyperlipidemia, insulin resistance, coronary artery disease, non-alcoholic fatty liver disease and diabetes. In recent years, molecular mechanisms of how circadian clocks regulate lipid metabolism have been studied extensively. It has been known for years that regulators of lipid metabolism oscillate in the liver [210]. Circadian clocks directly or indirectly regulate these regulators. For example, Schibler and colleagues have studied the modulation of SREBP in *Reverba* KO mice. SREBP targets such as *Hmgcr*, *Acas*,

## Chapter 1. Introduction

---

*Fas* and *Elo6* are involved in cholesterol and lipid metabolism. REVERB $\alpha$  repress *Insig2*, which sequester SREBP to the endoplasmic reticulum membrane and thereby interfere with the proteolytic activation of SREBP in the Golgi membrane. A rate-limiting enzyme of the conversion from cholesterol to bile acid, known as Cyp7a1, is modulated by REVERB $\alpha$ . This control is mediated by the stimulation of LXR by rhythmic production of oxysterols [189]. More recently SREBP binding sites have been identified genome-wide. Targets of SREBP were divided into clusters based on their expression pattern. One of the clusters showed a high enrichment of HNF4 binding sites and a phase shift of 8h compared to SREBP binding, which strongly suggest cross talk between these factors. Interestingly the expression of targets co-regulated by HNF4 and SREBP1 is strongly impaired in the Bmal1 KO mice [111].

Taken together, these researches done in the past decades revealed several key factors implicated in the cross talk between the nutrient cycle and the circadian rhythm.

## **1.6 Objectives**

At the molecular level, the relationship between the circadian cycle and nutrient response cycle (or metabolism) is poorly understood. Therefore, the CycliX consortium aimed at characterizing the changes in genomic states and transcriptional programs that are implicated in both cycles. The CycliX consortium studied the transcriptional and epigenetic basis of diurnal rhythms in mouse liver genome-wide, using temporal DNA occupancy profiles of RNA polymerase II (Pol II) as well as profiles of histone modifications [188]. Thus, they showed that the epigenetic landscape is highly dynamic and globally remodeled during the 24-hour cycle.

Currently, the circadian transcriptional regulation is mostly explained by three groups of transcription factors, bHLH proteins, such as BMAL1 or CLOCK, PAR b-Zip proteins, such as DBP, TEF or HLF, and nuclear receptors, such as ROR or REVERB, respectively acting at ZT8, ZT12, and ZT22. In addition, the contribution of the distal regulatory elements to circadian regulation as well as the tissue-specific regulation of the circadian cycle remains unclear.

Therefore we aimed at:

- Investigating the chromatin accessibility, as well as RNA polymerase II and histone 3 lysine 27 acetylation genome-wide profiles over 24h in the circadian context.
- Identifying the cis-regulatory modules and transcription regulators acting at the mapped cis-regulatory sites.
- Investigating genome-wide differences between *Bmal1*<sup>-/-</sup> and WT mice to understand the contribution of the circadian cycle and the nutrient response cycle regarding diurnal oscillations in mouse liver.
- Identifying of core clock and metabolism transcription factors as well as downstream effectors.
- Explaining phase-specific regulation by transcription factors.
- Identifying putative tissue-specific transcription factors by comparing NIH3T3 cells and the liver.
- Identifying putative co-regulatory elements involved in differential regulation of BMAL1-CLOCK binding in different contexts, namely NIH3T3 cells and mouse liver.

## **1.7 Achievements**

- Investigation of genome-wide DNase I hypersensitive sites (DHSs) in mouse liver. About 62000 DHSs were detected in mouse liver. In addition, 98000 Footprints were detected in about 3/5 of DHSs. We observed that DHS around TSS tend to contain more footprints.

## Chapter 1. Introduction

---

- Study of the temporal dynamic of Pol II loadings, H3K27ac and DNase I signal at DHSs in WT and *Bmal1*<sup>-/-</sup>. About 10 % of the DHSs depicted a circadian behavior.
- Analysis (and optimization) of transcription factor binding site using digital genomic footprinting methods in WT and at ZT6 in *Bmal1*<sup>-/-</sup>. Chromatin accessibility was dynamically influenced by the binding of transcription factors and DNase I cleavage pattern reflected DNA binding of protein complexes such as a BMAL1:CLOCK hetero-tetramer on double E-boxes.
- Study of putative BMAL1 cooperating transcription factor and tissue-specific factors in liver and fibroblasts. Our analysis revealed several putative BMAL1 co-regulators such as NRF1 and ZEB1, and tissue-specific factors such as HNF4, CEBP, FOXA and GATA1.
- Analysis (and optimization) of putative circadian regulators and regulators implicated in nutrient response cycle using a linear model approach in WT and *Bmal1*<sup>-/-</sup>. Our genome-wide description of WT compared to *Bmal1* KO mice revealed the importance of GR, FOX, and CREB in clock impaired mice under night restricted feeding. Thus, we observed a partial disorganization of transcriptional control in the absence of a peripheral clock, which underlined the strong effect of nutritional synchronization. Global *Bmal1* knockout in mice lead to important effects on metabolism and behavior that could be traced to dysregulation of BMAL1 target genes.
- Analysis of the contribution of distal regulatory element involved in the circadian transcription regulation. We observed that about 47% of DHS are located at more than 10 Kb from the closest active TSS. In addition distal DHSs up to 50 Kb improved the variance explained by our penalized linear model in WT and *Bmal1*<sup>-/-</sup>.
- Inspection of diurnal biological processes such as lipid and sugar metabolism and their regulation in WT and *Bmal1*<sup>-/-</sup>. Our analysis confirmed the implication of SREBP, ChREB, CREB and FOX in these processes.

## 2 Results

### Context

In the past decade, the emergence of various next-generation sequencing technologies and experimental procedures allowed querying our epi-genome and our gene expression with an unprecedented precision at a genome-wide or transcriptome-wide scale. These experiments includes chromatin immuno-precipitation followed by sequencing (ChIP-seq) [27] and DNase I-sequencing (DNase I-seq) [64] to study respectively DNA-binding proteins and chromatin accessibility. These techniques have gained a substantial interest thanks to consortiums like ENCODE, which have produced a tremendous amount of data in a wide variety of tissue and for a lot of different DNA-binding proteins [110, 237, 292, 311]. In the circadian field several groups have produced time series of liver ChIP-seq of histone marks as H3K27ac or H3K4me3, as well as core clock transcription factor (TF) as BMAL1, CLOCK, PER1,2, CRY1,2 ROR $\alpha$ ,  $\beta$ ,  $\gamma$ , REVERB $\alpha$ ,  $\beta$ , or metabolism related TF like CREB, SREBP, FOX, and more [171, 111, 188, 94, 95, 273, 334]. These studies have unveiled the dynamic nature of the chromatin landscape, upon time or stimulation or stress. Our goal in this project was to decipher the transcriptional regulatory logic of the diurnal cycle using a high-quality dataset produced by the CycliX consortium ,in WT and *Bmal1*<sup>-/-</sup> in mouse liver.

*Most of the results presented in the following sections were produced for the paper in preparation entitled "Transcriptional regulatory logic of the diurnal cycle in the mouse liver".*

### Abstract

Most forms of life exhibit temporal rhythms in gene expression that propel diurnal cycles in physiology. These rhythms are controlled by transcription-translation feedback loops of the core circadian clock and modulated by feeding-fasting rhythms. To better understand the regulatory interplay between the circadian clock and feeding-rhythmic metabolism, we examined DNase I hypersensitive sites (DHSs) in mouse liver during a diurnal cycle. DNase I signals cycled at a substantial fraction of all DHSs, suggesting that DHSs harbor regulatory elements controlling rhythmic transcription. Using ChIP-seq, we found that hypersensitivity

## Chapter 2. Results

---

cycled in phase with RNA polymerase II (Pol II) loading and H3K27ac. We exploited the DHSs to design a penalized linear regression model to infer the activity of transcription regulators using Pol II loading in WT and in *Bmal1*<sup>-/-</sup> mice. While our model identified most known circadian regulators, we also found motifs for the transcription factors (TFs) CREB, SREBP, FOX and GR, that exhibited diurnal activity both in WT and *Bmal1*<sup>-/-</sup> livers. Since these TFs regulate genes that display circadian oscillations due to food entrainment under night restricted feeding, our results suggest that these regulators are impacted by systemic cues or food driven in *Bmal1*<sup>-/-</sup> mice. In addition, we observed that hypersensitivity was only mildly affected genome-wide in arrhythmic *Bmal1*<sup>-/-</sup> mice, in contrast to BMAL1 binding sites that exhibited a strong reduction of hypersensitivity. Interestingly, though, nucleotide resolution DNase I footprints at locations harboring BMAL1 bound tandem E-box motifs changed in shape over the diurnal cycle, suggesting a transient hetero-tetramer binding configuration at those loci between ZT6 and ZT10. Overall, DNase I mapping provided significant additional insights into the mechanisms of diurnal transcription regulation in mouse liver.

## **2.1 Chromatin landscape in mouse liver**

### **2.1.1 Genome-scale mapping of DNase I hypersensitivity in mouse liver**

To comprehensively map putative regulatory elements genome-wide, we merged the seven DNase I hypersensitivity time points and performed peak finding (Methods section 4.2). This revealed 62'418 DNase I hyper sensitive sites (DHS), covering around 2% of the mappable genome (taking a width of 600bp for each DHS), which is comparable to previous studies across mouse tissues [352]. Because we aimed at associating DHSs with nearby genes to infer regulatory relationships, we first decided to remove from ENSEMBL annotations transcripts that were not expressed in our samples. Similar to our previous approach [188], we used histone modifications, Pol II profiles, and now also DNase I signals at transcription start and end sites of annotated transcripts to train a supervised learning method (support vector machine) that distinguishes expressed (active) from non-expressed genes (Methods section 4.8). To infer putative regulatory relationships, we then annotated each DHS to the nearest active TSS. Distances between DHSs and TSSs followed a bimodal distribution, with a first mode around 100bp from the TSSs and a second 10kb from the TSS (Figure 2.1 A).

Consistent with previous reports [320, 330], one third of our DHSs were found within 1kb of TSSs, while almost half were located more than 10kb from a TSS (Figure 2.1 B), suggesting that the identified DHSs contained both promoter-proximal and distal control regions. For the promoter proximal DHSs, the genomic distributions of DNase I cuts, Pol II, and H3K27ac signals (centered on TSSs) were consistent with the accessibility of DNA being determined by nucleosome displacement and Pol II complex assembly (Figure 2.1 C) [343]. At distal DHSs, profiles of H3K27ac showed a dip in the peak center, consistent with occupation by TFs and nucleosome displacement (Figure 2.1 D), while the weaker Pol II signals could reflect distal assembly of the transcriptional complex [170], or interactions between enhancer regions and the TSS through DNA looping [252, 263].

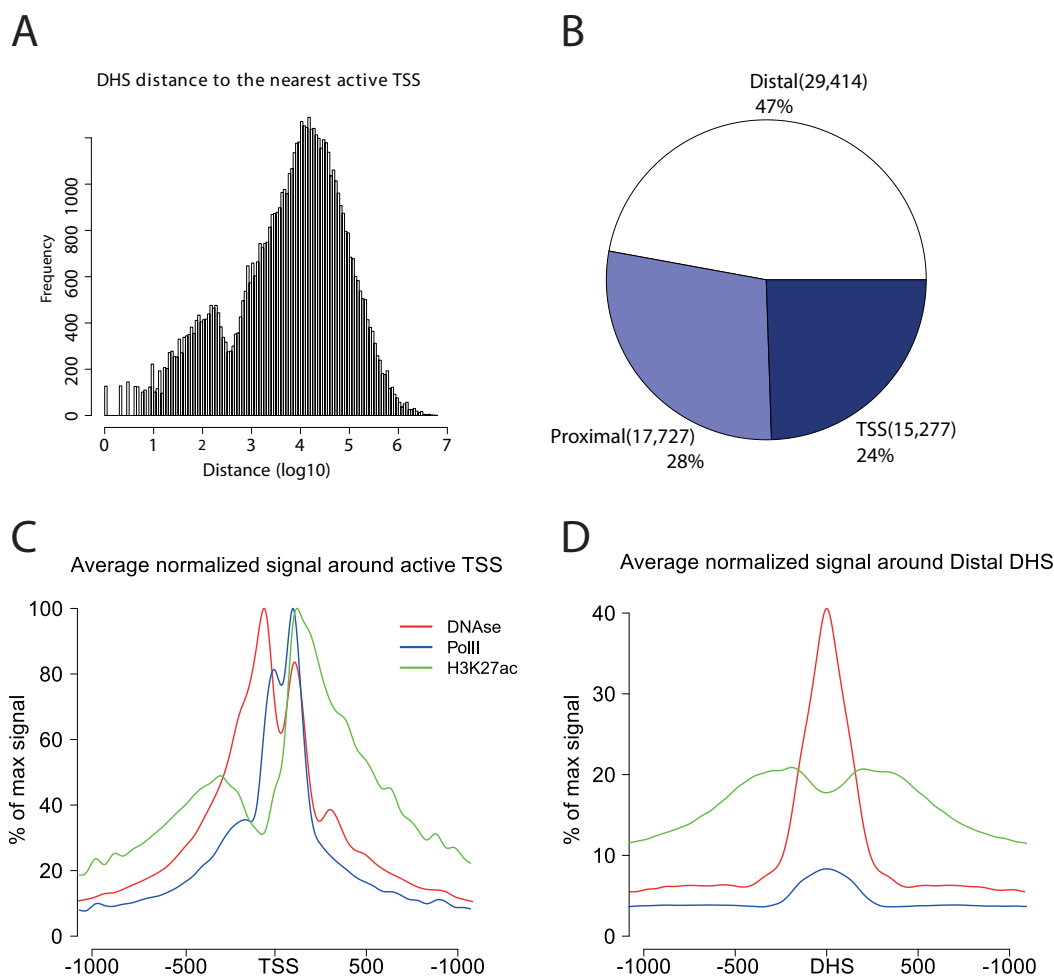


Figure 2.1 – Genome-wide localization of DNase I accessibility. **A**. Histogram of the distance between each DHS and the nearest active TSS. We observe a bimodal distribution, with a first mode corresponding to DHS in the promoter region (centered on 100bp from the TSS) and a second mode centered on 10kb from the TSS. **B**. Repartition of DHS within three classes depending on their distance from the nearest TSS: 47% are more than 10kb from a TSS and are classified as distal, 28% are between 1kb and 10kb away and are classified as proximal, and finally TSS DHS 1kb or less from a TSS represent 24% of all sites. **C**. Pol II, DHS and H3K27ac average signal around TSSs and distal DHSs. Profiles were adjusted so that the maximum around the TSS represents 100%. At the TSS, patterns of accessibility, Pol II loading and acetylated histone occupation correspond to previously observed characteristics [320, 58, 35]. **D**. At distal DHSs, DNase I signal is 60% lower on average. H3K27ac profiles display even further decrease. However they are also suggestive of nucleosome displacement at the center of the region. Corresponding centered signal of Pol II is also seen, albeit on average less than one-tenth the intensity of TSS Pol II.



### 2.1.2 DHS dynamics and correlation with Pol II and H3K27ac signals

To validate our assays, we examined the known circadian output gene *Dbp*, maximally transcribed at ZT8, to determine whether cutting frequency at DHSs exhibited diurnal variation. We detected several DHSs in the vicinity of *Dbp*, with high intensity and narrow signals surrounded by low noise levels. We observed that DNase I hypersensitivity was oscillating around *Dbp* locus (figure 2.2 A)

Ripperger *et al.* (2000) studied the *Dbp* locus using classical DNase I footprinting experiments on gel (figure 2.2 B). DBP Intronic DNase I hypersensitive regions embody E-box motifs that bind BMAL1/CLOCK. Because intragenic enhancer sequences appear to be essential for robust circadian *Dbp* expression, the hypersensitive regions located downstream of the cap site were investigated by [278] in greater detail. Ripperger's study uncovered seven putative *Dbp* regulatory sequences, of which two are located upstream (6 and 7) and four downstream (1–4) of the transcription initiation site. The sensitivity towards DNase I digestion of four of these regions (2, 4, 6, and 7) and that of the promoter region including the transcription initiation site oscillate with the same phase as *Dbp* transcription and high amplitudes are represented in green, whereas sites 1,3 and 5 with low amplitudes are represented in blue. Three of the putative regulatory sequences (2, 4, and 7) contain E-boxes that are binding sites for CLOCK and BMAL1. These sites were successfully detected in the current DNase-seq study with similar properties and we confirmed some of their hypothesis regarding BMAL1 binding at sites 6,4,and 2 using BMAL1 ChIP-seq from G. Rey study.

As exemplified by a DHS nearby the transcription start site (TSS) of *Dbp*, we observed that DHSs were located in regions with lower H3K27ac signals in between H3K27ac-enriched islands, suggestive of TF-induced nucleosome displacement [153, 131, 222]. Moreover, the DNase I hypersensitivity changed diurnally, notably at the TSS where the oscillations in DNase I hypersensitivity, Pol II, and H3K27ac peaked in sync at ZT10. However, all DHSs within 15 kb of the *Dbp* TSS displayed oscillations with the same phase as the TSS, suggesting regulatory relationships between these regions and gene transcription (see figure 2.3).

We next analyzed the *Npas2* gene (see figure 2.4), another known clock target [32]. *Npas2* is a target of RORs and peaks in the late night-time around ZT22 [66]. We detected several DHSs along the transcribed region of *Npas2* (Figure 2.4 A), including proximal (defined as 1-10kb from a TSS) and distal (defined as >10kb from a TSS) sites. The distal sites displayed high amplitude oscillations of DNase I cuts and H3K27ac (Figure 2.4 B). Normalized signals at the *Npas2* TSS also peaked at the expected phase with maximal signal at ZT22 for all three marks studied (Figure 2.4 c). Finally, all DHSs associated with *Npas2* (those having *Npas2* as their closest TSS), including numerous distal regions, likewise cycled with phases around ZT22 (Figure 2.4 D). Together, observations at the *Dbp* and *Npas2* loci show that DHSs were detected genome-wide with high resolution, and that the temporal patterns of DNase I cuts reflected diurnal activity of these elements at different times of the day.

## Chapter 2. Results

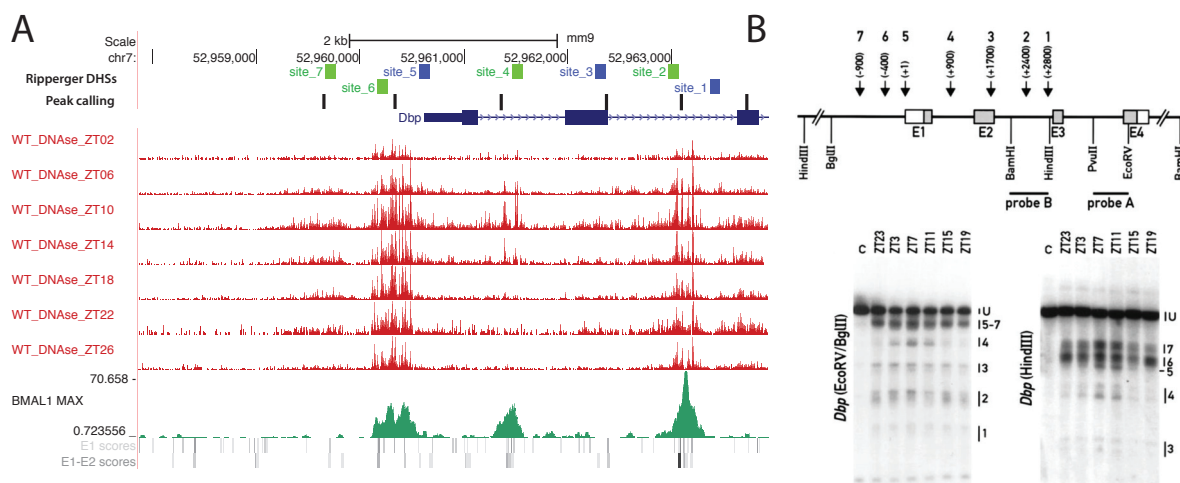


Figure 2.2 – DNase I accessibility at *Dbp* locus. DNase I hypersensitive sites within the *Dbp* locus. **A)** Detected DHS using DNase I-seq at ZT02-ZT26, overlapped with BMAL1 ChIP-seq, E-box prediction scores and previously detected hypersensitive sites by Ripperger *et al.* **B)** Original Ripperger analysis of *Dbp* locus: Schematic representation of the *Dbp* gene with its four exons (E1–E4) and three introns. The positions of the DNA hybridization probes and the restriction fragments used in the indirect end-labeling experiments are indicated. The approximate positions of the seven DNase I hypersensitive regions detected in both gels are depicted on top of the cartoon. Left gel: Mapping of DNase I hypersensitive sites starting from exon 4. Equal aliquots of liver nuclei harvested at the indicated times were treated with DNase I (ZT23 to ZT19) or without DNase I (lane C, derived from ZT7). After exhaustive digestion with EcoRV and BglII, the fragments were visualized using probe A. Right gel: Fine mapping of DNase I hypersensitive sites in the 5' moiety of the *Dbp* locus. The DNA was digested with HindIII and probed with probe B [278].

## 2.1. Chromatin landscape in mouse liver

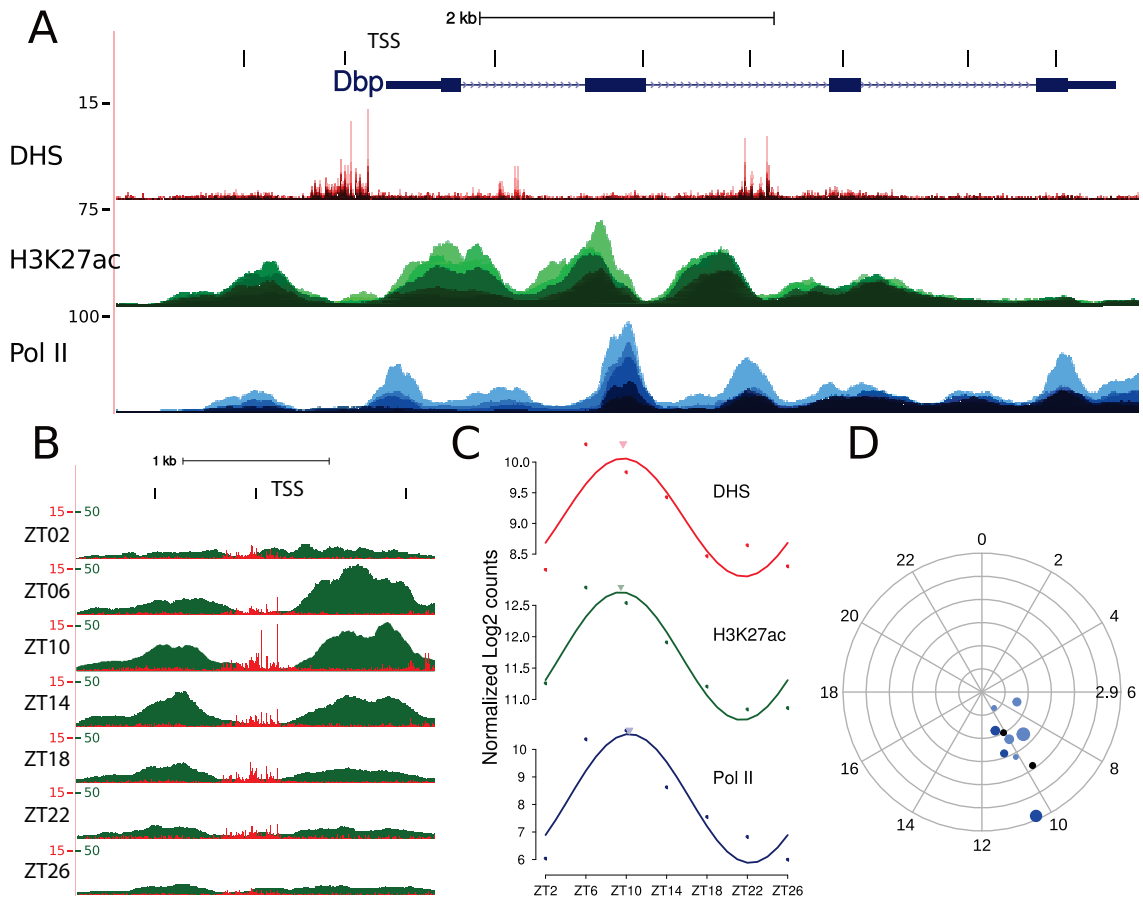


Figure 2.3 – DNase I accessibility, Pol II and H3K27ac dynamics at *Dbp* locus. **A**) DNase I hypersensitivity, Pol II, and H3K27ac enrichment at the *Dbp* locus. The DHS track shows a nucleotide resolved cutting frequency while H3K27ac and Pol II ChIP-seq signals are smoothed over 100bp. The seven time points are overlaid. **B**) A zoom-in around the DHS at the TSS of *Dbp* reveals DHS enrichment dynamics around the clock. Both DHS and H3K27ac are maximal at ZT10 and minimal at ZT22, consistent with BMAL1-mediated activation of *Dbp* transcription. Note that DHS signal is strongest in between adjacent histones, probably reflecting nucleosome displacement around transcription factor binding sites at the center of highly accessible regions. **C**) Read counts (in log<sub>2</sub> units) for DHS signal (in windows of +/- 300bp) centered on the *Dbp* TSS. Idem for Pol II and H3K27Ac ChIP-seq (in windows of +/- 1000bp) centered on the same DHS and cosine fits show a common peak phase around ZT10. Amplitudes are about 16 fold for Pol II, and approximately 4 fold for both DHS and H3K27ac. **D**) Phases and amplitudes of all DHSs that are located in the neighborhood of the *Dbp* gene (nearest-TSS association according to annotation). Each dot represents a region, with black for a TSS (x alternative TSS), dark blue for a near-TSS DHS, light blue for a proximal DHS and grey for a distal DHS. Log<sub>2</sub>-amplitudes are shown as distances from the center of the plot, phases as radial position clockwise from ZT0 at the top, and -log<sub>10</sub>(p-value) determine the size of the dots. We observed that all regions oscillate around a common phase at ZT10.

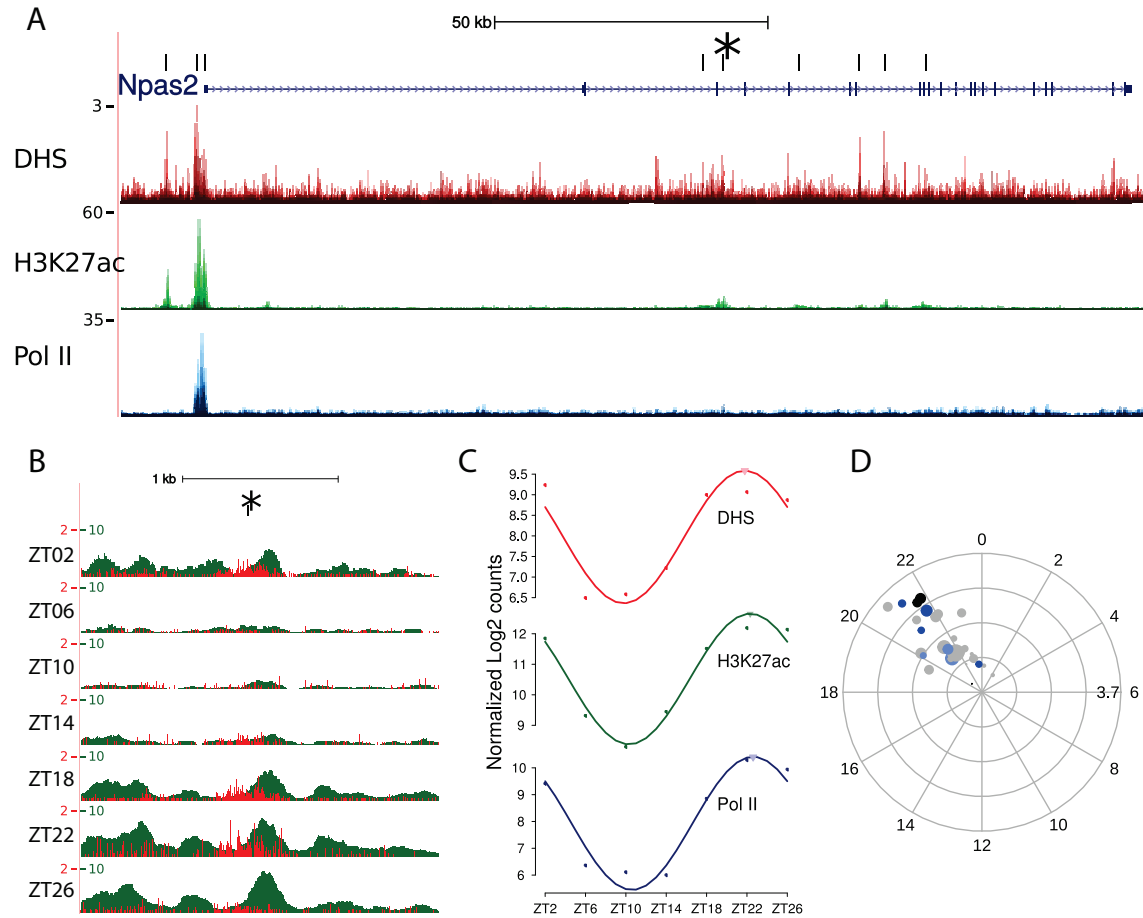


Figure 2.4 – DNase I accessibility, Pol II and H3K27ac dynamics at *Npas2* locus. **A**) DNase I hypersensitivity, Pol II, and H3K27ac enrichment at the *Npas2* locus. **B**) A zoom-in around the DHS at the enhancer of *Npas2* (starred) reveals DHS enrichment dynamics around the clock. Both DHS and H3K27ac are maximal at ZT22. **C**) Read counts (in log<sub>2</sub> units) for DHS signal (in windows of +/- 300bp) centered on the *Npas2* enhancer. Idem for Pol II and H3K27Ac ChIP-seq (in windows of +/- 1000bp) centered on the same DHS and cosine fits show a common peak phase around ZT22. **D**) Phases and amplitudes of all DHSs that are located in the neighborhood of the *Npas2* gene (nearest-TSS association according to annotation). Each dot represents a region, with black for a TSS (x alternative TSS), dark blue for a near-TSS DHS, light blue for a proximal DHS and grey for a distal DHS. Log<sub>2</sub>-amplitudes are shown as distances from the center of the plot, phases as radial position clockwise from ZT0 at the top, and  $-\log_{10}(\text{p-value})$  determine the size of the dots. We observed that all regions oscillate around a common phase at ZT22.

## 2.1. Chromatin landscape in mouse liver

---

We next studied whether DNase I hypersensitivity, accumulation of Pol II, and H3K27ac quantified at the identified DHSs displayed diurnal rhythms using harmonic regression (Methods section 4.4). The number of cyclic regions identified at different significance thresholds ( $p=0.01$ ,  $0.05$  and  $0.1$ , harmonic regression) clearly indicated that Pol II and H3K27ac oscillated at a larger number of DHSs compared to the DNase I signal itself, both for proximal (defined as 1-10kb) and more distal sites ( $>10$ kb) (Figure 2.5 A). To select high confidence rhythmically active regions, we assessed the combined rhythms of the three marks at each DHS using Fisher's method [188, 231], which yielded 4606 (7.3%,  $FDR < 0.05$ ) regions with diurnal patterns of activity. For all three signals, the amplitude of the oscillations was larger at distal sites (the median amplitude was twofold for DNase I and H3K27ac, and higher for Pol II) compared to TSSs, and Pol II had larger amplitudes than either DNase I or H3K27ac (Figure 2.5 B). Moreover, the phases (peak times) of the oscillations in DNase I signals were, except for some small deviations, similarly distributed as gene transcription and H3K27ac with a weak evening peak around ZT10 and a marked late night peak around ZT22 (Figure 2.5 C). We next considered the phase relationships of the DNase I, Pol II and H3K27ac rhythms.

## Chapter 2. Results

As previously demonstrated, chromatin marks exhibit diurnal rhythms that are tied to transcription [334, 171, 188, 277]. Similarly, enhancer RNAs (eRNAs) are transcribed in sync with their cognate transcripts [95].

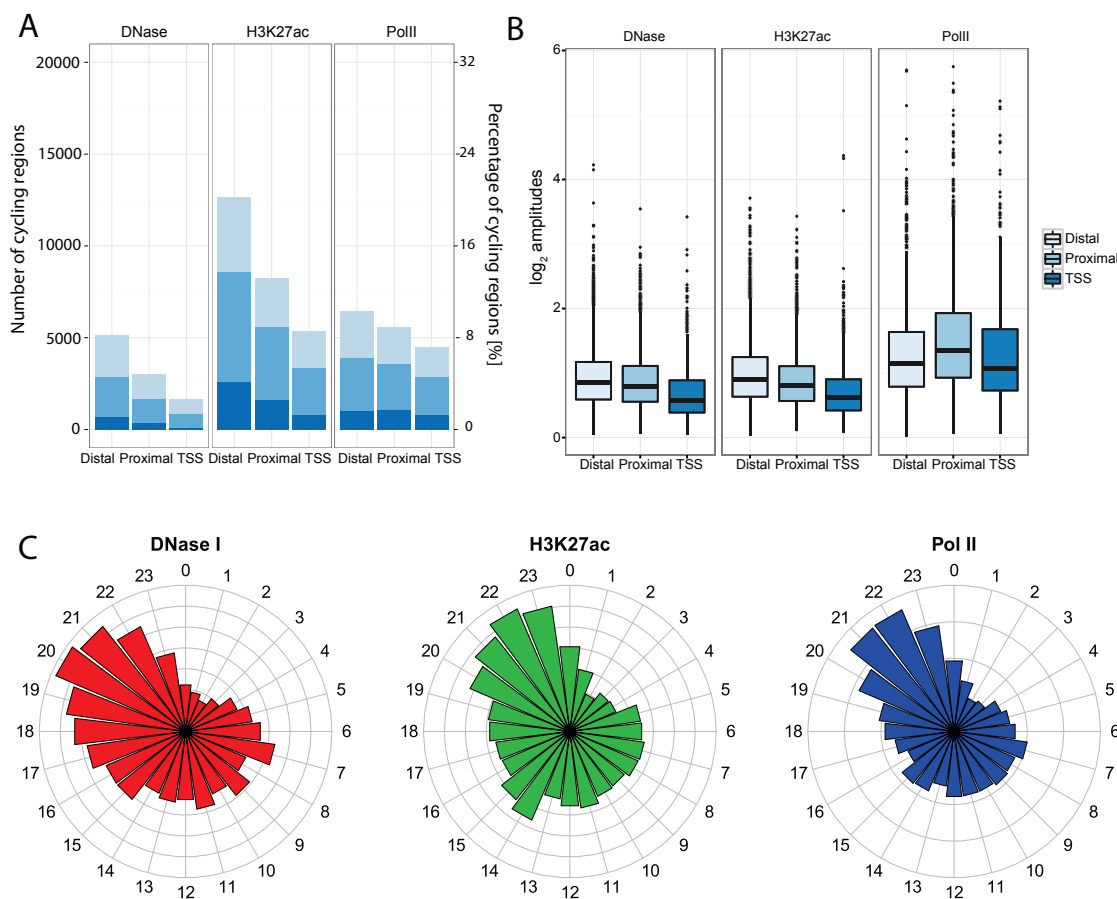


Figure 2.5 – Rhythmicity assessment, phases and amplitudes of DHS in DNase I, Pol II and H3k27ac signal. **A**) Number of DHSs with statistically significant cycling DNase I signal (left), H3K27ac-signal (middle) or Pol II signal (right) at three different thresholds ( $p < 0.1$ ,  $p < 0.05$  and  $p < 0.01$ ) partitioned according to their genomic location: TSS (1 kb), proximal (1-10 kb from TSS), or distal (>10 kb from TSS). **B**) Comparison of relative amplitudes for DHSs in each class (TSS, Proximal and Distal) and in each signal (Pol II, H3K27ac and DNase I). P-values were combined using Fisher's method and sites were selected with an FDR-corrected p-value threshold of 0.05. Higher amplitudes were observed in distal and proximal regions compared to TSS (t-test  $p < 2.2e-16$ ). In addition, Pol II patterns had higher peak-to-trough ratios than other signals. **C**) Circular histograms representing the distribution of phases for each mark at DHS selected as in **B**. All signals display the characteristic evening peak between ZT20 and ZT22, with DNase I peaking slightly in advance of other marks. The evening peak is also visible for DNase I data but appear weakly enriched in Pol II and H3K27ac.

We observed that DNA accessibility (DNase I cuts), Pol II, and H3K27ac displayed synchronous oscillations at DHSs (Figure 2.6 A). Such relationships were maintained after removing DHSs situated in the transcribed regions of active genes (data not shown), indicating that this is not a mere consequence of transcriptional elongation [275]. To test whether distal and TSS DHSs showed temporally correlated signals for each of the three signals, we examined pairs of oscillating DHSs (FDR adjusted  $p$ -value < 0.1, Fisher's combined test) of which one was located near a TSS (< 1kb) and one in an intergenic region required to be positioned at least 2kb and at most 20kb from any TSS. While no pair reached statistical significance for DNase I signals, probably reflecting that DNase I signals are noisier than the two other marks, we found 1611 pairs oscillating for H3K27ac and 630 for Pol II. The two phases were highly correlated with phase differences lower than one hour (Figure 2.6 B), suggestive of enhancer-TSS interactions [320].

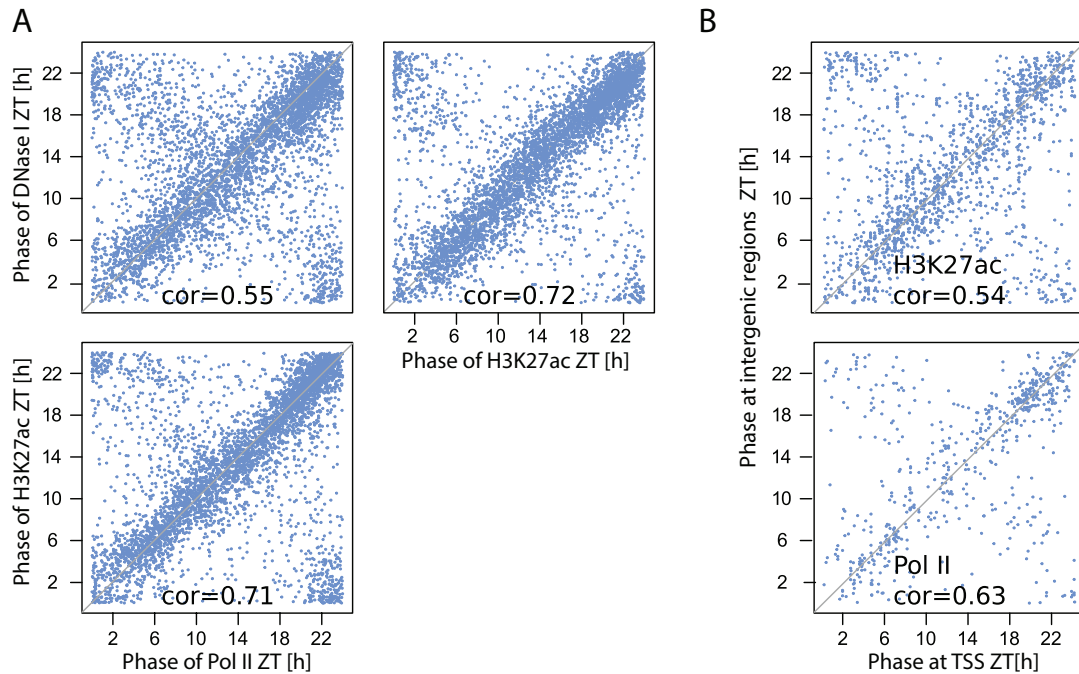


Figure 2.6 – Phase correlations of DHS in DNase I, Pol II and H3k27ac signal. **A)** Phase comparisons between DNase I, Pol II and H3K27ac at DHSs with a combined  $p$ -value threshold of 0.05. Scatter-plots display a concentration along the diagonal (shown in grey). Circular correlation values are indicated. Delays appear shorter than half an hour, which is within experimental variations, suggesting that the three marks peak synchronously. **B)** Phase relationship between intergenic accessible regions and their nearest TSS. Selection was done by combining  $p$ -values for cosine fits for each accessible regions and its nearest TSS for each signal and applying FDR correction, with a threshold of 0.1. Respectively 1611 and 630 significant pairs (intergenic accessible regions - TSS) were observed for H3K27ac and Pol II signals. The scatter-plots display a distribution along the diagonal. Circular correlation values are indicated and are high. This suggests that regulatory sites in the vicinity of active TSSs show coordinated temporal changes.



### 2.1.3 DNase I Footprint detection, localization and characteristics

To determine whether DHSs reflected DNA-bound transcription regulators, we searched for short windows protected from cleavage, also called footprints, using a published method [256] within a +/- 300 bp window around the center of each DHS. This identified previously reported footprints as illustrated for the well-characterized promoter of the Albumin (Alb) gene [196] (Figure 2.7).

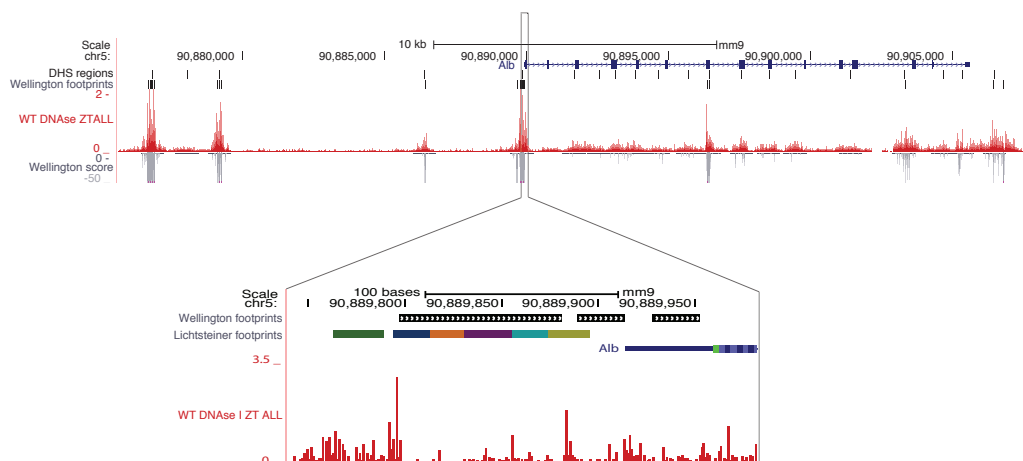


Figure 2.7 – DNase I accessibility and Footprint at *Alb* promoter. Albumin locus with a focus on promoter region displaying the DHS signal (bottom, red), the Wellington detected footprint regions (top, black) and the experimentally verified protein binding sites in the region (multicolor track, center). Many accessible intragenic regions in this locus does not display any footprint, but show a high DNase I signal, probably due to intense transcription of this gene in liver. The large protected region at the *Alb* TSS we detect in our data corresponds to previously established transcription factor binding sites [196].

In the promoter region of *Reverba* (*Nr1d1*), the detected footprints coincided with E-boxes and high BMAL1 ChIP-seq signals (Figure 2.8 A). Overall, the vast majority (70%) of DHSs near a TSS contained at least one footprint, while this proportion dropped to one half for the distal and proximal DHSs (Figure 2.8 B). Since transcribed DNA is known to be DNase I sensitive [310], the DHSs without footprints might reflect the process of transcription. To test this, we analyzed the number of footprints in DHSs outside of promoter regions and further marked with H3K36me3, a mark coinciding with transcribed gene bodies [119, 188]. Indeed, DHSs without footprints were frequently (90%) linked with highly transcribed regions (Figure 2.8 C). Thus, DHSs at TSS seem to contain more footprints than distal DHSs and highly transcribed regions explain why some of our DHS do not exhibit a footprint.



The region 8 Kbp upstream of *Reverba* (Figure 2.9) revealed a distal DHS with several footprints detected by Wellington algorithm [256]. This putative enhancer contains multiple predicted TFBS using FIMO [114] with the PWMs from Wang study [338]. We found several E-box motifs (with a similar consensus CACGTG sequence), such as USE, BHLHE40 or MAX, and other motifs such as HNF4 or CEBPB. Using various ChIP-seq data-sets from the literature, such as BMAL1 [273], HNF4A, CBP, CEBPA, p300, FOXA1 (also called HNF3 $\alpha$ ), FOXA2 and GABPA [96], we observed that this distal DHS is bound (directly or indirectly) by several transcription factors. This example illustrates the potential cooperative/competitive nature of the regulation of TFs at accessible regions. TFs rarely act alone at enhancers, their functions should be considered in a more integrated, combinatorial manner [308]. When the associated TFs are expressed in overlapping spatial domains, this combinatorial binding can result in discrete and precise patterns of transcriptional activity. For example, the recruitment of phase-specific activators and repressors gives rise to more refined expression patterns, such as a combination of day-time elements (D box) within the *Cry1*-proximal promoter and night-time elements (RREs) within its intronic enhancer gives rise to evening-time expression [326].

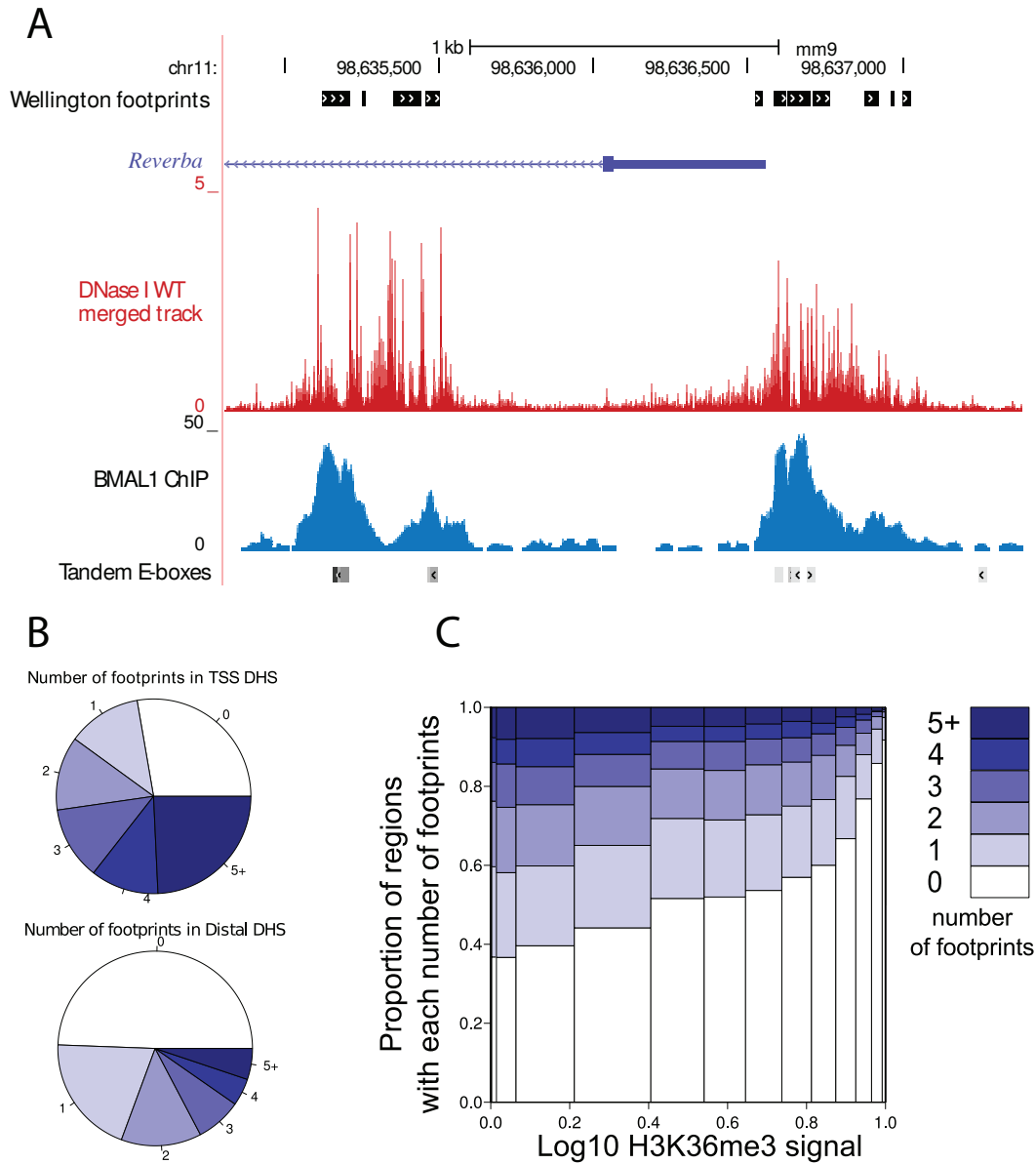


Figure 2.8 – Location-dependent footprint characteristics of DHSs. **A)** Visualization of DHS signal (red) around the *Reverba* promoter with the footprints detected by Wellington analysis annotated in black, on top. This region contains BMAL1 binding sites (blue) with one overlapping an E-box, annotated on the bottom line, that is marked by a characteristic footprint in which DNase I signal is virtually absent in the center, reflecting protection of the DNA from digestion, whereas high signal is observed on the edges of the binding site. **B)** Distribution of number of footprints within DHSs (+/- 300 bp around the peak center). TSS regions contain more footprints on average. More than half the distal regions contain a footprint, strongly supporting their identity as distal regulatory elements. **C)** Number of footprints detected in a given DHS as a function of H3K36me3 enrichment.

## 2.1. Chromatin landscape in mouse liver

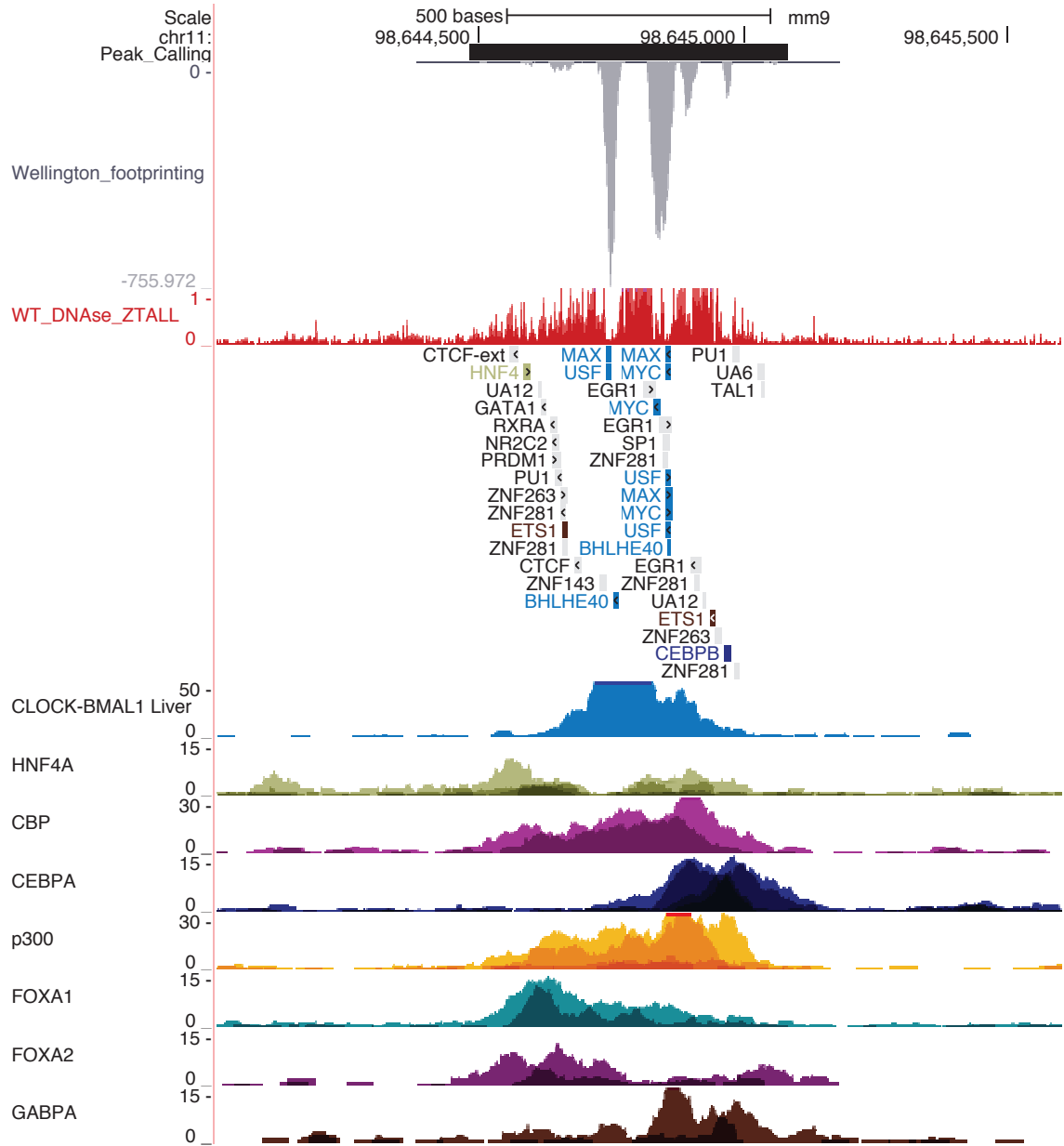


Figure 2.9 – *Reverba* enhancer footprint characteristics. In a region 8Kbp upstream of *Reverba* a DHS with multiple footprints was detected. These footprints contain several E-box related motifs, such as MAX, MYC, BHLH40 or USF (in blue). In addition, several predicted binding sites for HNF4, ETS1, CEBPB, and GATA1 were detected. The overlap with various ChIP-seq signal of BMAL1, HNF4A, CBP, CEBPA, p300, FOXA1, FOXA2 and GABPA revealed that multiple transcription factors can bind to this accessible region.

### 2.1.4 Genome wide impact of BMAL1 on DNA accessibility

We next sought to study how BMAL1 binding impacts DNA accessibility. We used a mouse model in which BMAL1 is unable to bind DNA (*Bmal1*<sup>-/-</sup>) [300] and performed DHS mapping at ZT6, near the maximal DNA binding activity of BMAL1 (Methods section 4.2). DNase I hypersensitivity at BMAL1-bound sites (detected in ChIP-seq) [273], such as the *Reverba* locus, was markedly decreased in *Bmal1*<sup>-/-</sup>, whereas control (unbound) regions like the *Gsk3* promoter showed no difference (Figure 2.10 A). Genome-wide, using the same methods as in wild-type mice (WT), we detected 56'201 DHSs at ZT6 in the *Bmal1*<sup>-/-</sup> animals. Of those, more than 38'000 sites were shared with the DHSs in WT. The imperfect overlap might largely reflect lower sampling when analyzing only one time point instead of the seven time points used for WT. Indeed, the overlap when considering only ZT6 in the WT was comparable (around 40'000 shared peaks).

Nevertheless, clear differences at DHSs with BMAL1 binding sites were observed. Namely, regions bound by BMAL1 [273] in the WT showed fewer DNase I cuts in the *Bmal1*<sup>-/-</sup> compared to the WT. This indicates that BMAL1 binding specifically impacts DNA accessibility at its target sites (Figure 2.10 B), consistent with the proposed pioneering function of BMAL1 [222]. While DNase I signals at those sites were also significantly lower at trough BMAL1 activity in the WT (ZT18), the KO showed even lower signals (Figure 2.10 C). The same analysis at sites bound by the E-box binding protein USF1 [301] did not show similar differences between WT and KO (Figures 2.10 D and E).

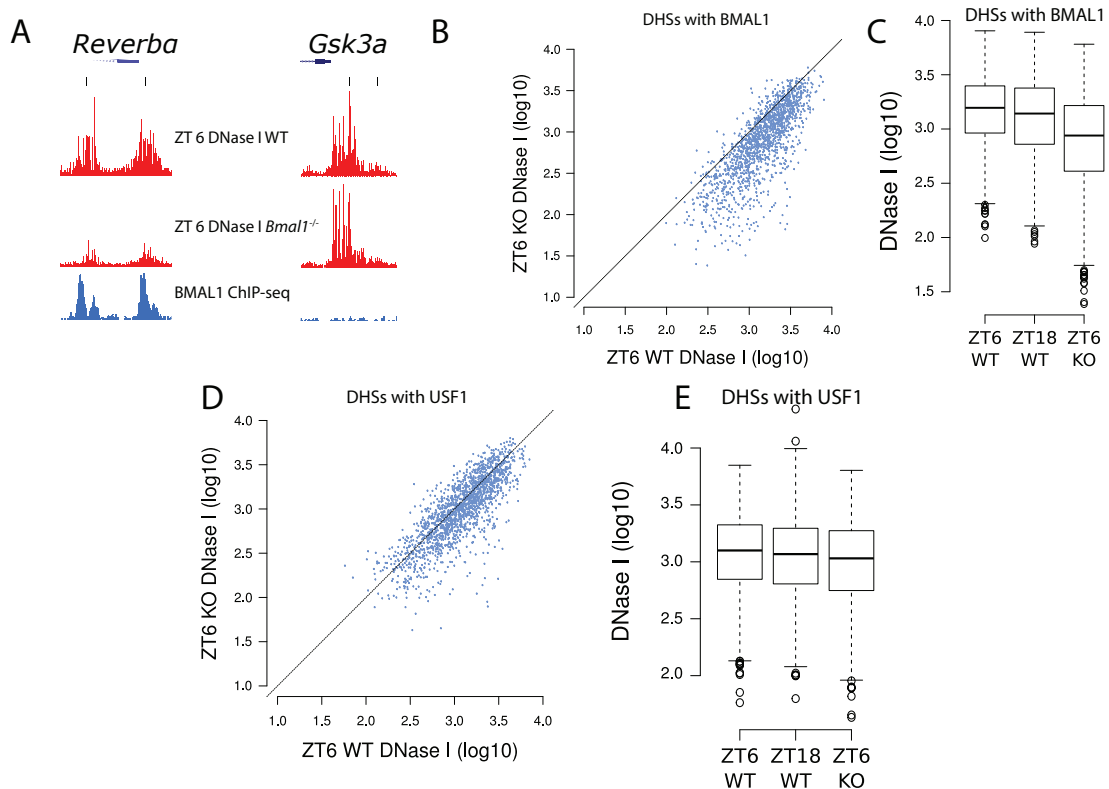


Figure 2.10 – Genome wide impact of BMAL1 on DNA accessibility. **A**) The *Reverba* (left) and *Gsk3a* (right) promoters where DHSs are indicated with black ticks at the top. DNase I signal (in red) is strongly reduced in the KO at BMAL1-bound sites (*Reverba* promoter, as indicated by *Bmal1* ChIP-seq signal in blue) but is similarly enriched in WT and KO at sites not bound by BMAL1 (at the *Gsk3a*) The vertical scale is the same for all 4 DNase I tracks, as well as for both *Bmal1* tracks.

**B**) Comparison of DNase I signal in *Bmal1*<sup>-/-</sup> versus WT at DHSs overlapping BMAL1 ChIP-seq peaks at ZT6. Almost all DHSs containing a BMAL1 binding site display lower accessibility in the *Bmal1*<sup>-/-</sup> suggesting fewer factors are bound to these regions at ZT6.

**C**) Boxplots showing DNase I intensity at the same sites as B, at peak (ZT6) and trough (ZT18) activities of BMAL1 in the WT and at ZT6 in the *Bmal1*<sup>-/-</sup>. Sites appear less accessible in the KO than during trough activity in the WT consistent with the lack of compensatory binding by other factors in the *Bmal1*<sup>-/-</sup>.

**D-E**) Same as **B-C** using overlap with USF1 ChIP-seq peaks to select DHSs. This protein targets similar motifs as BMAL1 (E-boxes). Lower signal in the KO is not observed at these sites.

### 2.1.5 Digital genomic footprinting and transcription factor binding around the clock

Owing to the 3D structures of protein-DNA interactions, genomic patterns of DNase I cleavage around transcription factor binding sites display factor-specific footprints [138, 237, 236, 132, 314, 350]. A previous study showed that BMAL1 binds DNA rhythmically, and that strong BMAL1 binding was frequently associated with tandem E-boxes [249] separated by 6 or 7 nucleotides bound by one or two BMAL1/CLOCK dimers, respectively [273]. Thus, we analyzed DNase I footprints at BMAL1 binding sites in function of time. Starting from BMAL1 ChIP-seq sites, we modified a mixture model for DNase I cuts to determine the optimal boundaries of the footprints at each time point, as well as the probability that the factor is bound to DNA (showed a footprint) for every site (see Methods section 4.6.1 and Appendix section A.1). We then analyzed footprints in BMAL1 binding sites containing tandem E-boxes separated by 6bp (E1E2-sp6). These DNA elements were previously shown to be bound *in vitro* by two BMAL1/CLOCK dimers [273]. Here, both E-boxes in the E1E2-sp6 motif appeared to be protected from digestion at ZT6, close to the maximal DNA binding activity of BMAL1, while at ZT18 only the 5' E-box displayed a footprint consistent with occupation by a transcription factor (Figure 2.11).

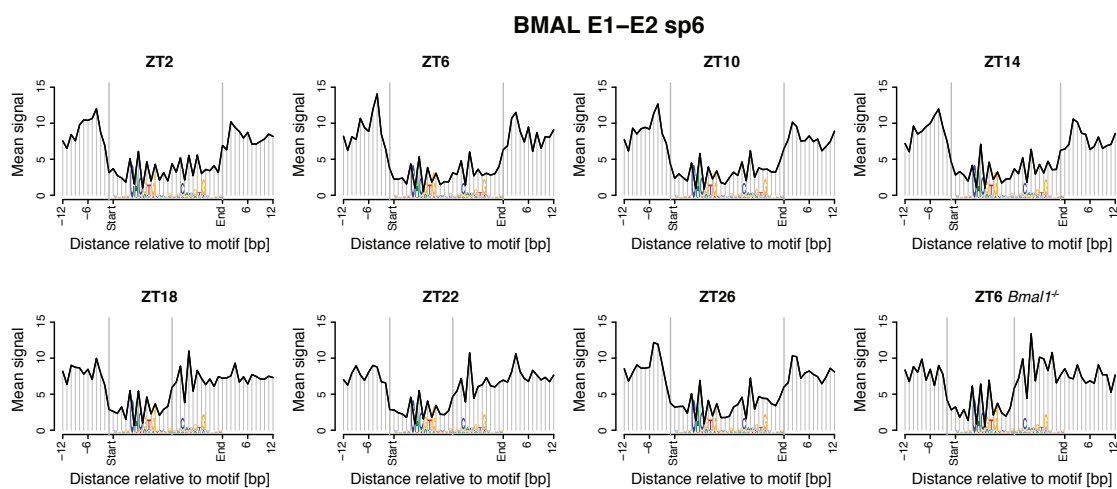


Figure 2.11 – Time-dependent footprint on double E-box motif with a spacer of 6bp overlapped by BMAL1 binding sites. DNase I hypersensitivity profile around double E-boxes with a spacer of 6 bp (E1-E2 sp6). We selected E1-E2 sp6 motifs overlapping a BMAL1 ChIP-seq peak and show the average of profiles classified as protected by the model. At ZT6, we observe that nucleotides of both E-boxes of the site are less accessible than flanking regions, indicating occupation by hetero-tetramers of CLOCK:BMAL1. In contrast, at ZT18, the width of the protected region is reduced by half, with the second, less conserved E-box no longer protected from digestion. In the KO, only the first E-box appears occupied.

Moreover, the footprint at ZT18 was undistinguishable from that in the *Bmal1*<sup>-/-</sup>, suggesting that other transcription factors bind BMAL1 sites when BMAL1 expression is low. The proportion of E1E2-sp6 motifs showing a footprint indicative of two BMAL1/CLOCK dimers varied across time points, with a maximum of 65% at ZT10, and minimum of 20% in the *Bmal1*<sup>-/-</sup>. Also, the binding dynamics of BMAL1 at E1-E2-sp7 was largely similar to that for E1-E2-sp6, though E1-E2-sp7 had both E-boxes predominantly protected only at ZT6, suggesting spacer-specific binding dynamics. In contrast, the footprints at BMAL1 binding sites with single E-boxes did not show significant changes in time or in the *Bmal1*<sup>-/-</sup>, again suggesting that other bHLH transcription factors compete with BMAL1. In fact, footprints of the bHLH transcription factor USF1 were largely similar to that of BMAL1 sites with single E-boxes, though the fraction of sites with clear footprints was reduced for USF1 compared to BMAL1 (Figure 2.12).

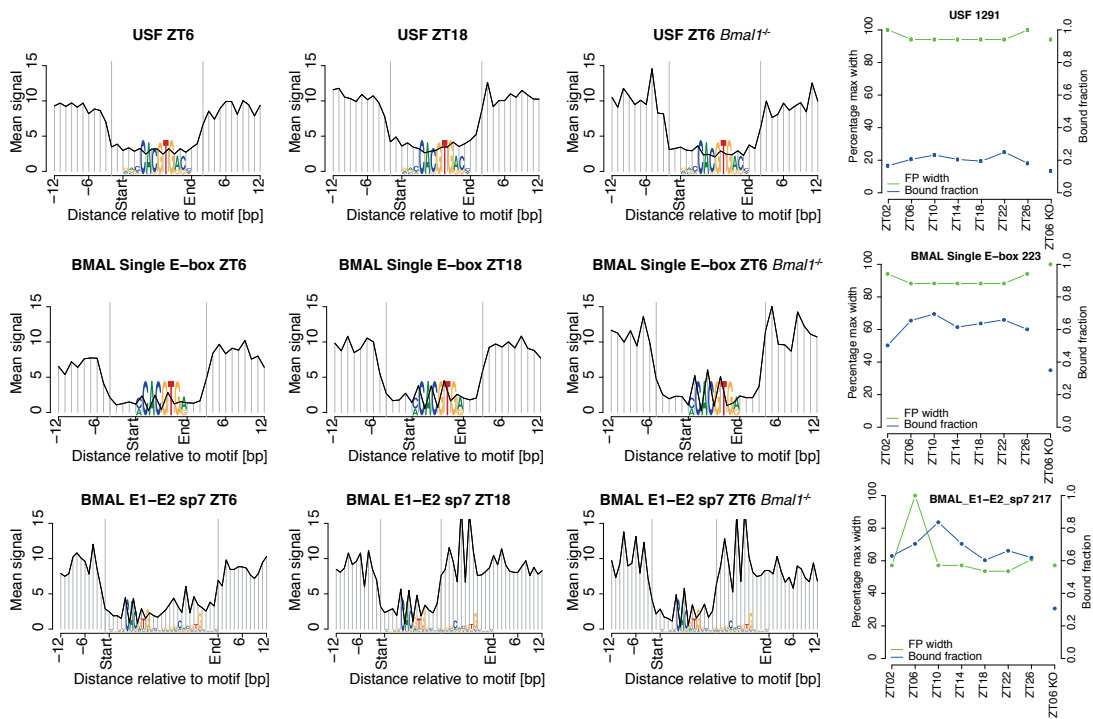


Figure 2.12 – Time-dependent footprint on single E-box of USF1 and BMAL1, and double E-box footprint with a spacer of 7 bp. Base-pair resolution patterns of DNA accessibility at USF1, BMAL1 single E-box, and BMAL1 double E-box motifs with a spacer of 7 bp, were studied to reveal dynamic changes. Using our mixture model to fit the width of protected regions at transcription factor motifs, within DHSs, overlapped by a high ChIP-seq signal ( $Z$ -score > 2) we computed the fraction, which appeared occupied at each time point. DNase I hypersensitivity average profile around each motif bound at ZT6, ZT18 and ZT6 in *Bmal1*<sup>-/-</sup>. The shape of the footprint reflects the protein-DNA interaction architecture. The bound fraction and the percentage of the maximum width are represented respectively in blue and green for each motif at each time point and at ZT6 in *Bmal1*<sup>-/-</sup> context.

## Chapter 2. Results

---

Finally, we examined temporal footprints at DHSs bound by other well-studied circadianly active TFs, ROR/REVERB, HSF1, SREBP and CREB (Figure 2.13). Interestingly, unlike for BMAL1/CLOCK, the shapes of the footprints for those factors did not change with time, and was also not affected in the *Bmal1*<sup>-/-</sup> condition. However, the fraction of sites showing footprints coincided well with the maximal transcriptional activity of the different factors. For example, footprints centered on REVERB $\alpha$ -bound ROR response elements (ROREs) showed the largest proportion of footprints at ZT22, which coincides with the phase of maximal ROR activity. We detected a low percentage of binding for ROREs, which is consistent with the fact that nuclear receptors have a low residence binding times [314] and therefore display a lower DNase I cleavage-protection pattern.

For HSF1, the number of footprints was maximal at ZT18, approximately four hours later than the previously reported peak activity [176], and for the feeding-induced SREBP this number peaked during the night, as expected [111, 211]. Finally, high confidence CREB binding sites [211] showed clearly marked and invariable footprints throughout the day, including in *Bmal1*<sup>-/-</sup> mice, consistent with the finding that CREB activity is regulated post-translationally on the DNA [94, 192, 298].



## 2.1. Chromatin landscape in mouse liver

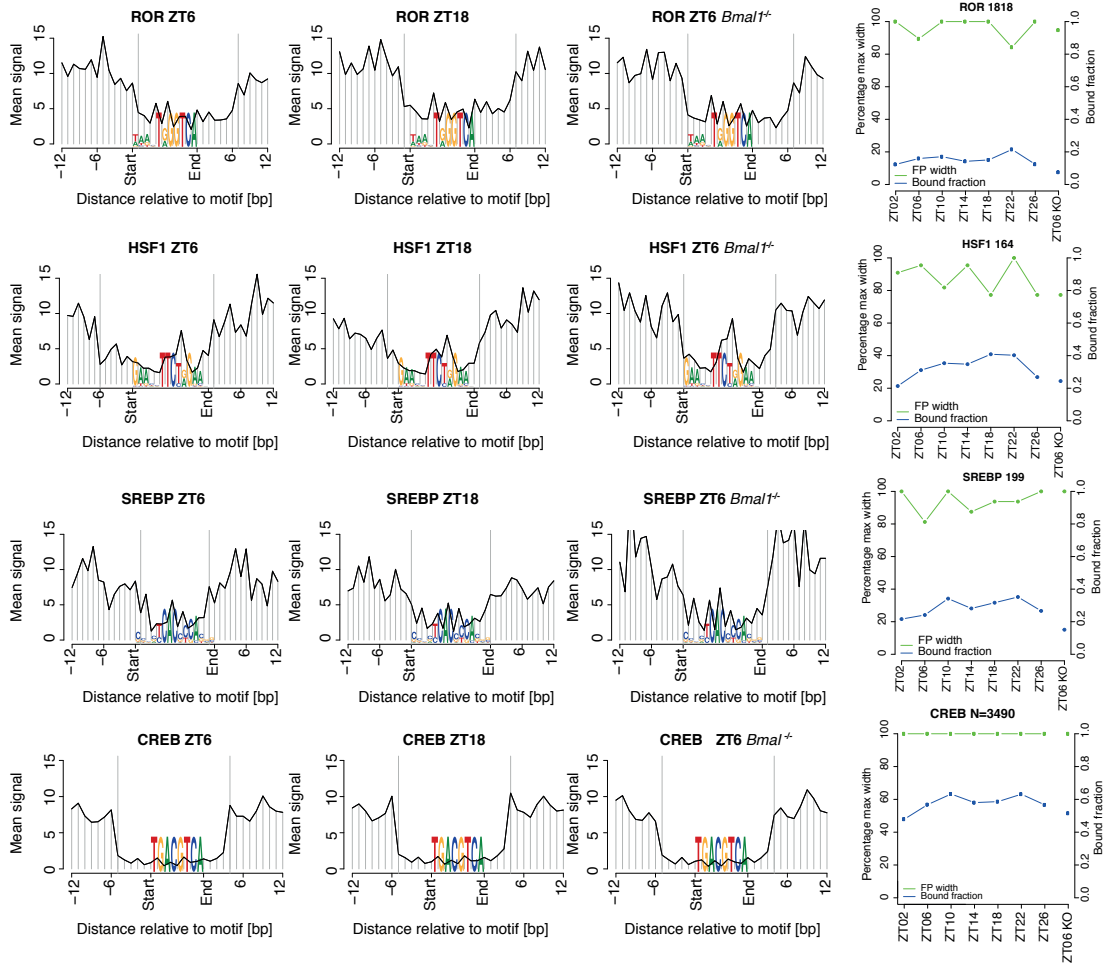


Figure 2.13 – Time-dependent footprint of ROR, HSF, SREBP and CREB. Base-pair resolution patterns of DNA accessibility at ROR, HSF, SREBP and CREB, were studied to reveal dynamic changes. Using our mixture model to fit the width of protected regions at transcription factor motifs, within DHSs, overlapped by a high ChIP-seq signal (Z-score > 2) we computed the fraction, which appeared occupied at each time point. DNase I hypersensitivity average profile around each motif bound at ZT6, ZT18 and ZT6 in *Bmal1*<sup>-/-</sup>. The shape of the footprint reflects the protein-DNA interaction architecture. The bound fraction and the percentage of the maximum width are represented respectively in blue and green for each motif at each time point and at ZT6 in *Bmal1*<sup>-/-</sup> context.

### 2.1.6 3D structure of BMAL1/CLOCK hetero-tetramer model bound on DNA

To better understand the time-varying footprint at BMAL1 sites and gain insights into how the BMAL1/CLOCK heterodimer occupies its tandem E-box-containing target sites, we used recently established 3D protein structures of single BMAL1/CLOCK complexes combined with structural modeling performed by Alexandra Styliani Kalantzi from Matteo Dal Peraro's lab (see Methods section 4.9). Two 3D models of the hetero-tetramer configuration were constructed. In a first model, the spacing between the two E-boxes was 6 bp (sp6) (figure 2.15 and figure 2.16) and in a second model the spacing was 7 bp (sp7). For the model of the single CLOCK:BMAL1 complex, we used the crystal structure of the heterodimeric BMAL1/CLOCK (pdb id: 4F3L) [141], in which we built the missing parts of the flexible loops. To link the single BMAL1/CLOCK model to the E-box, we used the complex crystal structure of BMAL1/CLOCK basic helix-loop-helix domains bound on the E-box (CACGTG) (pdb id: 4H10) [339]. We then superimposed the two single BMAL1/CLOCK E-box models, with the sp6 DNA and the sp7 DNA, forming the respective symmetric hetero-tetramer models. We found that the 6bp spacing between the two E-Boxes was suggesting an interaction between the two BMAL1/CLOCK complexes [234], although the 7 bp spacing seemed also favorable with a twist of 1° in the three interval base pairs. However, a conformation with base pair spacing less than 6 or more than 7 would make complex formation difficult because of conformational constraints. Thus, we found that the dynamics of transcription factor complexes on the DNA can change during the diurnal cycle, as reflected by the shape changes of DNase I footprints (section 2.1.5).

Using our 3D structure model with a spacer of 6 bp and molecular dynamic simulations (see Methods section 4.9), we were able to identify several residuals that are at the interface of the two hetero-dimers. These residuals are located in the PAS-B domain of CLOCK (table 2.1, figure 2.16). We observed two symmetrical interactions implicating a glutamic acid ( $E_{380}$ ) interacting with a glutamine ( $Q_{352}$ ) and a lysine ( $K_{335}$ ) interacting with a tyrosine ( $Y_{338}$ ) through a hydrogen bond. Moreover, the glutamic acid is negatively charged and the lysine is positively charged, suggesting a potential ionic interaction. There might be more hydrogen bonds forming between the two dimers on flexible loops on BMAL1 PAS-A domain. The flexible loops are more dynamic and therefore harder to characterize.

Table 2.1 – Residuals implicated in the interaction of the two hetero-dimers BMAL1:CLOCK

<i>Residuals A</i>	<i>Residual B</i>	<i>Distance (Å)</i>
$E_{380}$	$Q_{352}$	2.827
$K_{335}$	$Y_{338}$	2.835
$Q_{352}$	$E_{380}$	2.887
$Y_{338}$	$K_{335}$	2.882

Using our model with a spacer of 7bp and the molecular dynamic simulation, we calculated the distance of the center of mass between the two dimers (figure 2.14). The initial distance

## 2.1. Chromatin landscape in mouse liver

---

between the two dimers is 68.3 Å and after 130 ns of simulation, the distance decreased with a value of 62.4 Å, suggesting that the two BMAL1/CLOCK dimers are attracted by each other.

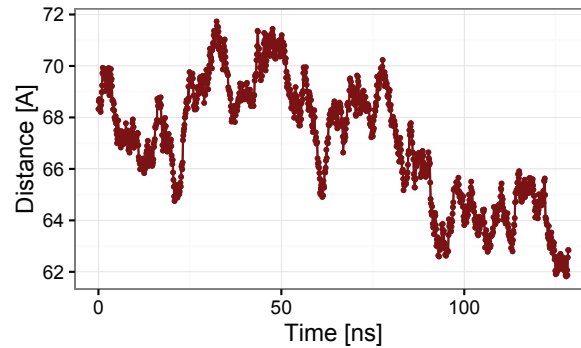


Figure 2.14 – Molecular dynamic simulation BMAL1/CLOCK hetero-tetramer model bound on DNA. Center of mass between the two BMAL1/CLOCK dimers calculated with a molecular dynamic simulation on the sp7 model.

Overall, these interactions might be critical for a functional molecular clock, and they deserve further experimental validations.

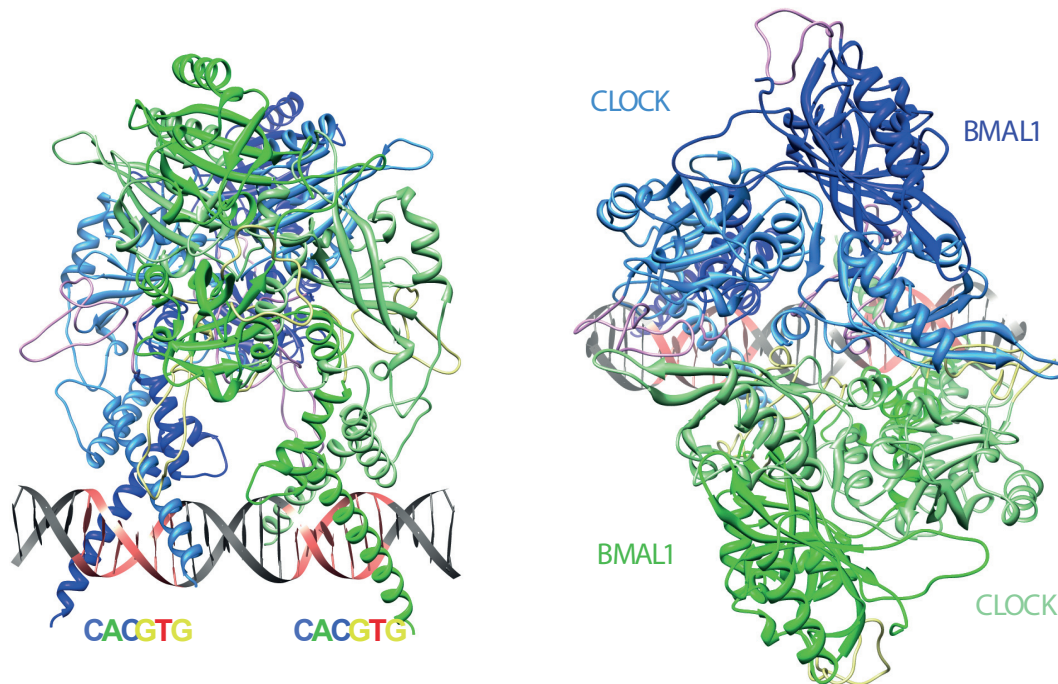


Figure 2.15 – 3D structure of BMAL1/CLOCK hetero-tetramer model bound on DNA. Two views of the 3D model of the CLOCK:BMAL1 hetero-tetramer based on the crystal structure of a dimer of two CLOCK:BMAL1 heterodimers occupying an E1-E2 sp6 site. Each heterodimer is drawn in a different color, green and blue respectively, while darker green and darker blue correspond to BMAL1 and lighter colors to CLOCK proteins. Information content along the DNA strands is shown in grey with highly constrained nucleotides of the motif in red. The proximity between two heterodimers suggests strong interaction between these complexes, and is consistent with the published data where cooperativity at tandem E-boxes has been shown by [249, 234], which has since been further corroborated by [273].

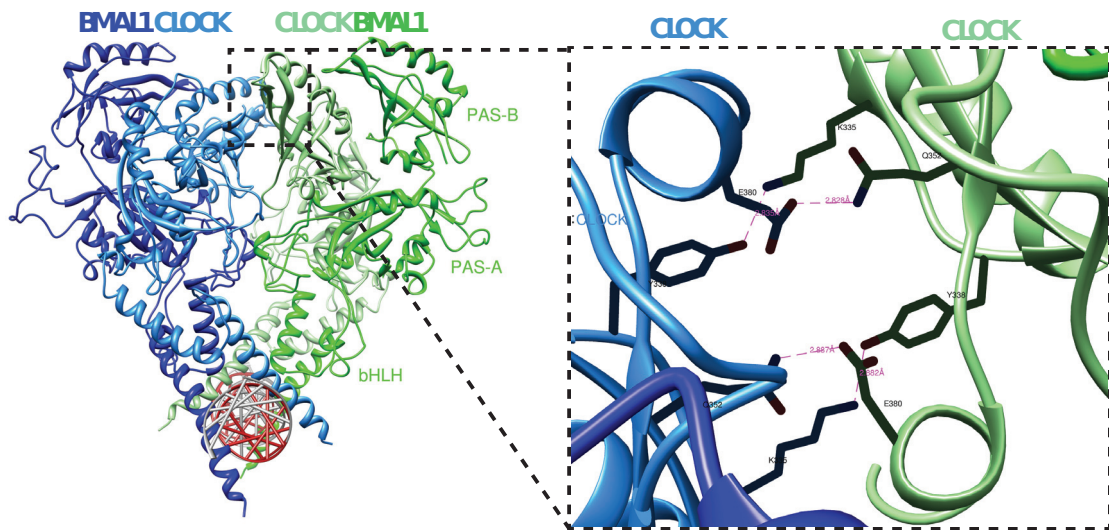


Figure 2.16 – BMAL1/CLOCK hetero-tetramer model bound on DNA: side view with interacting residuals. Side view of the 3D model of the CLOCK:BMAL1 hetero-tetramer with a zoom on the interface between the two CLOCK:BMAL1 hetero-dimers. The interacting residuals are mainly located on CLOCK. These residuals were identified thanks to molecular dynamics simulation [198] performed by Alexandra Styliani Kalantzi.

## 2.2 System specific chromatin accessibility and BMAL1 co-regulators

### 2.2.1 BMAL1 binding and DNase I hypersensitivity differ between fibroblasts and hepatocytes

As previously described, the rhythm in the core clock is produced and sustained by a transcription-translation feedback loop (TTFL), whose main activator is a heterodimer formed by two bHLH-PAS transcription factors, BMAL1 and CLOCK, which bind rhythmically to E-box motifs in the regulatory regions, such as TSS or enhancers of clock controlled genes [277, 108]. Genome-wide a large number of BMAL1-CLOCK targets have been identified by ChIP-seq in mouse liver [171, 273]. While the core TTFL is active and rhythmic genes have been observed in most cell types and tissues [312, 355], much less is known about the cell type-specific regulatory network and rhythmic output functions connected to it.

In a project in collaboration with Julia Cajan (a former PhD student in the Naef lab), we compared at BMAL1 ChIP-seq and DNase I hypersensitive sites in mouse liver and in NIH3T3 fibroblasts cells [273, 311].

To further identify the extent of circadian clock-controlled genes in NIH3T3 fibroblasts we set out to study DNA-binding of BMAL1 in dexamethasone (Dex) synchronized NIH3T3 cell cultures. Samples taken at the minimum (12h after Dex) and maximum (20h after Dex) binding time points were then submitted for deep sequencing (ChIP-seq). To assess similarities and divergences in BMAL1 binding in cultured NIH3T3 cells and mouse liver tissue (figure 2.17), the NIH3T3 fibroblasts data were compared to BMAL1 binding sites in liver from Rey *et al.* [273].

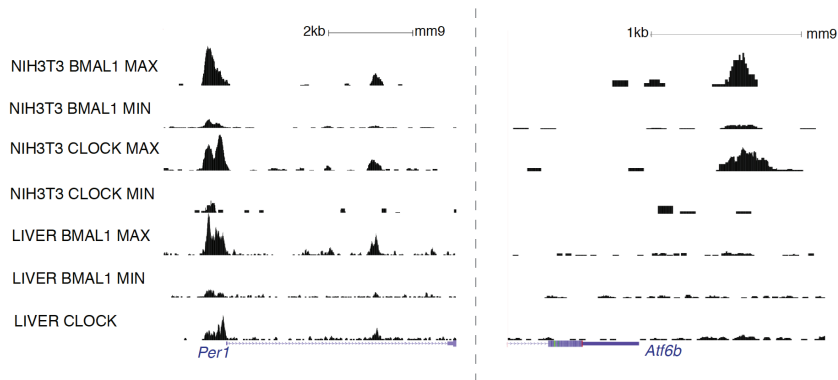


Figure 2.17 – Context specific binding of BMAL1/CLOCK. Examples of rhythmic BMAL1 and CLOCK. The left panel shows an example of a shared binding site in the promoter region of the *Per1* gene; the right panel shows the *Atf6b* promoter as an example of fibroblast-specific binding. Modified from Cajan, Sobel *et al.* (in preparation)

After standard peak detection, we filtered peaks that exhibited a low signal-to-noise ratio, which we also applied to the mouse liver data in [273] We obtained a list of 463 BMAL1 binding

## 2.2. System specific chromatin accessibility and BMAL1 co-regulators

---

sites in NIH3T3, and 1132 liver BMAL1 binding sites with an overlap of 319 shared sites. The observation that many BMAL1 sites were bound in a liver- or fibroblast-specific way raises the question whether the unbound sites are just weakly occupied or whether the chromatin surrounding those sites is in a poorly accessible, repressed state. We used published data sets of DNase I hypersensitive sites (DHSs) from the ENCODE consortium [311] to address this ambiguity. We re-analyzed DNase I signal to identify peaks in both conditions. We detected about 82000 DHSs in liver and 78000 DHSs in fibroblasts with an overlap of 26000 DHSs. This low overlap of DHSs between Liver and NIH3T3 suggest that the cellular regulatory programs are quite different between these two data sets. Moreover, we observed that 88% of BMAL1 binding sites were in a DHS in NIH3T3 (409/463) and 84% of BMAL1 binding sites were in a DHS in Liver (959/1132), which indicates that the chromatin accessibility is crucial for BMAL1 binding.

### 2.2.2 BMAL1 co-regulators are enriched in a cell-type specific manner

To identify putative BMAL1 co-regulators that could explain how BMAL1 can bind at different sites in hepatocytes and in NIH3T3 fibroblasts, we conducted a motif scan in a  $\pm 250$  bp window around the center of the binding sites. Since motif searches only indicate the occurrence of a certain sequence, but contain no information about the occupancy of the site, we also used DHS footprints (using Wellington algorithm from [256]) as a measure of bound (and hence presumably active) sites.

To control for general liver- or fibroblast-specific regulators, we used the combined set of DHSs from liver or fibroblasts [311], respectively, as input for the motif search since those should cover most active regulatory elements. This allowed to assess which motifs found around BMAL1 sites were associated with BMAL1 specifically in the liver or the fibroblasts (figure 2.18). As positive controls, E-box binding motifs (BHLHE40, MYC, MAX and USF) were clearly the dominant signal in both systems, although E-boxes seemed to be slightly overrepresented in NIH3T3 sites compared to liver.



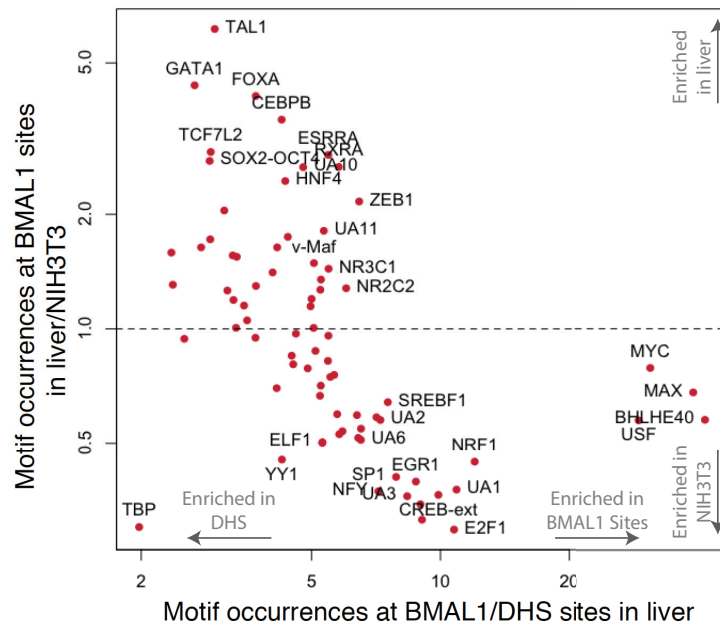


Figure 2.18 – Context specific-chromatin accessibility and BMAL1 co-regulators. A motif search using a PWM database derived from ENCODE ChIP-seq of 119 transcription factors [338] on all DHSs, as well as BMAL1 binding sites in liver and NIH3T3 fibroblasts, reveals motifs that are preferentially enriched in one dataset or the other and differentiates between general regulators (DHSs) or BMAL1-specific coregulatory elements. The X-axis represents the enrichment of motifs (in a footprint) in sites of BMAL1 versus DNase I hypersensitive sites, and the Y-axis represents the enrichment in Liver versus NIH3T3.

In the liver, we detected several motifs for known tissue-specific regulators that showed, however, low specificity for BMAL1 sites. These include RXRA (three-fold enriched in the liver and five-fold enriched in BMAL1 over all DHSs) or members of the FOXO family (four-fold enriched in liver, four-fold enriched in BMAL1 over all DHSs). Other factors, notably ZEB1, were more specifically associated with BMAL1 binding, but less specific for the liver (2.1 fold enriched in liver, seven-fold enriched in BMAL1 over all DHSs). However, in liver, specific recruitment of BMAL1 to DNA seems to be overall achieved mainly through general tissue-specific regulators. This is in contrast to fibroblasts where we observed stronger enrichment of specific motifs that were preferentially associated with BMAL1 sites. These motifs pointed to transcription factors such as E2F1 (3 fold enriched in NIH3T3, 15 fold enriched in BMAL1 over all DHSs) and NRF1 (2.1 fold enriched in NIH3T3, 15 fold enriched in BMAL1 over all DHSs). Overall, the overrepresentation of these motifs that coincides with a DNase I footprint indicates that a large percentage of these motif-containing sites were occupied and thus potentially active.



## 2.2. System specific chromatin accessibility and BMAL1 co-regulators

In order to confirm these observations, we analyzed the motif content of 100K random genomic locations, and we computed the ratio of the number of motif match over the sum of regions considered for several interesting candidates (figure 2.19 and 2.20). Again, we observed that EGR1, NRF1, BHLH40 and ZEB1 are highly enriched in BMAL1 sites in both conditions, and it seems that NRF1 and EGR1 are more enriched in fibroblasts. These motifs are not enriched in the random context which suggests a functional role of potential tissue-specific BMAL1 co-regulators. On the other hand, we detected several potential liver-specific co-regulators as DBP, NR3C1, HNF4, HNF1, HSF1, CEBPB, TCF7L2 and FOXO1. Interestingly some of these factor as HNF1 and HNF4 are already known as liver-specific factors, and FOXO1 and TCF7L2 are related to feeding-fasting cycle [241]. Moreover, DBP is a direct target of BMAL1 [278, 277] and its binding site (D-box) is upstream of many clock controlled genes, which are implicated in hormonal and metabolic control [296]. Finally, NR3C1 (also called Glucocorticoid receptor,GR),has been shown to synchronize hepatic circadian transcriptome [266].

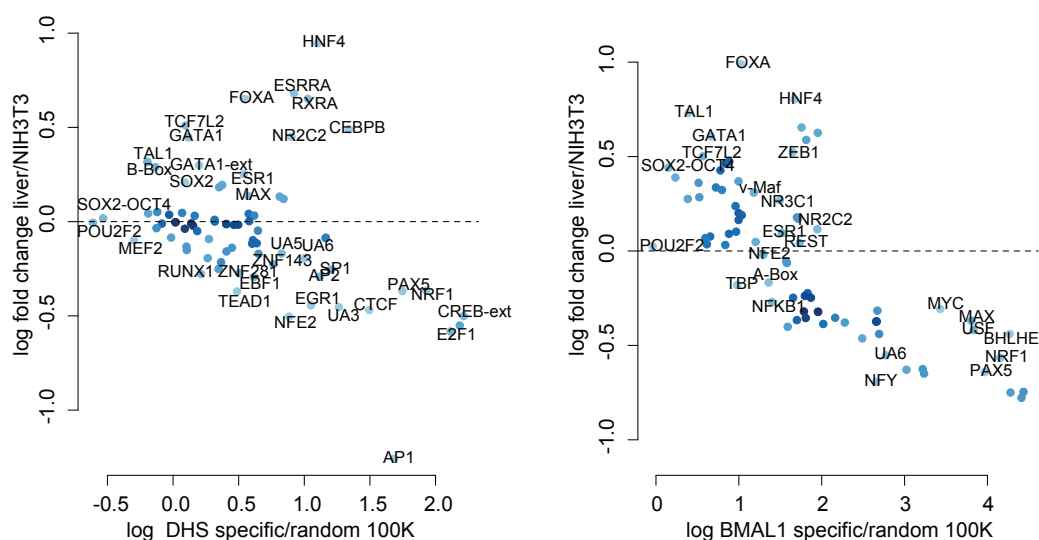


Figure 2.19 – Enrichment of liver/NIH3T3 specific regulators in DHS and in BMAL1 binding sites. Enrichment of TFBS in BMAL1 and DNase I hypersensitive sites, in liver or NIH3T3, over 100K random genomic locations as a control. The Y-axis represent the log fold change between liver and fibroblasts and the X-axis represent the log fold change between DHS or BMAL1 binding sites over the 100K the random genomic locations.

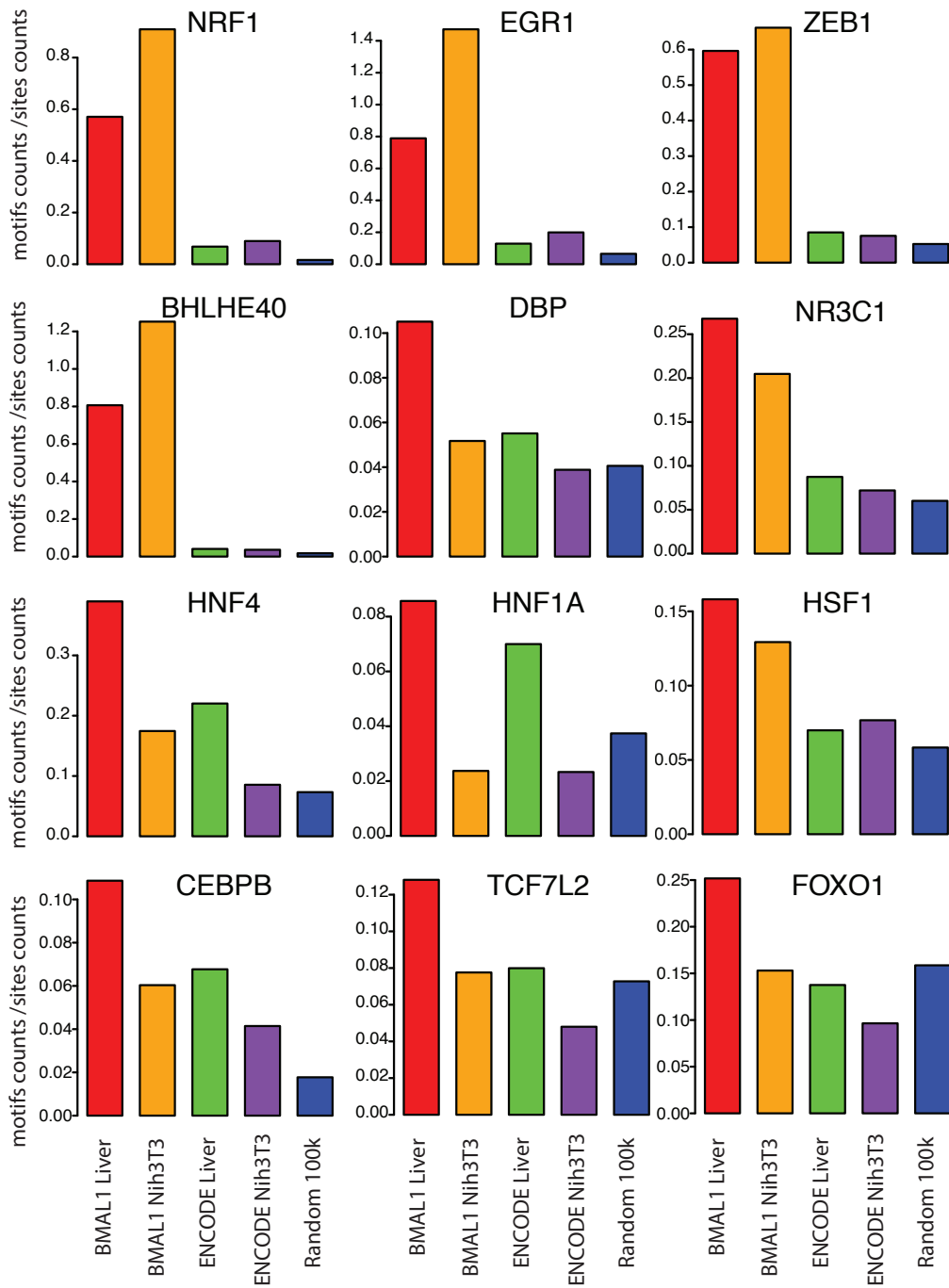


Figure 2.20 – Predicted liver/NIH3T3 specific regulators. Motif occurrence over total site counts for several candidates TF in BMAL1 and DNase I hypersensitive sites in liver ( $N_{BMAL1,liver} = 1132, N_{DHS,liver} = 81973$ ) or NIH3T3 ( $N_{BMAL1,NIH3T3} = 463, N_{DHS,NIH3T3} = 77618$ ) and in 100K random genomic locations as control.



### 2.2.4 ZEB1 was identified as a potential co-regulatory factor of BMAL1

Julia Cajan tested the potential implication of ZEB1 (Zinc Finger E-Box Binding Homeobox 1) in circadian gene regulation using ChIP-qPCR and shRNA. She found that this protein binds upstream and at the TSS of *Nr1d1* in a rhythmic manner (28h-32h after Dex) but in anti-phase of BMAL1 (20h after Dex) in NIH3T3 cells (figure 2.22 A and B). Moreover, the knock-down experiment of *Zeb1* mRNA using shRNA, monitored with a *Dbp*-luciferase reporter, resulted in a significantly prolonged period phenotype in NIH3T3 fibroblasts (figure 2.22 C), consistent with results of an RNAi screen in human U2OS cells [353]. In addition, ZEB1 was recently associated with adipogenesis [118] and may represent another link between the circadian clock and the metabolism.

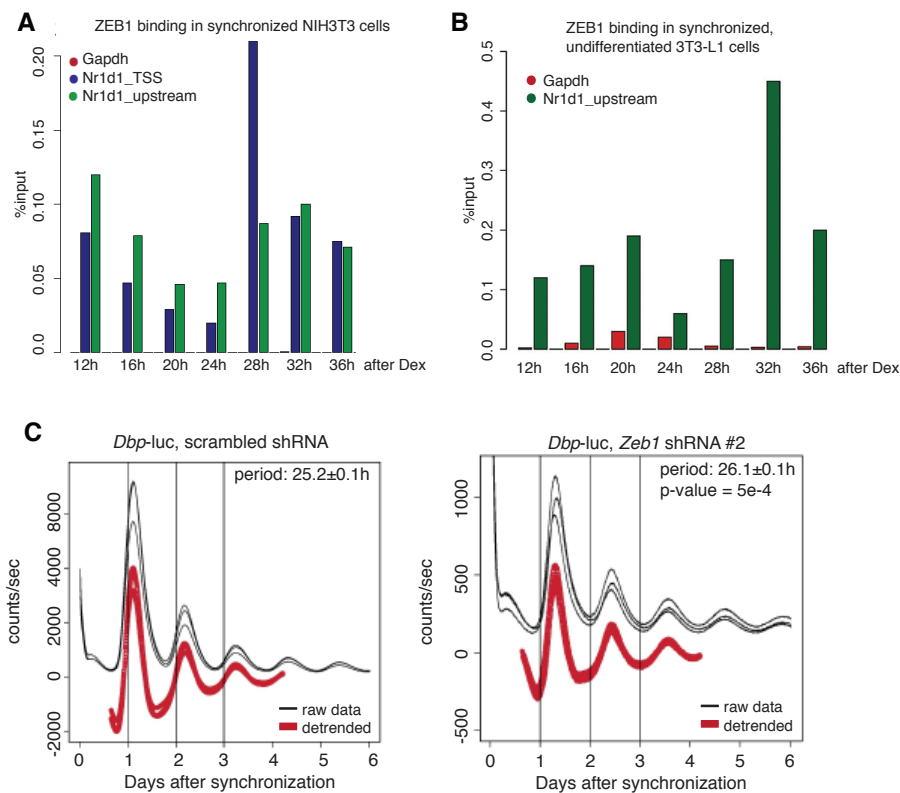


Figure 2.22 – ZEB1 was identified as a potential co-regulatory factor of BMAL1. (A) ZEB1 binds to shared BMAL1 sites rhythmically, but in antiphase to BMAL1, peaking at 28-32h after Dex in NIH3T3 and (B) undifferentiated 3T3-L1 cells. (C) *Dbp*-luc as a circadian reporter in NIH3T3 fibroblasts shows substantial period elongation upon *Zeb1* KD (mean for stable KD cell line: 26.1h ± 0.1h, control cell line: 25.2h ± 0.1,  $p$ -value = 5e-4,  $t$ -test, N=6). Modified from Cajan, Sobel *et al.* (in preparation)

**2.2.5 Digital genomic footprinting confirms predicted liver/NIH3T3 specific factor**

To study the transcription factor occupancy of the above motifs in more details, we analyzed nucleotide resolved DNase I cleavage pattern obtained for each motif (ENCODE DHS data, [311]) (figure 2.23) using the same sites in both tissues. DHS footprints detected at E-box sites (here BHLHE40) in liver and NIH3T3 fibroblasts show similar mean DNase signals and clear signatures of a bound protein. In contrast, footprints found at the Nuclear respiratory factor 1 (NRF1) and Hepatocyte nuclear factor 4 (HNF4) motifs show clear preferences for fibroblast- or liver-specific BMAL1 sites, respectively. Interestingly, a decrease of NRF1 expression was associated with insulin resistance in skeletal muscle in diabetic individuals [250]. Moreover, NRF1 targets encode key enzymes in oxidative metabolism and mitochondrial function.

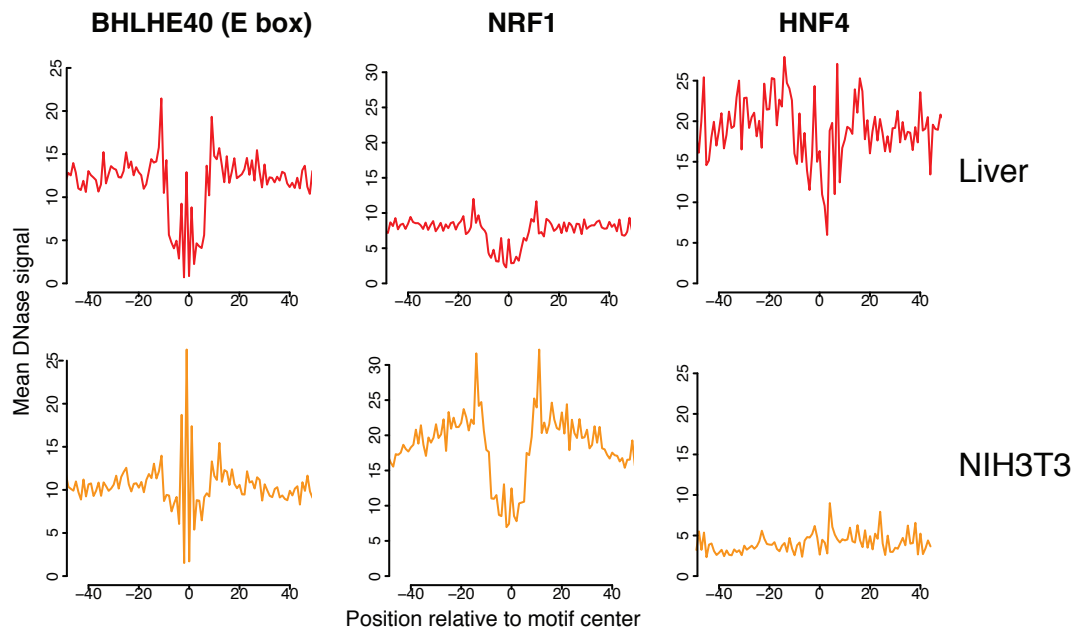


Figure 2.23 – Context-specific chromatin accessibility of E-box, NRF1 and HNF4. DNase I footprint in Liver or NIH3T3 around E-Box (BHLH40) a motif enriched in BMAL1 binding sites, NRF1 a motif enriched in NIH3T3 and HNF4 a liver-specific motif. These cleavage patterns were obtained using the same sites (PWM match in DHS in liver and/or NIH3T3 fibroblasts) in both tissues.

Taken together, these observations suggest that tissue-specific programs mediated by co-regulatory factors are necessary to support tissue-specific circadian gene regulation.

### 2.3 Phenotypic consequences of whole body *Bmal1* knockout

#### 2.3.1 Phase specific regulation by transcription factors in WT and in *Bmal1*<sup>-/-</sup> mouse liver under night restricted feeding

To further study how the core clock and feeding-fasting cycles regulate diurnal oscillations of gene expression in the liver at the transcriptional level, we combined temporal Pol II loadings at TSSs in WT and *Bmal1*<sup>-/-</sup> mice and transcription factor binding sites in accessible chromatin regions. Using DHSs and a collection of about 1900 position-weight matrices for TF-DNA affinities (Methods section), we identified DNA sequence motifs that explain rhythmic Pol II patterns in wild-type and KO. Therefore, we modified previously described linear regression models [273, 26, 44], but used sparse regression to identify a set of transcriptional activities (strictly speaking DNA motifs) represented by phase (time of maximal activity) and amplitude (Methods). In this model, motif activities are linearly combined according to the presence of the corresponding DNA motifs within DHSs, as in the phase vector model [326]. Specifically, we fitted the TSS Pol II signal of all genes using the motif content in all DHSs annotated to the corresponding genes, as described above. This enabled us to take into account, in addition to the proximal promoter, a collection of putative regulatory regions that may control the expression of the focal gene at a given vicinity of their respective TSS (Figure 2.24 A). We observed that the inclusion of Distal DHSs up to a vicinity of 50Kbp was increasing deviance ratio in both genotypes (Figure 2.24 B). This emphasized that enhancers (as represented by distal DHSs) play an important role in fine-tuning circadian gene transcription.

### 2.3. Phenotypic consequences of whole body *Bmal1* knockout

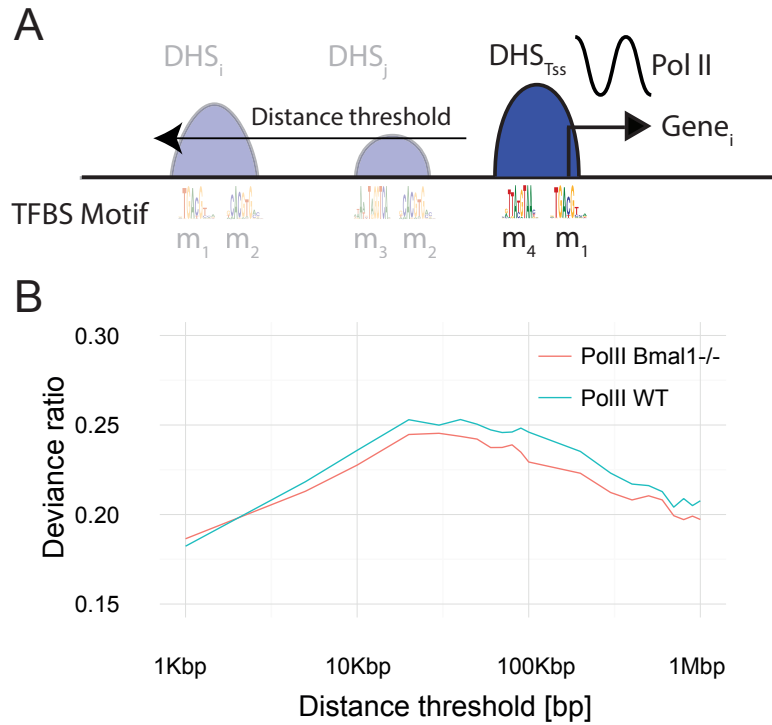


Figure 2.24 – Computational method for discovering phase specific activity of cis-regulatory elements. **A**) Scheme of the penalized generalized linear model. The distance threshold is symmetric regarding the TSS; it allows including intergenic or intragenic DHS motif content in the linear model. A DHS might be assigned to a multiple active gene thanks to the distance threshold around the TSS. **B**) Deviance ratio in WT and in *Bmal1*<sup>-/-</sup> for Pol II loading (at the TSS of all active genes) is represented in function of the distance threshold. The deviance ratio is defined as the fraction of (null) deviance explained by the model.

## Chapter 2. Results

In WT mice, our predictions revealed that core clock transcription factors were among the key phase determinants, and in particular Ror responsive elements (RRE), as reflected by the strong inferred activity of RRE motifs at late night using Pol II-bound regions (Figure 2.25). Other major motif groups included D-Box factors with a peak activity around ZT12 and E-boxes around ZT8, as previously described. However, motifs for factors that are not part of the core circadian clock also emerged in our analysis, as illustrated by Forkhead TF (FOX) motifs around ZT3, the CREB motif at ZT7, and GR motifs around ZT10. In *Bmal1*<sup>-/-</sup> mice (Figure 2.25), it was immediately apparent that the contribution from E-Box, RREs, and D-Box motifs was greatly reduced, and that other factors were responsible for phase-specific Pol II accumulation. One of the Forkhead box motif (FOX) appeared as the variable with the highest coefficient and a phase-specific contribution around ZT5 (FOXO). The FOX family contains 41 protein in mouse, and we can not know which protein of the family is responsible for this important phase-specific activity. Moreover we observed that some other FOX motifs from our database (FOXA1,FOXB1,FOXJ) had a small phase specific activity at ZT18. Interestingly, at the nuclear protein level, two FOX proteins, namely FOXA1 and FOXA3, were detected using mass-spectrometry with similar phase (personal communications with Jingkui Wang and Frederic Gachon). We also noted a larger relative role for the HSF, GR and CREB motifs in *Bmal1*<sup>-/-</sup>. On the other side, several motifs showed stable activity such as SREB.

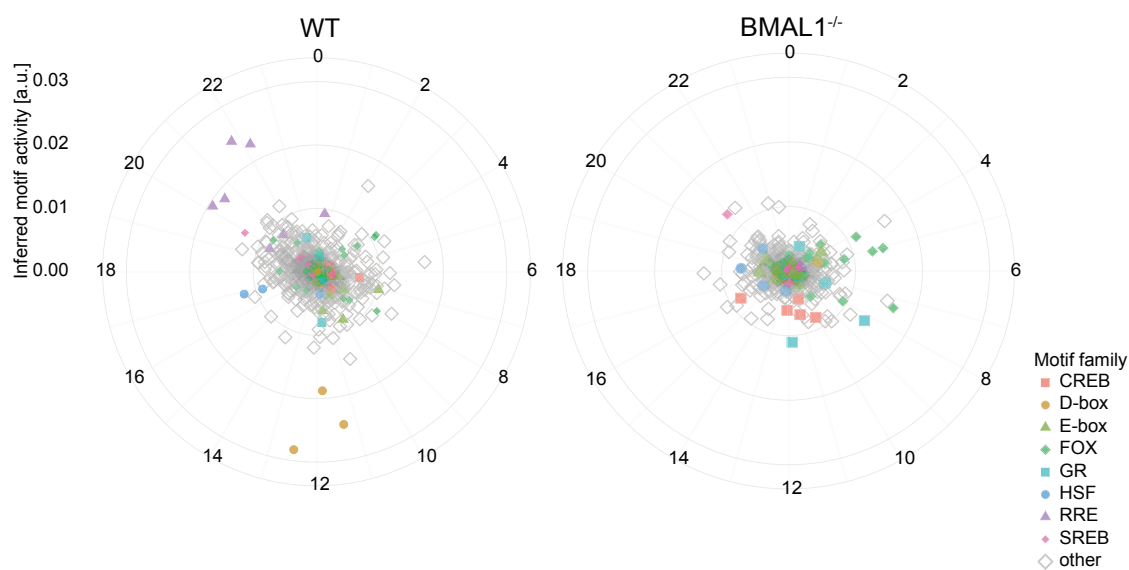


Figure 2.25 – Phase-specific activity of cis-regulatory elements in WT and *Bmal1*<sup>-/-</sup> mouse liver. We performed a penalized linear model in WT and in *Bmal1*<sup>-/-</sup> with a distance threshold of 50Kbp, and with a parameter  $\alpha$  of 0.1 (Elastic-net mode) and a  $\lambda$  of 0.1. We used motifs (PWMs) from four databases, notably JASPAR, SELEX, ENCODE (ChIP-seq) and TRANSFAC. In addition, We highlighted with the same color closely related motifs from the different sources (CREB, D-box, E-box, FOX, GR, HSF, RRE and SREB). The 40 motifs with highest inferred activity in WT and in *Bmal1*<sup>-/-</sup> are presented in table 2.2 and 2.3



### 2.3. Phenotypic consequences of whole body Bmal1 knockout

Table 2.2 – Top 40 inferred motif activity in WT mouse liver

PWM ID	Activity	Phase	Consensus	Sum in-formation content	Motif sum in DHS	Target gene sum	Source	Annotated Family
NFIL3_MA0025.1	2.826E-2	12.4	TTATGTAACGT	14.13	3597	2447	Jaspar	D-box
V_RORA2_01	2.454E-2	21.7	GTAACAT	11.92	5690	4005	Transfac	RRE
NFIL3_M135	2.443E-2	11.3	AGAGATAAGA	9.8	4455	3410	Selex	D-box
RORA_M208	2.268E-2	22.1	GGGGGGGGGCCA	14.23	8234	5079	Selex	RRE
V_RORA_Q4	1.937E-2	20.1	GGTAAGTA	6.50	7188	4786	Transfac	RRE
V_CREBP1_01	1.877E-2	11.8	TGACCTTTG	12.90	2287	1914	Transfac	D-box
RORA_1_MA0071.1	1.847E-2	20.5	CAAACGTAAACAAT	14.82	10192	5835	Jaspar	RRE
V_SOX13_03	1.710E-2	5.6	CACATTCCTCCG	15.67	3438	2546	Transfac	other
V_DMRT1_01	1.579E-2	2.0	AGGTCAG	8.34	6270	4288	Transfac	other
V_PAX2_01	1.471E-2	10.5	GGAGAAGCAG	6.44	5262	3781	Transfac	other
Cebpb_M130	1.354E-2	7.9	GGGAACAGATGGTCTC	15.89	3776	3007	Selex	other
SREBF1	1.286E-2	19.9	CTGAACCTTGACC	13.81	11144	6405	Encode	SREB
V_AR_02	1.243E-2	6.6	GATGAACCTCCGACCCGTTT	16.44	9443	5654	Transfac	GR
V_CEBPE_Q6	1.226E-2	19.1	CTTTGTG	7.77	6529	4405	Transfac	other
V_VDR_Q6_01	1.195E-2	12.1	ACCGTTAACGGT	15.28	14751	7165	Transfac	other
V_HSF1_Q6	1.193E-2	16.8	ACCGGAAGTG	10.25	11599	6735	Transfac	HSF
V_GCM1_04	1.189E-2	22.0	TTCCGAGAAGACCA	14.92	10327	6126	Transfac	other
V_PBX1_01	1.172E-2	4.3	GGGGGGGGGGGGCC	12.80	4635	3166	Transfac	other
V_MYB_03	1.156E-2	6.0	ATCCACAGGTGCGAAAA	12.01	8157	5325	Transfac	other
PAX1_M87	1.144E-2	13.7	GATTTAATGACC	13.62	6997	4893	Selex	other
V_FOXO3A_Q1	1.133E-2	8.2	CTTTAAGTACTTAATG	13.19	10652	5524	Transfac	FOX
V_CMAF_01	1.125E-2	21.0	GCCTCCCC	13.55	9443	5708	Transfac	other
V_ATF4_Q6	1.112E-2	8.5	CGGGGG	6.57	8126	5184	Transfac	other
V_FOXO4_01	1.103E-2	3.8	TATTGGTAATTACCTT	10.66	10040	5771	Transfac	FOX
V_MSX3_01	1.103E-2	6.9	GGTCCCGCCCCCTTCTC	14.38	3222	2558	Transfac	other
V_AIRE_02	1.099E-2	20.9	AGATAAG	8.22	4206	3250	Transfac	other
ZNF784_M30	1.075E-2	11.0	TACCCACAATGCATTG	22.05	7763	5056	Selex	other
V_DOBX4_01	1.073E-2	5.1	AACCCCTTGTATGC	13.70	3420	2758	Transfac	other
FOXB1_M138	1.065E-2	3.9	GGGGAATAGAGTTG	10.66	7497	4869	Selex	FOX
V_IPF1_05	1.047E-2	16.5	ACCGGAAGTG	11.35	4420	3230	Transfac	other
V_IK3_01	1.028E-2	11.7	TTGGCA	6.00	7618	5080	Transfac	other
V_MYCMAX_02	1.018E-2	7.0	TTATGTACTAATAA	16.78	8613	5412	Transfac	E-box
ESRRA_M190	1.012E-2	20.8	TTGGTGACGTCC	11.55	11845	6395	Selex	other
PBX1_MA0070.1	1.01E-2	9.1	CCATCAATCAAA	14.64	4844	3152	Jaspar	other
V_PBX1_02	1.010E-2	19.5	TTTGACCTCCAGTGACCCCC	17.17	5228	3392	Transfac	other
V_IRF2_01	1.007E-2	21.1	ACCGGAAGTA	10.85	8725	5377	Transfac	other
En1_MA0027.1	1.006E-2	7.2	AAGTAGTGTC	6.42	6931	4794	Jaspar	other
V_STAT5A_Q6	1.005E-2	15.0	AACCCGGAAGTG	16.25	12414	5612	Transfac	other
V_HNF3B_Q6	9.7E-3	6.8	AGATAA	8.19	14860	6475	Transfac	other
V_AR_Q6_01	9.6E-3	10.0	TAGCCAGACAG	9.99	14242	6919	Transfac	GR

## Chapter 2. Results

Table 2.3 – Top 40 inferred motif activity in *Bmal1*<sup>-/-</sup> mouse liver

PWMID	Activity	Phase	Consensus	Sum in-formation content	Motif sum in DHS	Target gene sum	Source	Annotated Family
V_FOXO4_01	1.664E-2	7.3	TATTGGTAATTACCTT	10.66	10040	5771	Transfac	FOX
V_DMRT7_01	1.628E-2	4.1	AGGTCAG	8.34	6270	4288	Transfac	other
V_SOX12_03	1.544E-2	7.1	GATTTCCCATCATGCCTTGC	17.54	4107	3065	Transfac	other
V_FOXO1_Q5	1.453E-2	5.0	GAAACTAGTTAACATC	12.86	17895	7098	Transfac	FOX
Ar_M188	1.368E-2	8.2	TAAACCGGAAGTTCGT	16.94	5199	3708	Selex	GR
FOXb1_M138	1.289E-2	5.0	GGGGAATTAGAGTTG	10.66	7497	4869	Selex	FOX
SREBF1	1.26E-2	20.9	CTGAACCTTTGACC	13.81	11144	6405	Encode	SREB
V_ZEC_01	1.265E-2	17.8	AGCTCATTAT	11.69	7146	4752	Transfac	other
V_COUPTF2_Q6	1.262E-2	21.3	GGTACAGGGTGTCT	14.12	24206	8644	Transfac	other
V_DOBX4_01	1.141E-2	8.4	AACCCTTTGTATGC	13.70	3420	2758	Transfac	other
V_FOXJ3_06	1.139E-2	4.1	AAAAGCGGATATTG	14.29	10263	5586	Transfac	FOX
Ar_MA0007.1	1.102E-2	11.8	ATAAGAACATCGTGTACCCGCC	15.70	8149	5122	Jaspar	GR
V_HDX_01	1.089E-2	22.9	TCAGGAGTTCGAGACC	15.76	4598	3434	Transfac	other
RARA_M207	1.072E-2	15.9	TTGTTGTTTACATA	16.38	15370	7296	Selex	other
V_PITX2_Q2	1.043E-2	17.2	CATTTCCGTT	11.63	6023	4144	Transfac	other
V_PPARG_Q3	1.009E-2	9.3	GATTTCCGGAAAATG	15.21	17565	7729	Transfac	other
V_HNF1_Q6_01	1.002E-2	23.4	CCACACCCTG	13.40	6169	4030	Transfac	other
V_KAIS0_01	9.879E-3	19.8	AAAGGTCAAA	9.62	15679	7585	Transfac	other
NR1H2::RXRA_MA0115.1	9.4E-3	9.65	AAAGGTCAAAGGTCAAC	27.87	16448	7417	Jaspar	other
V_PAX_Q6	9.425E-3	15.6	GGCCCGGGCCG	14.31	15042	7272	Transfac	other
V_CDP3R3HD_01	9.385E-3	19.2	TCTACCGGAAGTGGGT	10.95	2812	2206	Transfac	other
V_FOXO3A_Q1	9.345E-3	8.0	CTTTAAGTACTTAATG	13.19	10652	5524	Transfac	FOX
V_XBP1_Q2	9.234E-3	14.4	TATTACATAACA	11.55	4097	3250	Transfac	other
V_HNF3G_Q4	9.1E-3	7.4	TTGTTA	4.56	14879	6504	Transfac	other
V_AR_Q6_01	9.099E-3	9.3	TAGCCAGACAG	9.99	14242	6919	Transfac	GR
V_ICSBP_Q6	9.080E-3	4.9	GCCGCCATTTTG	16.30	9755	5777	Transfac	other
V_SF1_Q6_01	8.629E-3	7.7	CTCAGCCAATCAGCGC	12.92	16821	7412	Transfac	other
V_FOXK1_Q4	8.506E-3	5.1	TGAAGGGATTAATCATC	12.98	7917	4252	Transfac	FOX
V_RSRFC4_01	8.430E-3	10.6	AGATAG	6.62	3075	2477	Transfac	other
V_CREB_Q2_01	8.238E-3	10.0	GCTGAC	6.23	10250	5886	Transfac	CREB
V_HOXA7_01	8.175E-3	5.3	ACCGGAAGTG	9.81	3810	3001	Transfac	other
V_HES1_Q6	8.080E-3	13.5	CCCTCCCCCA	12.67	19505	8557	Transfac	other
V_HNF6_Q6	7.906E-3	1.5	TGCGTGGGGCT	14.61	4803	3361	Transfac	other
V_IRF4_Q4	7.569E-3	10.7	ACCGGATGTA	10.23	7264	5030	Transfac	other
Ddit3::Cebpa_MA0019.1	7.2E-3	16.8	AGATGCAATCCC	11.65	4990	3756	Jaspar	other
V_NFYC_Q5	7.200E-3	17.9	CGACCAACTGCCGTG	12.31	12888	6811	Transfac	other
V_FREAC2_01	7.154E-3	1.9	AGTAATTAATTACTTC	13.78	6198	4334	Transfac	other
V_HSF1_Q6	7.102E-3	18.2	ACCGGAAGTG	10.25	11599	6735	Transfac	HSF
V_HNF4_Q6	7.021E-3	21.3	ATTGATTGACAGGG	13.90	15858	7449	Transfac	other
V_ATF_01	6.963E-3	11.0	GAAGTGTGTTTCAGAC	17.72	7111	5004	Transfac	CREB

## 2.3. Phenotypic consequences of whole body *Bmal1* knockout

### 2.3.2 CREB phosphorylation is still oscillating in *Bmal1*<sup>-/-</sup> mouse liver

During fasting, mammals maintain normal glucose homeostasis by stimulating hepatic gluconeogenesis. Hepatic gluconeogenesis is known to be regulated by hormonal cues as circulating glucagon and more recently by the circadian clock, which coordinates glucose metabolism with changes in the external environment [154, 162]. CREB activity during fasting was shown to be modulated by CRY1 and CRY2, which are rhythmically expressed in the liver [354].

In our analysis, CREB was found among the most delayed TF activities inferred by the generalized linear model in *Bmal1*<sup>-/-</sup> mouse liver. To test this prediction, Benjamin Weger (A post-doc from the Gachon lab) measured nuclear levels CREB and pCREB using Western blots in nuclear extract from four independent livers every two hours in WT and *Bmal1*<sup>-/-</sup> mice (figure 2.26 A and B). In this analysis we were able to replicate the pattern of CREB activity in WT as in [354] and we showed that CREB activity is still oscillating in the *Bmal1*<sup>-/-</sup> context. On average, we observed a phase delay of approximately two hours in *Bmal1*<sup>-/-</sup>, though this was not significant ( $p=0.5$ , Chow test), presumably owing to individual feeding patterns [354]. Our results confirm that CREB is regulated by the food related signaling in clock impaired mouse upon night restricted feeding regimen. The phase delay of two hours suggests that the circadian clock is implicated in the fine tuning of the hepatic glucose metabolism.

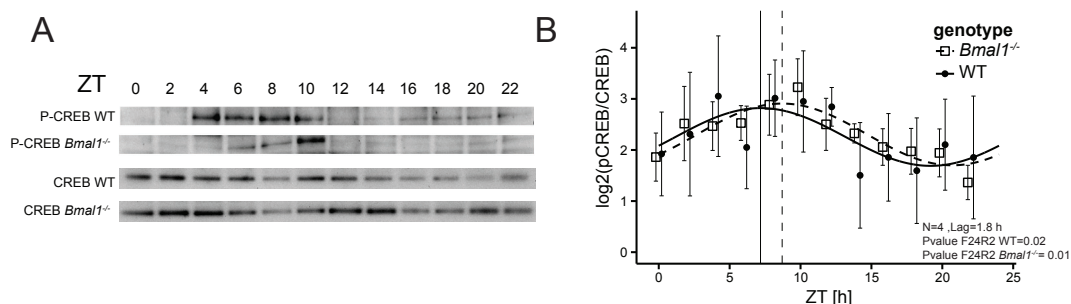


Figure 2.26 – CREB validation by western blot in WT and in *Bmal1*<sup>-/-</sup>. **A**) Representative Western blot of CREB and pCREB in WT and *Bmal1*<sup>-/-</sup> genotypes. **B**) Quantification of log<sub>2</sub>(pCREB/CREB) Western blots in WT and *Bmal1*<sup>-/-</sup> genotypes. Nuclear extracts from four independent livers (n=4) were harvested every two hours. Both genotypes showed a significant oscillation ( $p<0.05$ , Harmonic regression) of the mean signal from the four mice. Though the peak time in *Bmal1*<sup>-/-</sup> is delayed, the comparison of the rhythm in the two conditions is not significant ( $p=0.49$ , Chow test).

### 2.3.3 Oscillating gene expression and related pathway in WT and *Bmal1*<sup>-/-</sup> mouse liver

In order to understand how the circadian clock and the feeding-fasting cycle control diurnal gene expression in liver, we studied differences in mRNA expression and between *Bmal1*<sup>-/-</sup> and wild-type mice subject to the same night restricted feeding regimen. We looked at oscillating genes in WT and in *Bmal1*<sup>-/-</sup> ( $p$ -value < 0.05, log<sub>2</sub> amplitude > 0.5, FDR < 0.3) and we observed that a substantial fraction of genes is still oscillating in the *Bmal1*<sup>-/-</sup> genotype (figure 2.27 A). In addition the overlap of genes cycling in WT and *Bmal1*<sup>-/-</sup> was of intermediate size indicating that genes with a diurnal expression pattern are quite different between WT and *Bmal1*<sup>-/-</sup> mice. By computing the cumulative count from the genes with the highest amplitude to the lowest (figure 2.27 B), we observed that genes in WT had bigger amplitudes than in *Bmal1*<sup>-/-</sup> as expected. The phase distribution of mRNA in WT and in *Bmal1*<sup>-/-</sup> suggest a phase delay of gene expression in the absence of the clock (figure 2.27 C).

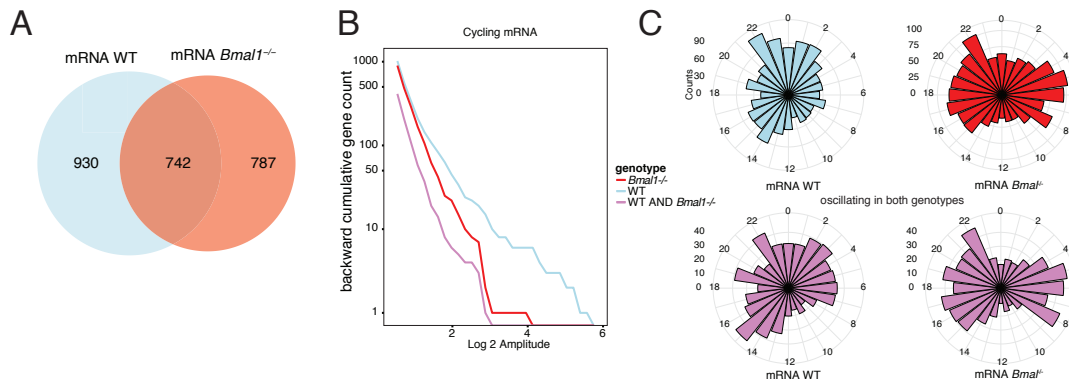


Figure 2.27 – Rhythmicity analysis of mRNA from microarray in WT and *Bmal1*<sup>-/-</sup>. **A)** Venn diagram of oscillating genes in WT and in *Bmal1*<sup>-/-</sup> using mRNA expression. **B)** Cumulative count of oscillating genes (with a multi linear regression  $p$ -value below 0.05 and a log<sub>2</sub> amplitude greater than 0.5) in *Bmal1*<sup>-/-</sup> and WT with log<sub>2</sub> amplitude greater or equal to x. **C)** Phase histograms of oscillating genes in WT, in *Bmal1*<sup>-/-</sup> and in both genotypes.

Subsequently, we wanted to identify the function of genes oscillating in WT, oscillating in *Bmal1*<sup>-/-</sup> and oscillating in both conditions. Therefore we did a functional pathway analysis with g:Profiler [269] using Kegg and Reactome pathway annotations (figure 2.28, table 2.4). In the WT context, genes annotated for circadian rhythm as well as lipid and sugar metabolism related pathway annotations were enriched. In the *Bmal1*<sup>-/-</sup>, we observed that pathway related to sugar, and lipid metabolism are still oscillating. We found that amplitudes of genes related to the glucose-dependent ChREBP signaling were increased in the absence of the clock and the amplitudes of genes annotated with bile acids secretion are decreased in the *Bmal1*<sup>-/-</sup> genotype. ChREBP has emerged as a central regulator of glycolysis and *de novo* fatty acid in liver [148]. Taken together, these changes proved once more that the clock is implicated in the precise regulation of the liver metabolism.

### 2.3. Phenotypic consequences of whole body *Bmal1* knockout

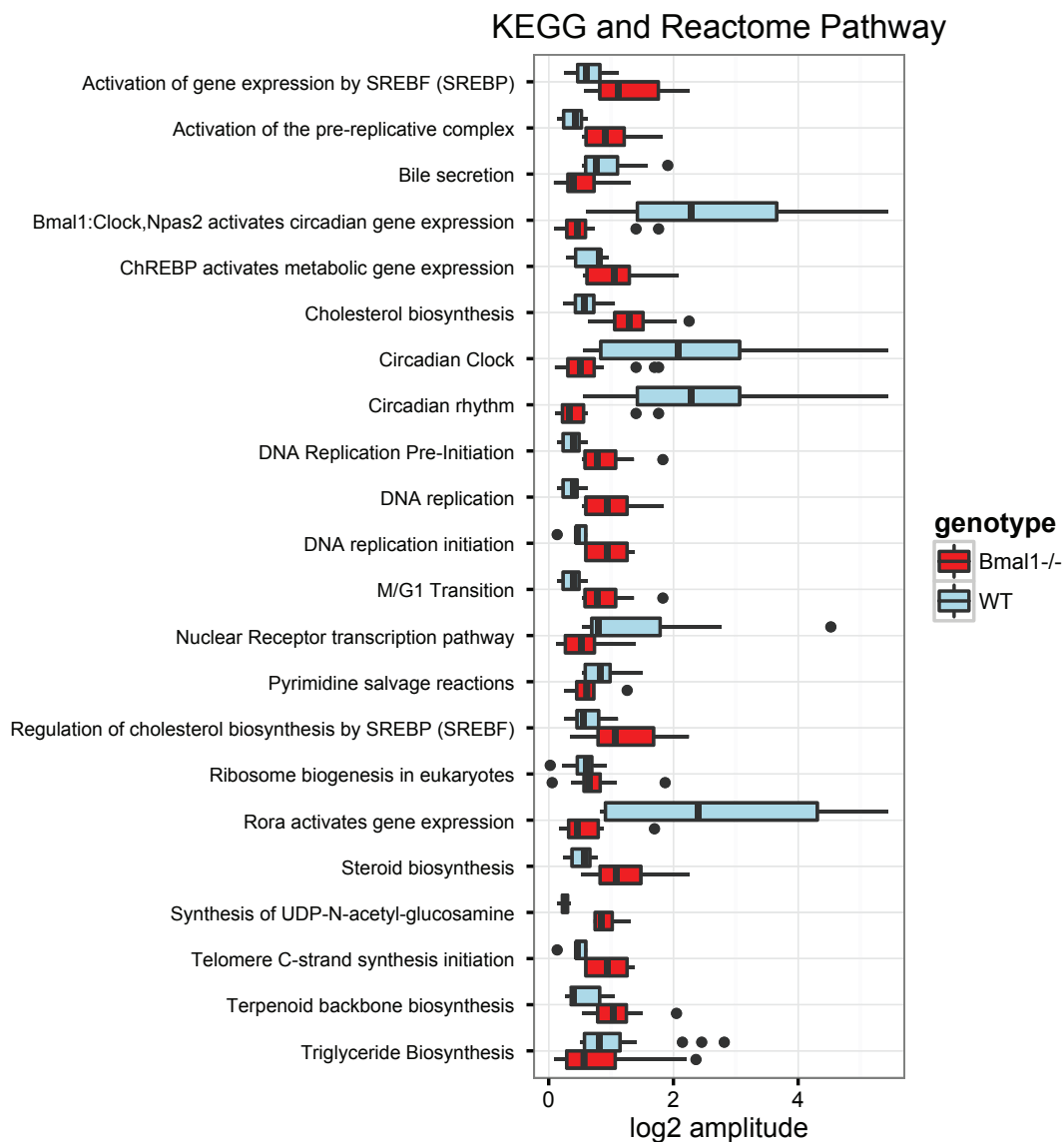


Figure 2.28 – Kegg and Reactome Pathway enrichment and amplitudes analysis. Boxplot of log<sub>2</sub> mRNA amplitude of genes from significantly enriched Kegg or Reactome pathway in WT and in *Bmal1*<sup>-/-</sup> genotypes. These pathways were retrieved using g:Profiler with a *p*-value threshold below 0.1 in one or both genotype. Boxplots were generated with the amplitude of each oscillating genes annotated with a specific pathway. Genes used for each annotation are reported in the table 2.4.

## Chapter 2. Results

Table 2.4 – Reactome and Kegg pathway analysis of oscillating genes in WT and *Bmal1*<sup>-/-</sup> mouse liver

<i>genotype</i>	<i>p-value</i>	<i>Source ID</i>	<i>Pathway</i>	<i>Genes</i>
WT	7.84E-8	REAC:508751	Circadian Clock	<i>CRY1, SREBF1, NAMPT, NR1D1, PER1, FBXL3, PPARA, CPT1A, NPAS2, CLOCK, BHLHE40, BHLHE41, ELOVL3, ARNTL, PER2, DBP, CRY2</i>
WT	4.92E-7	REAC:1368110	Bmal1:Clock,Npas2 activates circadian gene expression	<i>CRY1, NAMPT, NR1D1, PER1, PPARA, NPAS2, CLOCK, BHLHE40, BHLHE41, ARNTL, PER2, DBP, CRY2</i>
WT	3.69E-5	KEGG:04710	Circadian rhythm	<i>CRY1, NR1D1, PER1, FBXL3, NPAS2, RORC, PER3, CLOCK, BHLHE40, BHLHE41, ARNTL, PER2, CRY2</i>
WT	3.71E-4	REAC:7083403	Triglyceride Biosynthesis	<i>ELOVL1, LPIN1, ACLY, ACOT3, ACOT12, LPIN2, FASN, AGPAT2, ACOT7, AGK, AGPAT6, ELOVL5, AGPAT1, ELOVL3, ELOVL6, GPD1L, ACOT4, LCLAT1, ACOT1</i>
WT	6.92E-4	KEGG:04976	Bile secretion	<i>ADCY6, SLC10A2, ABCG8, ATP1B1, ABCB11, CAR2, NCEH1, CYP7A1, SLC2A1, SLC01A4, AQP8, LDLR, AQP9, ATP1A1, NR0B2, ABCG5, ABCB1A, NR1H4, SLC4A4</i>
WT	7.61E-4	REAC:7084372	Regulation of cholesterol biosynthesis by SREBP (SREBF)	<i>KPNB1, MVD, SREBF1, SQLE, SREBF2, FASN, PMVK, SCAP, LSS, ELOVL6, MVK, ACACB, DHCR7</i>
WT	1.14E-3	REAC:7084024	Nuclear Receptor transcription pathway	<i>PPARD, NR1I3, NR2C2, RXRG, ESRI, NR1D1, NR1D2, PPARA, NR4A1, RORC, NR3C2, NR0B2, AR, NR1H4, THRA, NR2C2AP</i>
WT	3.34E-3	REAC:7084371	Activation of gene expression by SREBF (SREBP)	<i>MVD, SREBF1, SQLE, SREBF2, FASN, PMVK, LSS, ELOVL6, MVK, ACACB, DHCR7</i>
WT	1.95E-2	REAC:7083346	Pyrimidine salvage reactions	<i>UPP1, TYMP, TK1, UCK2, UPP2, CDA</i>
WT	3.07E-2	KEGG:03008	Ribosome biogenesis in eukaryotes	<i>NOB1, GTPBP4, RIKO2, RCL1, NOP58, NAT10, NOP56, SPATA5, POP4, MPHOSPH10, WDR43, CIRH1A, GNL3, RPP38, HEATR1, UTP18, UTP14B</i>
WT	3.98E-2	REAC:1368092	Rora activates gene expression	<i>SREBF1, NR1D1, CPT1A, NPAS2, CLOCK, ARNTL</i>
WT	4.68E-2	REAC:7083820	Cholesterol biosynthesis	<i>MVD, FDFT1, SQLE, PMVK, NSDHL, SCAMOL, LSS, MVK, DHCR7</i>
<i>Bmal1</i> <sup>-/-</sup>	1.89E-13	REAC:7084372	Regulation of cholesterol biosynthesis by SREBP (SREBF)	<i>KPNB1, CYP51, MVD, NFYB, ACACA, SREBF1, SQLE, SREBF2, NFYA, GPAM, FASN, SCD2, PMVK, RAN, SC5D, LSS, ELOVL6, MVK, ACACB, DHCR7, FDPS</i>
<i>Bmal1</i> <sup>-/-</sup>	3.52E-13	REAC:7084371	Activation of gene expression by SREBF (SREBP)	<i>CYP51, MVD, NFYB, ACACA, SREBF1, SQLE, SREBF2, NFYA, GPAM, FASN, SCD2, PMVK, SC5D, LSS, ELOVL6, MVK, ACACB, DHCR7, FDPS</i>
<i>Bmal1</i> <sup>-/-</sup>	9.31E-8	REAC:7083820	Cholesterol biosynthesis	<i>CYP51, MVD, FDFT1, HMGC, SQLE, HSD17B7, PMVK, NSDHL, SCAMOL, SC5D, LSS, MVK, DHCR7, FDPS</i>
<i>Bmal1</i> <sup>-/-</sup>	1.15E-6	KEGG:00100	Steroid biosynthesis	<i>CYP51, FDFT1, SQLE, SOAT2, HSD17B7, CYP2R1, NSDHL, SCAMOL, SC5D, LSS, DHCR7</i>
<i>Bmal1</i> <sup>-/-</sup>	1.22E-6	REAC:7083228	Activation of the pre-replicative complex	<i>MCM2, MCM5, CDT1, POLA1, POLE, RPA3, MCM4, POLA2, PRIM1, PRIM2, MCM6, MCM8, MCM7, MCM3</i>
<i>Bmal1</i> <sup>-/-</sup>	3.70E-6	REAC:7083665	ChREBP activates metabolic gene expression	<i>MLXIPL, MLX, ACACA, ACLY, FASN, AGPAT1, ACACB</i>
<i>Bmal1</i> <sup>-/-</sup>	6.51E-5	KEGG:03008	Ribosome biogenesis in eukaryotes	<i>NOB1, XPO1, GTPBP4, RCL1, GNL3L, SBDS, WDR75, NOP58, NOP56, SPATA5, RAN, MPHOSPH10, WDR43, CIRH1A, GNL3, CSNK2A2, RPP38, HEATR1, MDN1, UTP14B</i>
<i>Bmal1</i> <sup>-/-</sup>	6.71E-5	KEGG:03030	DNA replication	<i>MCM2, MCM5, POLA1, POLE, RPA3, POLD2, MCM4, POLA2, PRIM1, PRIM2, MCM6, MCM7, MCM3</i>
<i>Bmal1</i> <sup>-/-</sup>	3.72E-3	REAC:7083225	DNA replication initiation	<i>POLA1, POLE, POLA2, PRIM1, PRIM2</i>
<i>Bmal1</i> <sup>-/-</sup>	3.72E-3	REAC:7083769	Telomere C-strand synthesis initiation	<i>POLA1, POLE, POLA2, PRIM1, PRIM2</i>
<i>Bmal1</i> <sup>-/-</sup>	8.09E-3	KEGG:00900	Terpenoid backbone biosynthesis	<i>MVD, HMGC, NUS1, ACAT2, PDSS1, PMVK, MVK, FDPS</i>
<i>Bmal1</i> <sup>-/-</sup>	3.88E-2	REAC:7083217	M/G1 Transition	<i>MCM2, MCM5, CDT1, POLA1, POLE, UBC, RPA3, E2F2, MCM4, POLA2, PRIM1, PRIM2, MCM6, PSMD14, MCM8, MCM7, MCM3</i>
<i>Bmal1</i> <sup>-/-</sup>	3.88E-2	REAC:7083216	DNA Replication Pre-Initiation	<i>MCM2, MCM5, CDT1, POLA1, POLE, UBC, RPA3, E2F2, MCM4, POLA2, PRIM1, PRIM2, MCM6, PSMD14, MCM8, MCM7, MCM3</i>
<i>Bmal1</i> <sup>-/-</sup>	4.15E-2	REAC:7084173	Synthesis of UDP-N-acetylglucosamine	<i>UAP1, GFPT1, GNPAT1, PGM3</i>
WT and <i>Bmal1</i> <sup>-/-</sup>	2.33E-7	REAC:7084372	Regulation of cholesterol biosynthesis by SREBP (SREBF)	<i>KPNB1, MVD, SREBF1, SQLE, SREBF2, FASN, PMVK, LSS, ELOVL6, MVK, ACACB, DHCR7</i>
WT and <i>Bmal1</i> <sup>-/-</sup>	3.13E-7	REAC:7084371	Activation of gene expression by SREBF (SREBP)	<i>MVD, SREBF1, SQLE, SREBF2, FASN, PMVK, LSS, ELOVL6, MVK, ACACB, DHCR7</i>
WT and <i>Bmal1</i> <sup>-/-</sup>	2.62E-5	REAC:7083820	Cholesterol biosynthesis	<i>MVD, FDFT1, SQLE, PMVK, NSDHL, SCAMOL, LSS, MVK, DHCR7</i>
WT and <i>Bmal1</i> <sup>-/-</sup>	2.84E-4	KEGG:00100	Steroid biosynthesis	<i>FDFT1, SQLE, SOAT2, NSDHL, SCAMOL, LSS, DHCR7</i>
WT and <i>Bmal1</i> <sup>-/-</sup>	6.35E-4	KEGG:03008	Ribosome biogenesis in eukaryotes	<i>NOB1, GTPBP4, RCL1, NOP58, NOP56, SPATA5, MPHOSPH10, WDR43, CIRH1A, GNL3, RPP38, HEATR1, UTP14B</i>
WT and <i>Bmal1</i> <sup>-/-</sup>	1.35E-2	REAC:7083665	ChREBP activates metabolic gene expression	<i>ACLY, FASN, AGPAT1, ACACB</i>

## 2.3. Phenotypic consequences of whole body *Bmal1* knockout

---

### 2.3.4 Glucose, lipids and fatty acids metabolism in WT and *Bmal1*<sup>-/-</sup>

During fasting, glucose homeostasis is achieved by triggering expression of gluconeogenic genes in response to glucagon and glucocorticoids. Glucagon exerts its strongest action when mice wake up and begin to eat. The response to glucagon by liver cells varied depending on a direct circadian regulation of the transcriptional activity of cAMP response element-binding protein (CREB), which was inhibited by CRYs [354]. CREB was found to induce expression of the gluconeogenic program through the nuclear receptor coactivator PGC-1 [137].

As demonstrated in a previous section (see 2.3.2) CREB is still rhythmic in *Bmal1*<sup>-/-</sup> genotype upon night restricted feeding. In addition, our analysis (section 2.3.3) showed that ChREBP and SREBP signaling were affected by the *Bmal1*<sup>-/-</sup>. These two transcription factors are implicated in glucose metabolism and lipid synthesis in liver [75].

#### Glucose metabolism in WT and *Bmal1*<sup>-/-</sup>

As developed in the introduction (section 1.5.2), the glucose homeostasis is controlled by the circadian clock. Thus, during the feeding period blood glucose is mainly of dietary origin, while during the starvation period glucose is produced in the liver through gluconeogenesis.

Therefore, we wanted to study gene related the glucose metabolism at Pol II and mRNA levels. We looked (figure 2.29 and gluconeogenesis pathways **A**) at the gluconeogenesis and at the glycolysis pathways and we labeled several irreversible steps at physiological conditions that are specific to one or the other pathway. Indeed the conversion of the oxaloacetate to phosphoenolpyruvate by the phosphoenolpyruvate carboxykinase (PCK1) is one key regulatory step of the gluconeogenesis. Furthermore, the conversion of fructose 1,6-bisphosphate to  $\beta$ -D-fructofuranose 6-phosphate by the enzyme fructose-1,6-bisphosphatase (FBP1,FBP2) and the conversion of D-glucopyranose 6-phosphate to D-glucopyranose by glucose-6-phosphatase (G6pc) are located at the end of the pathway. These two reactions are critical steps of the gluconeogenesis pathway. Interestingly, in the glycolysis the inverse conversion steps are regulated by different enzymes with an irreversible action notably the  $\beta$ -D-glucopyranose 6-kinase (Gck) and the 6-phosphofructokinase (Pfkf). The conversion of phosphoenolpyruvate to pyruvate by the pyruvate kinase (Pklr) is the last irreversible reaction of the glycolysis. These key regulatory steps are essential to allow the temporal separation of glycolysis and gluconeogenesis in the liver. Several enzymes of these two pathways were detected with a circadian behavior in Pol II loading at the TSS and mRNA accumulation in WT and in *Bmal1*<sup>-/-</sup> (figure 2.29 **B**). Our data suggest that these enzymes are co-regulated by the circadian clock and by feeding cues at the transcriptional level, as the oscillations were maintained in *Bmal1*<sup>-/-</sup> context but with different amplitudes and phases compared to the WT genotype. More precisely, these genes depict lower amplitudes in *Bmal1*<sup>-/-</sup> (figure 2.30), notably *Gck* and *Pck*. In addition the phases of these genes are quite impacted by the impaired clock genotype, while the mesor is not affected.



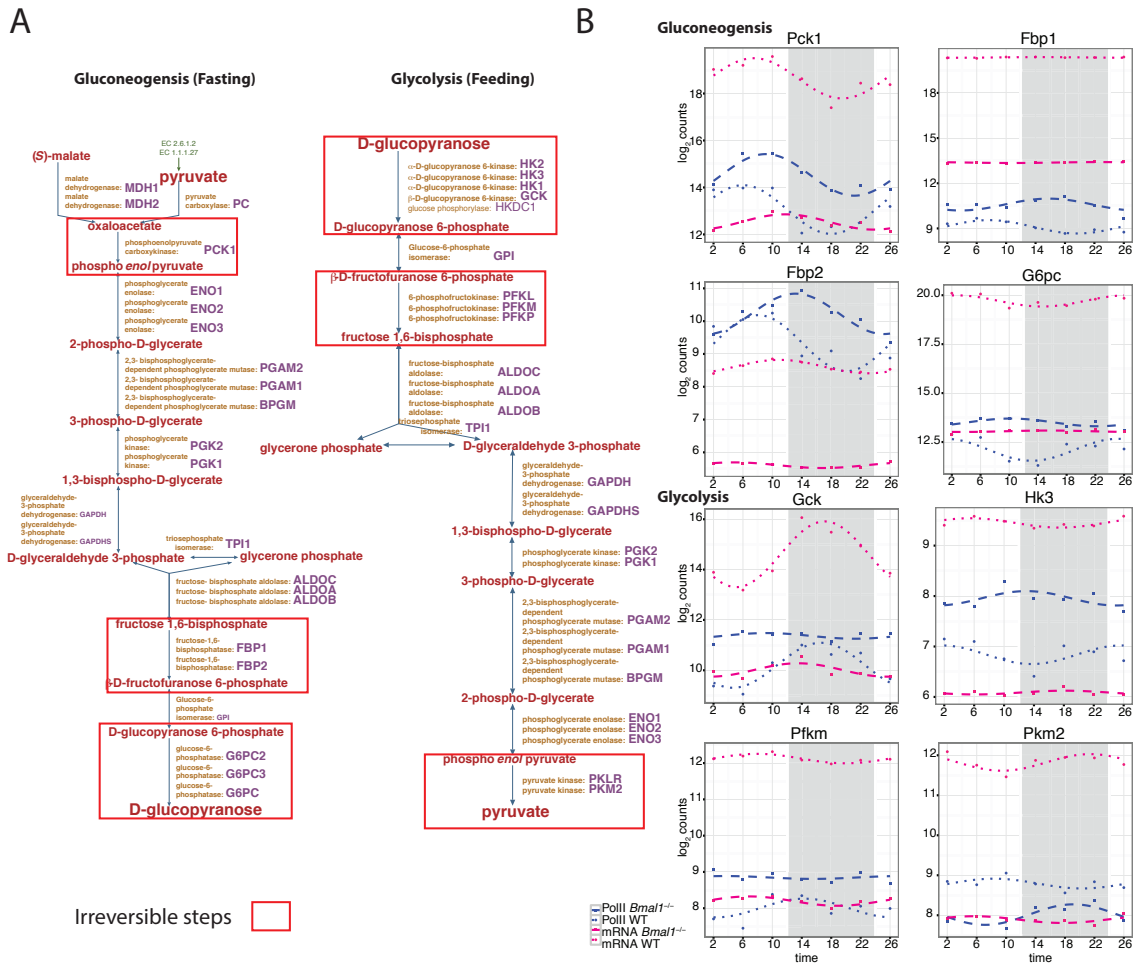


Figure 2.29 – Glycolysis and gluconeogenesis pathways. **A)** Gluconeogenesis is a mechanism which maintains blood glucose levels, avoiding low levels (hypoglycemia) during fasting phase. It is a pathway consisting of a series of eleven enzyme-catalyzed reactions. Many of the reactions are reversible steps, and irreversible steps are important regulatory steps highlighted using red boxes. Glycolysis converts glucose to pyruvate by via a series of intermediate metabolites. Each chemical modification (arrows) is performed by a different enzyme encoded by different genes (purple). **B)** Temporal profiles in mRNA (in pink) and Pol II loadings at TSS (in blue) of key regulators of the glycolysis/gluconeogenesis in WT (dotted line) and in *Bmal1*<sup>-/-</sup> (dashed line) context.



### 2.3. Phenotypic consequences of whole body *Bmal1* knockout

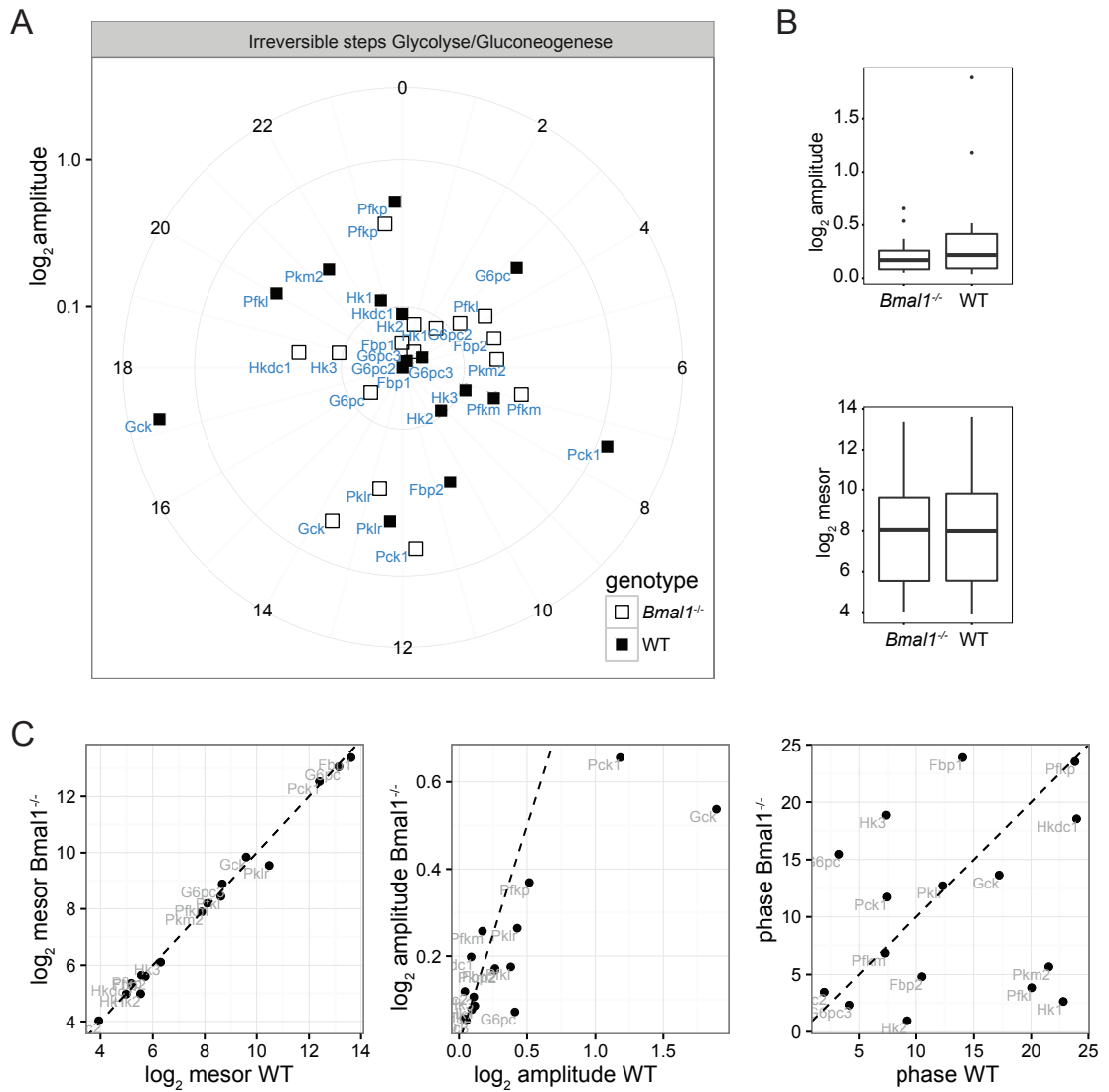


Figure 2.30 – Effect of *Bmal1* knockout on key steps of glycolysis/gluconeogenesis. **A)** Phase and amplitude of key steps of the glycolysis/gluconeogenesis in WT and in *Bmal1*<sup>-/-</sup> reported in a polar scatter plot. **B)** boxplot of log<sub>2</sub> amplitudes and mesor in WT and in *Bmal1*<sup>-/-</sup>. **C)** Mesor, amplitudes and phase in WT versus *Bmal1*<sup>-/-</sup>.

## Chapter 2. Results

---

In order to explain changes in phase and amplitude of *Gck* and *G6pc* we looked at the promoter region of these two enzymes, which play an opposite role during glycolysis/gluconeogenesis (figure 2.31). We did a PWM scan with FIMO and Wang PWMs [338, 114] and we overlapped DNase I-seq signal which was analyzed with Wellington algorithm to detect TFBS footprints [256]. We observed that *Gck* contained several E-box motifs (MYC/MAX) that are under an important footprint about 100bp upstream the TSS. Moreover, in the same DHS we found a motif of CEBPB, which was linked already with the regulation of gluconeogenesis [13] as an essential factor underlying glucocorticoid-dependent activation of *Pepck* gene transcription. In the DHS 1 Kbp upstream of *Gck* promoter, we found motifs of CREB, GR (NR3C1) and HSF1, which were TFBS motifs with a high-activity in the *Bmal1*<sup>-/-</sup> when we did our penalized linear model. These motifs might be responsible for the remaining oscillations in *Gck* transcription and mRNA accumulation. The analysis of *G6pc* promoter revealed a CREB and a TCF7L2 motif a 100bp upstream of the TSS. In addition, TCF7L2 has been recently shown to inhibit adjacent promoter occupancies of CREB, CRT2, and FOXO1, which are critical transcriptional modules in hepatic gluconeogenesis [241]. Taken together, our motif predictions may explain the transcriptional regulation of *Gck* and *G6pc*. However, this analysis was centered on the promoter of these two genes, which are regulated as well by distal cis-regulatory elements and the phase of expression of these genes is quite difficult to explain only by looking at the TFBS in the promoter. As demonstrated before (section 2.3.1), distal sites might add useful information to the fine tuning of the gene expression. In addition some transcription factors depict a weak footprint that might be difficult to detect, such as nuclear receptors. Therefore, some of the motifs inside DHS without footprints might play a role in the transcriptional regulation. Moreover in glycolysis/gluconeogenesis the protein expression and activation, such as phosphorylation of F6P during glycolysis for example, are key regulatory events of these pathway. Therefore the results obtained at the transcriptional level are not as accurate as at the protein activity level, although the regulation of transcription seems to play a role. In conclusion, we confirmed that the glucose metabolism is fine tuned by the circadian clock and that this regulation is performed at least partially at a transcriptional level.

### 2.3. Phenotypic consequences of whole body Bmal1 knockout

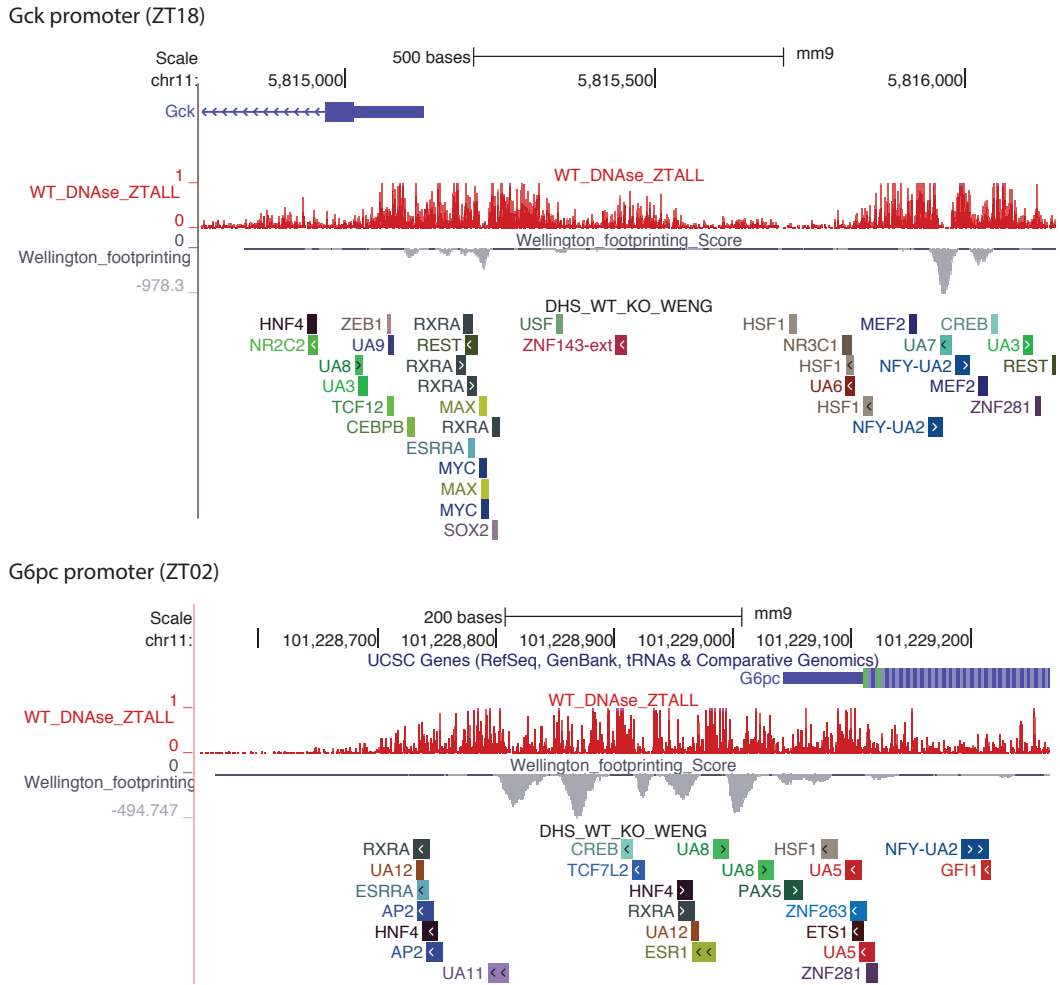


Figure 2.31 – *Gck* and *G6pc* promoter: example of transcriptional control of glucose metabolism. Promoter structure of **A)** *Gck* and **B)** *G6pc*, two key enzymes of gluconeogenesis/glycolysis. UCSC genome browser screenshot of the promoter region overlapped with our PWM scan using Wang [338] motifs and with DNase I signal analyzed with Wellington algorithm for digital genomic footprint detection [256].

## Chapter 2. Results

### lipids and fatty acids metabolism in WT and *Bmal1*<sup>-/-</sup>

As lipid metabolism is affected by the circadian clock and the feeding-fasting cycle [111] and as ketone bodies are produced by the liver through fatty acid oxidation during fasting conditions [238], we wanted to study free fatty acids (FFA),  $\beta$ -hydroxybutyrate (BHB) and triglycerides (TG) temporal pattern in blood of WT and *Bmal1*<sup>-/-</sup> mice.

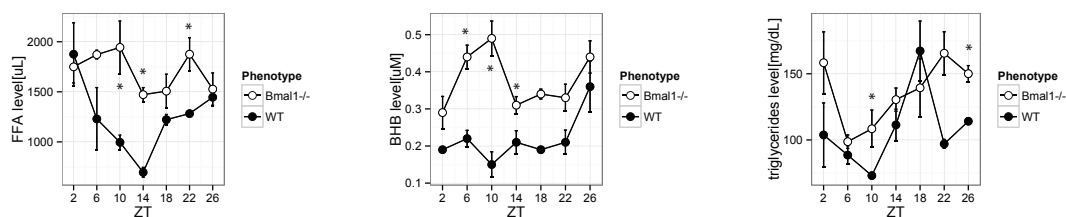


Figure 2.32 – Free Fatty acids, triglycerides and ketone bodies in WT and in *Bmal1*<sup>-/-</sup>. FFA TG and BHB were monitored in blood of n = 3 - 5 mouse per time point at ZT 2,6,10,14,18,22 and 26 in wild-type and *Bmal1*<sup>-/-</sup> by Dr. Federica Gilardi at UNIL. ZT0 represent the switch between dark and light phase.

In wild-type condition FFA level in blood decreases during the light phase (ZT0-ZT12), where the mice are fasting and sleeping, and increase during the dark phase when the mice are awake and eating. The *Bmal1*<sup>-/-</sup> condition exhibits significantly higher level of FFA (ZT10, ZT14 and ZT22) throughout the day. BHB levels are significantly higher in *Bmal1*<sup>-/-</sup> during the light phase at ZT6 and ZT10 and the beginning of the dark phase at ZT14. The wild-type temporal pattern of triglycerides show a low concentration during the light phase and a peak during the dark phase at ZT18 with a rapid decrease at ZT22. Interestingly, in the *Bmal1*<sup>-/-</sup> context, the level of triglycerides is slightly more elevated during the day, especially at ZT 10. During the night phase the *Bmal1*<sup>-/-</sup> mice didn't succeed to reduce the concentration of triglyceride notably at ZT 22 and in the beginning of the light phase at ZT 26. This effect might be due to a phase shift of the *de novo* lipogenesis in the *Bmal1*<sup>-/-</sup>.

Elevated levels of FFA are usually associated with obesity and insulin resistance. Higher BHB levels in *Bmal1*<sup>-/-</sup> suggest increased FFA oxidation and diabetic ketoacidosis, leading to reduced mitochondrial redox state [329]. Interestingly, a recent study showed that liver-derived ketone bodies such as BHB are necessary for food anticipation. They have shown that *Per2* regulates  $\beta$ -hydroxybutyrate (BHB) and that food anticipation originates in the liver [56]. Finally, the level of TG suggests a potential shift of 2h in *de-novo* lipogenesis in the *Bmal1*<sup>-/-</sup> genotype. Potential relationships with diabetes and non-alcoholic fatty liver disease are suggested by the literature [232, 289, 214], but in our night restricted feeding regimen the *Bmal1*<sup>-/-</sup> did not exhibit liver steatosis, as reported by Dr. Federica Gilardi. In conclusion, the night restricted feeding regimen seems to counteract deleterious effect induced by a defect of the molecular clock.

## 3 Discussion

### 3.1 Abundant diurnal DHSs in the mouse liver

This work is to our knowledge the first study to map genome-wide DNase I hypersensitivity (DHS) in the mouse liver as well as to profile DHS sites (DHSs) around the clock with 4h resolution. The resulting data set provides an unprecedented view of circadian chromatin accessibility. Other studies [334, 171, 65] profiled the enrichment of the active chromatin mark H3K27ac for a similar purpose. However, using new and previously published H3K27ac enrichment data, we found that the genomic resolution and extent of regulatory insight offered by DHS profiling is superior to that of H3K27ac mapping [237, 320, 311] because DHS enables higher resolution in mapping transcription factor binding sites and allows dynamic footprinting analyses. Nevertheless, these molecular signals are complementary, and display a high correlation with each other and Pol II genome-wide and across circadian time. In contrast to the observed delay between H3K4me3 enrichment and Pol II, reported previously [188], no significant delays are observed between accessibility as measured by DHS and H3K27ac enrichment at a given site, which likely reflects the fact that histone acetylation turnover is faster than that of histone methylation [213].

Comparing average signals, we observed that transcription start sites (TSSs) displayed higher intensity for all marks studied compared to distal or proximal DHSs; however, relative amplitudes in oscillations appeared higher in non-TSS sites. This suggests that TSS accessibility is less variable than putative enhancer accessibility, which likely reflects increased protein binding dynamics at putative enhancers compared to TSSs. This suggests that enhancers may have a dominant role in the control of diurnal gene expression, consistent with the notion that modulation of histone marks and accessibility of chromatin at enhancers is one of the primary regulatory mechanisms underlying cell type specificity [134]. This hypothesis is further supported by our observation of close phase relationships between putative enhancers and the nearest TSS. As such, our data constitute a valuable resource to explore the involvement of putative enhancer regions in mediating cyclic recruitment of Pol II at the TSSs and subsequent transcription of the respective target genes.

The core clock factor BMAL1 is essential for rhythmic transcription in mammals. Liver-specific and global *Bmal1* knockout in mice lead to important effects on metabolism and behavior that can be traced to dysregulation of BMAL1 target genes [300, 286, 251]. Importantly, comparing clock wild-type with BMAL1 mutant suggested that feeding, rather than the clock, is the main driver of rhythmic translation in the liver [17]. It has been suggested that BMAL1 is a pioneer-like factor enabling easier access of regulatory sites by other TFs [222]. Its dimerization partner CLOCK has been shown to have histone acetyltransferase activity [80]. Therefore, we sought to verify whether its presence in the mouse liver led to large-scale differences in DNA accessibility as measured by DNase I digestion.

### 3.2 Limited impact of BMAL1 removal onto the liver chromatin landscape

Comparing DHS signals between wild-type and *Bmal1*<sup>-/-</sup> samples at ZT6 revealed relatively few changes as the signal correlation between the two samples was above 0.8. Thus, only a minority of DHSs appear to be affected, consistent with the limited number of target sites for BMAL1 [171, 273]. However, the majority of these BMAL1 binding sites showed a decrease in DHS signal, which may be consistent with a pioneer-like function for BMAL1 [222]. Higher resolution, footprint- and protein structure-based analyses further revealed a CLOCK:BMAL1 DNA binding pattern at tandem E-box sites consistent with a hetero-tetramer configuration at peak activity of these TFs [273, 249, 234]. Interestingly, at the daily minimum of BMAL1 activity, only the strongest E-boxes still display a footprint, further supporting the notion that other TFs, such as USF1, compete with CLOCK:BMAL1 to occupy similar target sites, although in distinct biophysical fashion [301, 76].

In order to compare the binding pattern of BMAL1:CLOCK with another important TF in the liver we looked at the rhythm of the CREB footprint. CREB binds to cAMP responsive elements (CRE motif) in a strong/stable manner, which explains why we do not see any change in the DNase I pattern during the diurnal cycle [211]. Our data support the hypothesis that CREB is constitutively bound to its cognate DNA binding site and that its circadian activation is mediated through other mechanisms, such as phosphorylation in a circadian manner in the liver by PKA, PKC or CK2 [93, 107].

Studying other transcription factor binding sites such as HSF, SREBP and RRE, we found that the number of bound targets changed diurnally, but that the footprints remained constant. Heat shock transcription factor 1 (HSF1), which drives the expression of heat-shock proteins at the onset of the dark phase [270, 291], had an apparent enrichment of bound sites between ZT14 and ZT22. The number of targets of HSF1 was low, suggesting that this regulator is highly specific for the temperature-stress-related pathway [270]. REVERB $\alpha$  participates in the circadian modulation of sterol regulatory element-binding protein (SREBP) activity, and thereby in the daily expression of SREBP target genes involved in cholesterol and lipid metabolism [189]. The SREB motif showed a complex circadian pattern with two modes

at ZT10 and ZT22. However, the temporal mRNA expression of SREBP1 targets was not always synchronized with its binding pattern. In particular, different expression phases were described for SREBP1 target genes depending on their function, indicating the implication of other transcription factors in their regulation [111]. The DNase I footprint at RRE (ROR responsive element) showed two modes of activity with a small peak at ZT10 (binding of the repressor RevErb) and a higher peak at ZT22 (binding of the activator ROR). This bimodal activity suggests a possible dynamic exchange of TFs, namely RORs and REVERBs, on their cognate binding sites [120].

### 3.3 Tissue specific regulation of the circadian clock

The liver tissue, composed mainly of hepatocytes, is subject to systemic cues such as feeding/-fasting cycle, body temperature variations and hormonal cues, while circadian oscillation in NIH3T3 fibroblasts (cultured in a dish) originate only from the endogenous molecular clock. In our comparative study of liver tissue and NIH3T3 fibroblasts, we discovered and described the varying motif content and chromatin landscape found around BMAL1 binding sites. We observed that BMAL1 sites are accessible, active and in turn repressed in a context-specific manner. This observation raised the question of how tissue specific regulation of the circadian clock is achieved.

The result of our analysis of BMAL1 sites located in accessible regions (DHS), supports the hypothesis that the chromatin state determines context specific binding. Our motif analysis indicated that co-regulatory elements in NIH3T3 fibroblasts seemed to be more specifically linked to BMAL1 sites rather than being overall cell-type specific factors. For instance, members of the E2F family appear to colocalize strongly with BMAL1. E2F factors are notably involved in cell cycle regulation [271], which is consistent with fibroblasts still actively cycling, as opposed to most of the cells in the liver. Another interesting factor enriched in fibroblast is NRF1, which targets key enzymes in oxidative metabolism and mitochondrial function [293].

On the other hand, in liver we found previously reported tissue-specific regulators potentially driving the context-specific expression of clock gene. Among the identified factors we found FOXA, a well-known pioneer factor [61, 117, 206], HNF4 and several nuclear receptors (as reported in [349]). Strikingly, in liver, one of the most enriched factors with a high percentage of bound motifs was GATA1. A recent publication [220] describing the tissue-specific binding of CLOCK and CYCLE in *Drosophila* identified serpent (SRP), which is a member of the GATA family in flies, as a cis-regulatory factor necessary for driving CLOCK-CYCLE mediated transcriptional activation in *Drosophila* body. The majority of its targets is linked to metabolism and detoxification and is active in the fat body, midgut and Malpighian tubules. Thus, there might be an evolutionary link to the enrichment of GATA1 in liver. In addition, several other studies [38, 39, 341] have reported a list of motifs overrepresented in promoter regions of clock-controlled genes, many of which were detected in our analysis. While this gives confidence in our motif finding, we further extended their analysis by appointing context-specificity



and distal regulatory elements to the motifs discovered. In conclusion, it will be appealing to study further the networks containing the circadian clock machinery as well as putative co-regulatory factors, including distal regulatory elements, in their respective cell type and possibly considering an evolutionary perspective.

### 3.4 Regression models uncover factors which maintain cycling patterns in gene expression in the absence of BMAL1 activity

Our genome-wide study of the oscillating genes using Pol II loadings at TSSs or mRNA expression in WT and *Bmal1*<sup>-/-</sup> genotype revealed that a similar proportion of genes ( about 10% of genes expressed in the liver) depict a diurnal pattern in both contexts, but the overlap is surprisingly small ( 20%). As expected, we found that amplitudes are higher in WT than in *Bmal1*<sup>-/-</sup> genotype, but the number of oscillating genes is slightly more important in *Bmal1*<sup>-/-</sup>. These observations were consistent between mRNA and Pol II loadings. Thereby we observed that the molecular clock has a dual effect on gene regulation probably through BMAL1 activation and PER1/2 repression [85] and that restricted feeding can restore circadian expression of hundred of transcripts [333]. Using a penalized regression model, we further uncovered an important role for the TFs FOXO1, Glucocorticoid Receptor (GR), CREB and NFY in mediating robust cyclic transcription in mice lacking BMAL1. GR activity seemed to be unaffected by the *Bmal1*<sup>-/-</sup> genotype, whereas FOX and NFY showed increased activity, with a similar phase in WT and in *Bmal1*<sup>-/-</sup>. Interestingly, NFY is a binding partner of SREB and Sp1 in the regulation of lipid metabolism and cholesterol biosynthesis [268]. Therefore, our data suggest that lipid metabolism and cholesterol metabolism do not require a functional clock under night restricted feeding, as previously described [333]. Moreover, restricted feeding has recently been shown to counteract the effect of a high-fat diet in WT mice [130, 55]. In the case of CREB, we found an important phase delay in *Bmal1*<sup>-/-</sup>, implying that the circadian clock regulates hepatic gluconeogenesis [354]. Forkhead box factors like FOXO1 have been implicated in cell cycle regulation, oxidative stress, and are negatively regulated by insulin signaling [187]. Analogously to the core clock, FOXO1 and FOXO6 also regulate the expression of key enzymes implicated in gluconeogenesis [166, 68], collectively supporting our identification of FOXO TFs as effectors of metabolic rhythms in liver cells. Similar to FOXO TFs, Glucocorticoid hormones (GCs) also control glucose metabolism and stimulate gluconeogenesis or fat breakdown. Interestingly, [190] showed that GCs could counter the action of the circadian clock in resetting the phase of peripheral oscillators. Recently GR and FOXO were shown to be common regulators of many genes involved in gluconeogenesis and cell cycle progression, and a synergistic activity of these two TFs might exist [101]. In *Bmal1*<sup>-/-</sup> mice, under a night-restricted feeding regimen, insulin release in blood is still rhythmic [111]. In addition, [242] demonstrated that up to a hundred circadian transcripts are under glucocorticoid control and that these are distinct from clock-controlled circadian genes. GCs regulate the expression of the circadian TF PER2 [59], which is probably essential in mediating this circadian transcription. Indeed, GCs have been shown in rat to synchro-



### 3.5. Food entrainment and circadian regulation of lipid and sugar metabolism in liver

---

nize circadian gene expression in cultured fibroblasts and to affect the phase of circadian gene expression in the liver [24]. Our observations could therefore be compatible with a model in which two separate circadian signals are received from the system, resulting in the entrainment of the core clock by GCs through PER2. However, GCs can directly affect transcription of other targets through GR. Lastly, we observed that many diurnally transcribed genes in *Bmal1*<sup>-/-</sup> mice also exhibited a one-hour delay compared to the WT condition. This implies that a large proportion of regulatory elements would be activated sooner in the presence of the clock, consistent with the emerging notion that the circadian clock also acts to anticipate daily occurrences of feeding/fasting rhythms by activating relevant genes ahead of time [251, 185]. This is further consistent with observations that restricted feeding restores the circadian transcription of many metabolic genes [333], that feeding cues dominate light sensing as a Zeitgeber [70, 290], and that even in *Bmal1*<sup>-/-</sup> mice, food anticipatory behavior exists [51]. It is thereby interesting to note that the phase distribution in stringently selected cyclic regions is equally polarized at the two most prominent phases of expression, ZT6-10 and ZT18-22, in the *Bmal1*<sup>-/-</sup> compared to the WT condition. These observations suggest that in the *Bmal1*<sup>-/-</sup> genotype the molecular clock is not able anymore to regulate the metabolism in a proactive way whereby the production of some enzymes is initiated before the presence of food in order to optimize the energy consumption of the cell [337]. This goes against the idea of widespread disorganization of transcriptional control in the absence of a peripheral clock and underlines the strong effect of nutritional synchronization.

### 3.5 Food entrainment and circadian regulation of lipid and sugar metabolism in liver

The liver is a central player in adjusting metabolic processes to daily feeding–fasting cycles. This role is demonstrated by the circadian expression of many liver genes implicated in the metabolism of lipids, proteins, carbohydrates, and xenobiotics. The food entrainment of the liver clock has been widely studied using restricted feeding or inverted feeding [55, 333]. The clock impaired mice depict severe metabolic disorders such as type 2 diabetes and a reduced lifespan [102, 214]. In BMAL1 knockout mice, several studies showed that these mice were exerting glucose intolerance, increased respiratory quotient, reduced fat storage, increased circulating fatty acid, increased ectopic fat formation in liver and muscles, and hypoinsulinemia [185, 300, 12, 214]. Inactivation of the known clock components *Bmal1* and *Clock* suppress the diurnal variation in glucose and triglycerides [300, 1, 286]. However, we observed that the feeding regimen could partially restore some of these rhythms and counteract deleterious effect of the knockout. In the night restricted feeding Gilardi *et al.* reported that the glucose and insulin levels were close to normal [111], but in our analysis, we found that key enzymes of the glycolysis/gluconeogenesis had decreased amplitudes, and their phases were quite impacted by the *Bmal1* knockout. This almost normal glucose levels might be explained by CREB and FOXO1 signaling as these regulators are still oscillating in the impaired-clock genotype. Indeed, thanks to our linear model we observed that these two

regulators still depict an important activity in *Bmal1*<sup>-/-</sup> mice and we confirmed by western blot that the phosphorylation of CREB at the serine 133 is still rhythmic with a non-significant delay of two hours.

In addition we observed that circulating free fatty acids and  $\beta$ -hydroxybutyrate depict a higher concentration in blood in *Bmal1*<sup>-/-</sup> mice. Moreover  $\beta$ -hydroxybutyrate concentration increase during the fasting phase suggesting increased FFA oxidation and ketoacidosis, leading to reduced mitochondrial redox state [329]. The elevated circulating free fatty acids are consistent with previous reports on *Bmal1*<sup>-/-</sup> mice [286, 300, 185]. The level of triglyceride was slightly more elevated in *Bmal1*<sup>-/-</sup> and suggested a potential shift of 2h in *de-novo* lipogenesis.

Despite these metabolic stress, upon night restricted feeding the *Bmal1*<sup>-/-</sup> mice did not exhibit liver steatosis. In conclusion, the night restricted feeding regimen seems to counteract deleterious effect induced by a defect of the molecular clock.

### 3.6 Limitations of our approach

#### 3.6.1 Enhancer-TSS mapping

A large fraction of BMAL1 binding sites (60%) is located more than 10kb away from the nearest TSS [273], raising questions of circadian enhancers effect on transcriptional regulation. Interestingly, recent RNA-sequencing experiments revealed the presence of thousands of circadian enhancer-RNAs (eRNAs) in mouse liver [95]. As is often done in similar studies, candidate enhancers are assigned to target promoters according to genomic proximity. However, physical distance (measured in 3D) would be functionally more relevant.

An important issue in genomics is to find which enhancer regulates which promoter. Indeed there is a "many-to-many" relationship between promoters and enhancers. Enhancers can be associated with their target using several methods, such as nearest gene annotation or co-variation of gene expression and enhancer-associated histone modifications dynamics. However, these approaches are based on indirect evidence and are unable to detect "long-range interactions" [92]. Chromosome conformation capture experiments, which evaluate the frequency of cross-links between genomic regions, notably Hi-C which monitors interactions between pairs of genomic sites genome-wide [197] could provide experimental proofs of enhancer-promoter interaction or highlight co-regulated genes through hypothetical "transcription factories".

Importantly, functional interactions between genomic loci, such as promoter-enhancer pairs, mainly occur within a genomic scale typical of topologically associating domains TADs, i.e., dozens to hundreds of kilobases, and can be cell-type specific [142, 226, 295]. Consistently, our analysis using the distance threshold with our penalized linear model showed that enhancers are generally in a vicinity of 50 Kbp.

### 3.6.2 Characterization of transcription factor binding sites

In mouse, about 4800 proteins are annotated with the gene ontology term "DNA binding". On the other side, there are about 830 binding profiles (PWM) that have been characterized using a high-throughput method called HT-SELEX in human [156]. In addition, the last release of JASPAR database in 2016 contains 519 PWM for vertebrates [215]. Based on these numbers about 15-20% of DNA-binding motifs have been characterized. Moreover, some DNA binding proteins do not bind in a sequence specific manner [87], but they recognize open chromatin regions (and/or histones modification, or methylated DNA) or specific shape of the 3D structure of the chromatin.

In addition, ChIP-seq data set availability is quite limited in mouse liver with few dozens of transcription factors thanks to mouse ENCODE efforts [311], and recent studies on the transcriptional regulation of the circadian clock and the feeding-fasting cycle [171, 222, 273, 96, 95, 43, 93]. Furthermore, the data quality and the downstream analysis varies a lot between the different labs involved in this field. Moreover, a fraction ChIP-seq peaks frequently appear due to indirect binding of the factor and thus doesn't contain a DNA-binding motif [22]. Thus ChIP-seq is an excellent high-throughput method, which allows studying the genome-wide binding of a protein of interest, but the characterization of a DNA-binding motif using this technique is not always straight forward.

Accordingly, our approach is limited by the knowledge of transcription factor binding sites (PWMs). *De novo* motif finding cannot be applied to DNase I hypersensitivity because these classical algorithms are not designed to find every PWM at once. Moreover, there is an "underdetermination" of PWMs as several proteins can bind to the same binding site.

Overall, recent efforts in the study of transcription binding sites have drastically increase our knowledge of how proteins can bind DNA in a sequence specific manner [156]. Interestingly, a model that combines the PWM with a DNA shape feature-based regulatory potential score has been recently developed and improved the accuracy in detecting binding sites [347]. Another recent strategy to improve transcription factor binding site detection is based on hidden Markov models and called transcription factor flexible models. TFFMs are flexible, and can model both position interdependence within TFBSs and variable length motifs within a single dedicated framework [216]. A last improvement of transcription factor binding site detection implies *deep learning* techniques, which offer a scalable, flexible and unified computational approach for ascertaining sequence specificities and pattern discovery from experimental data [7].

## 3.7 Conclusions

At the molecular level, the relationship between the circadian cycle and nutrient response cycle is unclear. Moreover, the role of distal regulatory elements remains elusive. Therefore, the CycliX consortium aimed at characterizing the transcriptional programs that are implicated

### Chapter 3. Discussion

---

in both cycles. Thus, our study shed light on the interrelationships between the nutrient-response cycle and the circadian clock as well as the contribution of the distal regulatory elements to the circadian control. In addition, we investigated some aspects of the tissue-specific regulation of the circadian cycle.

Our investigation of genome-wide DNase I hypersensitive sites (DHSs) in mouse liver led to the discovery of 62000 accessible sites consistent with the previous reports on chromatin accessibility [311]. In addition, 98000 footprints were detected in about 3/5 of these accessible regions, which indicate that some of these accessible sites are potentially due to a high transcription such as in the albumin gene. These DHSs that does not contain any footprints and are probably not regulatory elements, despite the fact that some transcription factors depict a weak footprint, which is difficult to detect [224]. Furthermore, we observed that DHS around TSS tend to contain more footprints, which indicates that the promoters are critical for the control of the transcriptional regulation.

Our Study of the temporal dynamic of Pol II loadings, H3K27ac and DNase I signal at DHSs in WT and *Bmal1*<sup>-/-</sup> revealed that about 10 % of the accessible sites depicted a circadian behavior. Interestingly, this proportion is similar to the percentage of oscillating transcripts in mouse liver [17, 188], which suggest their implication in the circadian transcriptional regulation.

Our analysis of transcription factor binding sites using digital genomic footprinting methods in WT and at ZT6 in *Bmal1*<sup>-/-</sup> showed that the chromatin accessibility was dynamically influenced by the binding of transcription factors, as previously reported [20, 237]. In addition, we observed that DNase I cleavage pattern reflected DNA binding of protein complexes such as a BMAL1:CLOCK heterotetramer on double E-boxes. This result increases the evidence of the functional role of these double E-boxes [273, 234].

The investigation of putative BMAL1 cooperating transcription factor and tissue-specific factors in liver and fibroblasts revealed several putative BMAL1 co-regulators such as NRF1 and ZEB1, and tissue-specific factors such as HNF4, CEBP, FOXA, and GATA1, consistently with the promoter analysis performed by Bozek *et al.* [38, 39]. Our results contradict somehow the pioneering-like function of BMAL1 proposed by Menet *et al.* [222]. Indeed, the circadian clock probably requires tissue-specific pioneer factor to support tissue-specific functions. Otherwise, the transcriptional regulation of downstream genes by the clock would be similar in every tissues and BMAL1 would bind at the same locations.

Our study of putative circadian regulators and regulators implicated in nutrient response cycle using a linear model approach in WT and *Bmal1*<sup>-/-</sup> revealed the importance of GR, FOX, and CREB in clock impaired mice under night restricted feeding. These factors have been already related with systemic cues that affect the liver synchronization [266, 82, 93]. Thus, we observed a partial disorganization of transcriptional control in the absence of a peripheral clock, which underlined the strong effect of nutritional synchronization. Besides, global *Bmal1* knockouts in mice lead to important effects on metabolism and behavior that could be traced

to dysregulation of BMAL1 target genes [273].

In this study, we accumulate some evidence of the contribution of distal regulatory element involved in the circadian transcription regulation. Thus, we observed that about 47% of DHS are located at more than 10 Kb from the closest active TSS. Moreover, distal DHSs up to 50 Kb improved the variance explained by our penalized linear model in WT and *Bmal1*<sup>-/-</sup>. Such regulatory elements were investigated in the circadian context using maps of the activity related chromatin mark H3K27ac [334], as well as enhancer RNAs (eRNAs) [95]. These studies consistently identified thousands of putative enhancers with a broad range of peak activity times, and were associated with distinct DNA regulatory motifs and TF binding patterns.

Finally, our inspection of diurnal biological processes such as lipid and sugar metabolism and their regulation in WT and *Bmal1*<sup>-/-</sup> confirmed the implication of SREBP, ChREB, CREB and FOX in these processes, as previously suggested. [111, 93, 240].

## 3.8 Perspectives

### 3.8.1 Benefits of chrono-nutrition and chrono-therapy

A diurnal rhythm of eating-fasting promotes health [112]. In addition, it was recently demonstrated that the restricted feeding regimen can compensate the effect of high-fat diet in mouse [130]. In our results, we observed that in BMAL1 knockout upon night restricted feeding, a large part of the transcriptome still depict diurnal variations, and we did not observe previously reported hepatic steatosis [300, 324]. Overall, Well-regulated eating habits actively contribute to better lipid metabolism, even if animals ingest a high-fat diet. Recent findings show that time-restricted feeding during the active phase amplifies circadian clocks and decrease metabolic disorders induced by a high-fat diet without caloric reduction, whereas irregular food intake causes various metabolic dysfunctions. Such evidence from nutritional studies that consider circadian system (chrononutrition) has rapidly accumulated. Furthermore, in the near future, chrononutrition might eventually be used as a new tool in medical protocols to treat metabolic syndrome or diabetes.

Epidemics of metabolic diseases (e.g., cardiovascular diseases, type 2 diabetes, obesity, metabolic syndrome and certain cancers) have become major contributors to the cost of poor health, and they are presently accelerating, in most developing countries. In addition, light at night (or light pollution) disrupt circadian rhythms and is commonly referred to chronodisruption. These chronodisruption might impact the hormonal signaling. Numerous hormones depict circadian regulation that significantly impact physiology and pharmacology [69]. Melatonin, a circadian hormone of the pineal gland, influences various aspects of retinal [322] and cardiovascular [81] function and affects local clocks in diverse brain regions [126]. Melatonin rhythmicity has important roles in a variety of metabolic functions as an anti-oxidant, anti-inflammatory chronobiotic and possibly as an epigenetic regulator [174].

## Chapter 3. Discussion

---

Circadian regulation of the adrenal gland leads to diurnal secretion of glucocorticoid hormone, which impact metabolism and directly affects 60% of the liver transcriptome [266]. Circadian control of gastrin, ghrelin, and somatostatin, along with direct regulation by autonomous clocks within the gastrointestinal tract, enable circadian influences on digestive function [173]

Current research highlights that circadian clocks in peripheral tissue can respond to different cues and thereby show different phase relationships. Thus, full prediction of chronopharmacology in pathological contexts will probably require a systems biology approach that considers chronointeractions among the various clock-regulated systems [69].

In most cases, autonomous circadian clocks not only within the gastrointestinal tract but also in various other tissues have significant impact on physiology and metabolism. For instance, ablation of clocks in pancreatic islets leads to diabetes due to defects in the coupling of  $\beta$  cell stimulus to insulin secretion [214], and tissue-specific targets of the clock includes multiple ion channels and kinases in the heart that influence cardiac function and triglyceride metabolism [169, 323]. Recent transcriptome studies have identified widespread tissue-specific circadian regulation not only in heart but also in skeletal muscle and fat, indicating that clocks in these tissues directly regulate physiology [40].

Considering the broad scope of circadian (patho)physiology, it is logical that the pharmacokinetics and pharmacodynamics of many drugs would be circadian and thus that drug efficacy and safety profiles would also change with time of day. Nevertheless, this variation is poorly considered by clinicians, drug developers, and regulators. In part, this lack of interest arises from a lack of awareness of the molecular mechanisms underlying this control. However, two decades of intensive research have revealed - not only basic mechanisms of circadian clocks but also about how these mechanisms impact physiology and disease [69].

### 3.9 More circadian layers of regulation in cells

In addition to circadian transcription, recent research has revealed extensive evidence of circadian post-transcriptional regulation in mammals, along with translational control [157], control of transcription termination and/or elongation [247], and circadian control of splicing [219]. Thus, the current number of transcripts exhibiting circadian abundance is significantly greater than the number of genes transcribed in a circadian manner [171, 188, 223], and the number of proteins that are expressed in a circadian manner is greater than the number of transcripts [265]. Major signaling molecules such as cAMP show circadian variations that both control clock output and play a role within the clock [243], and recent links between clocks and sirtuins suggest a similar influence of redox potential [16]. Furthermore, a significant fraction of histone posttranslational modifications varies in a circadian fashion at a large number of loci [171, 188]. Overall, through a collection of different mechanisms, a significant amount of cellular physiology is regulated by the circadian clock. However, additional regulations, such as the rhythmic degradation of intermediate products, can further modulate gene expression, e.g., by adjusting the amplitude or shifting the phase of the driving rhythm



[209]. For instance, MiR-122 is important for the hepatic circadian functions and targets genes involved in cholesterol and lipid metabolism [106]. Using an inducible DICER knock out, an recent study assessed the global contribution of miRNAs to circadian gene expression in mouse liver [83]. Although a non-negligible fraction of circadian transcripts was affected by miRNAs, this study emphasized the resilience of the circadian clock to perturbations in miRNA biogenesis. Furthermore, in cultured mouse cells, the temperature entrainment of the clock was shown to involve post-transcriptional mechanisms. Notably, the Cold-Inducible RNA-binding protein (CIRBP), possibly by binding to transcripts encoding circadian oscillator proteins, such as Clock [230]. Interestingly, CIRBP and RBM3, another RNA-binding protein, guide the choice of polyadenylation sites of target transcripts in a circadian manner [201]. Nevertheless, underlying mechanisms are still poorly understood, and further efforts are needed to appreciate better the various mechanisms controlling circadian gene expression at multiple levels.

#### 3.9.1 Emerging trends in functional -omics and in (chrono)biology

##### Fundamental research

In the post-genomic era, the circadian transcription has been systematically studied using various experimental design, such as inverted feeding, night restricted feeding, in light-dark (LD) or dark-dark (DD) conditions. Moreover, the availability of several mouse strain with a defect of several core clock genes enabled the investigation of different component of the core clock [210]. The technological improvement concerning sequencing technologies [307] as well as mass spectrometry [9] have allowed the study of the circadian cycle with an unprecedented throughput. Today it possible to study the whole transcriptome, as well as a large fraction of the proteome [218] or the lipidome [1] and the metabolome in a single experiment. In addition, transcription factor binding sites or regulatory elements , as well as binding sites of RNA binding proteins on mRNA can be systematically investigated [53]. These experiments are a valuable source of information to understand the molecular changes during the diurnal cycle. Cutting edge research are currently deciphering tissue or cell types specificities [355], as well as the sub-cellular level with targeted researches on the nuclear proteome and lipidome or in organelles such as the mitochondrial lipidome [19]. A general trend in biology is the shift towards single cell analysis in order to study the heterogeneity of the gene/proteins expression in a cell population or tissue. For instance, new methods allow the investigation of mRNA [344] and chromatin accessibility [42] in single cells. Interestingly, novel methods of microscopy allow the study of single molecules in a living cell [183]. These techniques are critical to study biological processes such as transcription in a quantitative manner even with "near endogenous" conditions.

### **Translational research**

Last but not least, the heterogeneity in a population of individuals in terms of genome sequences with, for instance, single nucleotide polymorphisms (SNP), might be another source of useful information in order to associate a macroscopic phenotype to molecular mechanisms [63]. Recent efforts of Biobanks in terms of human sample and research agreement collection [229] will enable population-scale analysis such as the 1000 genome project [18]. However, the genome sequence, it-self, is not sufficient and should be integrated with other data types such as activity, medical records (diseases, response to medication) or behavior, as well as other molecular data types (epigenetics, transcriptomics, metabolomics) in order to pave the way for personalized medicine.



## 4 Methods

### 4.1 Experimental design, protocols and data set quality control

The experimental design and the bench work was done by the Cyclix Consortium. More precisely Nouria Hernandez Mauro Delorenzi, Bart Deplancke, Béatrice Desvergne, Nicolas Guex, Winship Herr, Felix Naef, Jacques Rougemont and Ueli Schibler designed the experiments and supervised the whole Cyclix project. The animal breeding, the conditioning and the collections of biological material were performed by Teemu Andersin, Pascal Cousin, Federica Gilardi, Pascal Gos, Gwendal Le Martelot and Fabienne Lammers. The chromatin immunoprecipitations, the libraries preparation and the gene expression arrays were done by Donatella Canella, Federica Gilardi and Sunil Raghav.

The bioinformatics analysis was mainly a collaboration between Irina Krier, and me. This work was supervised by Felix Naef and Bart Deplancke, with the help of Jacques Rougemont.

#### Animals breeding and liver sample preparation

C57BL6 and *Bmal1*<sup>-/-</sup> male mice 12–14 weeks old (at time of sacrifice) were housed in a 12 h light/12 h dark (LD) regimen. They were then entrained to a 12 h/12 h LD regimen with water ad libitum but food access only between ZT12 and ZT24 for Pol II ChIP-seq and H3K27ac ChIP-seq and microarray for 7 d (ZT, Zeitgeber time; ZT0 is defined as the time when the lights are turned on and ZT12 as the time when lights are turned off) for 7 days before the sacrifice. Mice used for DNase I-seq were entrained to a 12 h/12 h LD regimen with water and food ad libitum. At each ZT2, ZT06, ZT10, ZT14, ZT18, ZT22, and ZT26, 3–5 mice were anesthetized with isoflurane and decapitated. The livers were perfused with 2 ml of PBS through the spleen and immediately collected. A small piece of liver tissue (approx. 100 mg) was snap-frozen in liquid nitrogen and kept at 80°C for RNA extraction. The remaining liver tissue was immediately homogenized in PBS containing 1% formaldehyde for chromatin preparation. All animal care and handling was performed according to the State of Geneva's law for animal protection.

### **Chromatin immuno-precipitation of RNA Polymerase II followed by sequencing**

Perfused livers were processed for chromatin preparation as described in [277]. The chromatin samples from the five mice were then pooled, frozen in liquid nitrogen, and stored at  $80^{\circ}\text{C}$ . For the ChIP experiments, the following antibodies were used: anti-RPB2 (Santa Cruz Biotechnology, sc-673-18), anti-H3K4me3 (Abcam, ab8580), and anti-H3K36me3 (Abcam, ab9050). To determine the optimal amounts of each antibody, we performed pilot ChIP assays and determined the enrichment for a set of promoters by real-time qPCR according to [277]. A total of 1 ml of each chromatin suspension (containing about 60  $\mu\text{g}$  of DNA) was incubated with 10  $\mu\text{g}$  of anti-RPB2, 1.5  $\mu\text{g}$  of anti-H3K36me3, or 1.5  $\mu\text{g}$  of anti-H3K4me3 in buffer A (20 mM Tris/HCl (pH 7.5), 150 mM NaCl, 2 mM EDTA) overnight at  $4^{\circ}\text{C}$  on a rotating wheel. Ten  $\mu\text{l}$  of protein A bead suspension (25% slurry in buffer A), pre-blocked with 10  $\mu\text{g}/\text{ml}$  of salmon sperm DNA and BSA at  $4^{\circ}\text{C}$  overnight, was then added and the incubation was continued for 1 h at room temperature on a rotating wheel. The beads were then washed with dialysis buffer and ChIP wash buffer as described in [239]. Protein–DNA complexes were eluted from the beads, de-cross-linked, and treated with RNase A and, subsequently, with proteinase K, as described in [277]. The DNA concentration was determined by fluorometry on the Qubit system (Invitrogen). A total of 10–12 ng DNA were used for the preparation of the library. Libraries for ultra-high throughput sequencing were prepared with the ChIP-Seq DNA sample kit (Illumina) as recommended by the manufacturer.

### **Chromatin immuno-precipitation of Histone 3 lysine 27 acetylated followed by sequencing**

ChIPs were performed according to the method described by [267] with a few modifications. The 100  $\mu\text{L}$  chromatin aliquots was used for each IP and diluted with 900  $\mu\text{l}$  of RIPA buffer (1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS in PBS at pH 7.4) and added to dynal magnetic beads conjugated with (Sheep-antimouse IgG dynabeads, Invitrogen, Cat no: 110-31) pre-treated with 3  $\mu\text{l}$  of polyclonal antibody for H3K27ac (Active motif, Cat no: 39135) for immunoprecipitation of specific complexes. The samples were incubated overnight at  $4^{\circ}\text{C}$  on rotator, then magnetic beads washed 7 times with lithiumchloride wash buffer (100mM Tris at pH 7.5, 500mM LiCl, 1%NP-40 and 1%sodiumdeoxycholate) and finally once with 1X TE buffer (10mM Tris-HCl at pH 7.5, 0.1mM Na<sub>2</sub>EDTA). The chromatin complex was eluted using elution buffer (1% SDS, 0.1MNaHCO<sub>3</sub>) for 1 h at  $65^{\circ}\text{C}$  using eppendorf thermo-mixer. The chromatin was then de-crosslinked overnight at  $65^{\circ}\text{C}$  and ChIP DNA purified using Qiagen PCR purification kit and eluted in 50  $\mu\text{l}$  of elution buffer. For qPCR reaction 1.5  $\mu\text{l}$  of 1/10 diluted ChIP DNA is used. Libraries for ultra-high throughput sequencing were prepared with the ChIP-Seq DNA sample kit (Illumina) as recommended by the manufacturer.

## 4.1. Experimental design, protocols and data set quality control

---

### DNase I-seq

Mouse liver nuclei were prepared as described in [321]. Freshly prepared nuclei were suspended in ice-cold  $\psi$ -buffer (11 mM KPO<sub>4</sub> pH 7.4, 108 mM KCl, 22 mM NaCl, 5mM MgCl, 1 mM CaCl<sub>2</sub>, 1 mM DTT) and pelleted. 5x10<sup>6</sup> nuclei were suspended in 200  $\mu$ l of  $\psi$ -buffer supplemented with 0.2% of NP40 and 1  $\mu$ /ml of DNase I (DPFF Worthington Biochemical Corporation). DNase I digestion was performed for 6 minutes in room temperature and the reaction was stopped by adding 200  $\mu$ l of lysis buffer (50mM Tris-HCl pH 8, 20 mM EDTA, 1% SDS, 200  $\mu$ g/ml proteinaseK). Protease digestion was performed overnight at 55°C. RNaseA (100  $\mu$ g/ml) was then added and samples were incubated at 37°C for an hour. DNA was then extracted twice with phenol-chloroform and precipitated with isopropanol in the presence of 0.5 M NaCl. DNAs were dissolved in 5 mM Tris-HCl pH 8. DNAs from 4 animals were pooled and 75  $\mu$ g of DNA was loaded on 11 ml 10%-50% sucrose gradient in STE buffer (1M NaCl, 20 mM Tris-HCl pH 8, 5 mM EDTA) and centrifuged at 30000 rpm for 16 hours at 20°C (SW 40 Ti rotor, Beckman Coulter Inc). The sucrose gradients were then fractionated and DNA was precipitated by two volumes of ethanol in the presence of 5  $\mu$ g of glycogen. Fractions containing DNA sized around 300bp were pooled and used for Illumina library preparation.

### Quality control (QC)

Sequenced reads alignments, from our illumina libraries to mouse genome (mm9 assembly), were performed using HTSstation [72]. This web interface enables sequencing and mapping quality control through automatic generation of reports. The alignment software used in the HTSstation pipeline is Bowtie, with default parameters [186]. After mapping, the number of reads that were obtained for each sample is indicated in table 4.1. Every sample used, had more than 70% of mapped reads (see Table 4.1).

Most of the CycliX libraries were single-end sequencing. But in few cases paired-end sequencing was used, it was processed as paired-end reads, but later quantified as single-end reads, i.e. only the beginning of a read was used.

The Fastqc quality control report (used in HTSstation pipeline) was decent for all libraries produced in this project, although heterogeneity in proportions of duplicated reads was observed, especially for the H3K27ac ChIP-seq in *Bmal1*<sup>-/-</sup> genotype.

For high depth libraries (with more than 100 million reads), it is expected that stacking of reads will occur independently from PCR-amplification bias. This is in particular the case for DNase I data, where the enzyme cleavage site determines the mapped read start and is expected to occur preferentially at the edge of a transcription factor binding site. Cleavage efficiency by the enzyme determines the read depth at these positions and constitutes the signal which we would like to measure. Therefore, we did not apply methods to remove stacked reads. Multiply mapping reads with more than 5 mapping positions were discarded. For quantification, reads with N alignments were counted as 1/N reads mapping at each position. The sequencing

## Chapter 4. Methods

---

libraries used in this project were large, containing each between 100 million and 200 million reads (see Table 4.1).

We therefore used a binary format conversion to save time for follow up computations, developed by Nicolas Guex for the CycliX consortium along with a tool for quantifying signal at genomic locations of interest. This tool enables fast computing, thanks to Vital-it Cluster infrastructure, by using efficiently the large memory capability of the cluster and by loading each chromosome at once. Binary files were generated using only the 5' end of reads corresponding to the DNase I digestion position or the break induced by sonication for ChIP-seq. BigWig files were subsequently generated using the HTSstation [72] with centering parameters computed from forward strand versus reverse strand cross-correlation maximization on ENSEMBL67 TSS for each ChIP-seq sample separately.

In order to control and compare the quality of the DNase I time points we used the percentage of reads in the enriched regions (called Percentage of Tags in Hot-spots, PTIH) [155]. This control is done after the peak calling. The PTIH is simply the number of reads in peaks divided by the number of reads that were mapped to the genome.

These quantification are reported in table 4.2. Briefly, we observed that ZT02 was of lower quality compared to the other DNase I samples.

Several other web-interface, have been recently developed to perform quality controls, quantification and analysis of high-throughput data, as Galaxy [2], Chipster [161] or the ChIP-seq server from the Computational Cancer Group at EPFL [10]. I was able to test as well some open access command line tools, and R or python libraries to perform these quantifications quite efficiently, notably with NGSplot [202].

#### 4.1. Experimental design, protocols and data set quality control

Table 4.1 – the number of reads in each sample and the number of mapped reads on the mm9 assembly are reported.

Library	Number reads	Numbers mapped	%mapped
DNase I WT			
ZT2	249350929	174147376	70
ZT6	317274205	252864063	80
ZT10	269194001	210684885	78
ZT14	204984161	159264085	78
ZT18	194932779	146829749	75
ZT22	191842671	167303697	87
ZT26	190983184	161216246	84
Pol2WT			
ZT2	220172973	203597437	92
ZT6	189640820	168723064	89
ZT10	162253851	145450110	90
ZT14	196057079	176252350	90
ZT18	254879645	234013533	92
ZT22	142117355	122741922	86
ZT26	191096558	168972009	88
H3K27acWT			
ZT2	552441072	476877194	86
ZT6	569270357	496184118	87
ZT10	586341426	515957037	88
ZT14	461236082	371945828	81
ZT18	568258389	492287638	87
ZT22	535729511	466984801	87
ZT26	407807224	344924054	85
Pol2 Bam1 KO			
ZT2	509081190	442613199	87
ZT6	456861768	379886582	83
ZT10	615333770	560876871	91
ZT14	538202935	492830944	92
ZT18	567313327	518551916	91
ZT22	567987125	506596171	89
ZT26	605314961	424032099	70
H3K27ac Bam1KO			
ZT2	208280807	158653084	76
ZT6	207101049	155183877	75
ZT10	208889844	166330282	80
ZT14	251404424	186223464	74
ZT18	206733328	166182481	80
ZT22	208941050	160127444	77
ZT26	173757120	134496020	77
DNase I Bam1-			
ZT6	287086306	206398188	71

## Chapter 4. Methods

---

Table 4.2 – the percentage of tags mapped in the mm9 genome that are enriched in peaks (see peak calling section).

DNase	Quality Control
Sample	Percentage of tag in DHS
DNase I WT ZT2	29
DNase I WT ZT6	38
DNase I WT ZT10	57
DNase I WT ZT14	50
DNase I WT ZT18	54
DNase I WT ZT22	58
DNase I WT ZT26	46

## 4.2 Peak calling optimization

### State of the art

Peak calling is a critical step in the processing of ChIP-seq or DNase I-seq data. Its role is to determine respectively the actual binding sites or accessible sites from sequenced DNA fragments mapped onto a reference genome sequence [193, 21, 285, 315]. In ChIP-seq or DNase I-seq experiments, only reads with a single matching locus on the genome are taken into account. The positions of mapped reads are subsequently used as input for peak-calling algorithms, which predict protein binding sites or accessible sites in DNase I-seq. The main difficulty of peak calling is to determine the significance of peaks. Due to a high variability in the background distribution of reads and the relatively small amount of enriched DNA, one strategy consists of modeling the background distribution statistically and searching the data set for local deviations from this distribution. Several models describing reads densities in the background have been considered including the Poisson distribution [357], the Negative Binomial distribution [152] and Hidden Markov Models [309].

A current approach used to determine background distributions of reads, is to work with mock immuno-precipitation, notably using a non-specific antibody or without antibody as a control [164]. In this input DNA or control data there is no enriched binding fragments. These controls can be used to empirically estimate false discovery rates of the real ChIP data [345].

Another important point is to exclude 'non-mappable' regions, i.e short subsequences occurring multiple times in the genome or regions containing copy number variants (CNV). These regions may be artificial source of read densities [328]. A last point for the peak-calling is to consider the wideness of the considered protein. For instance the shape of peaks for a transcription factor or DNase I accessible sites can be of few hundreds of base pairs, and in the other side the shape of peaks of histones modification might be much larger [21]. Recent approaches imply deconvolution of overlapping peaks or modeling of peak shape and estimation of peak parameters using an Expectation–Maximization algorithm [356, 221, 273]. A number of recent reviews focus on ChIP-seq technology, applications and software, and several studies worked on benchmarking the peak calling by producing gold standard ChIP-seq samples used to compare different algorithms and software [287, 315, 125].

### This study

In our study, we started with the mapping of DNase I reads to the mm9 mouse assembly and we merged all time points together before doing the peak calling. Then, we compared several peak-calling tools as [357], HOMER [135], ChIP-Peak and ChIP-Part [10]. Each of these tools was performing well with some adjustments of their parameters, but we decided to use ChIP-Peak because we could separate efficiently single peaks of DNase I signal with a specified peak width and an exclusion parameter called vicinity instead of calling large hypersensitive regions. We found that the optimal parameter was to use a 600 bp peak width and a vicinity of

## Chapter 4. Methods

400bp. Overall, we detected 62000 peaks in WT (C57BL6) mouse liver, which covers 1.2 % of the mouse genome.

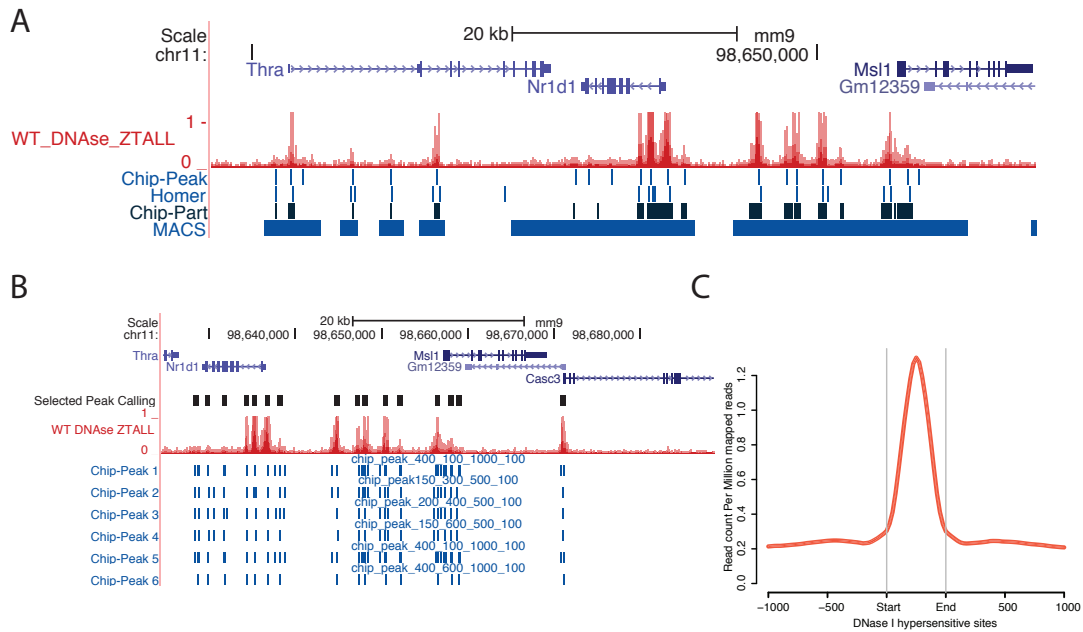


Figure 4.1 – Peak calling strategy for DNase I data: A) ChIP-peak, ChIP-Part, HOMER and MACS peak calling on DNase I-seq signal, B) parameter optimization of ChIP-Peak, C) Average normalized DNase I signal on 62000 DHS



## 4.3 Signal normalization

### State of the art

The normalization is a critical step in DNase I-seq or ChIP-seq or any NGS data analysis. Several methods have been used in recent studies, as scaling to total amounts of reads (i.e., normalizing for sequencing depth), quantile normalization or locally weighted regression (LOcally WEighted Scatterplot Smoothing ,LOESS) normalization [188, 233, 195, 317, 171]. A variation of the scaling to total amounts of reads where a scale factor is estimated in a region ( 10 Kb) using a linear regression was applied in several ChIP-seq studies and are improving peak calling (see peak-calling section). Quantile and LOESS normalization are based on the assumption that the effect of biological condition change does not cause global binding alterations. This assumption might be not valid, when comparing samples with different stages of disease progression, or on samples before and after a certain treatment.

Another scaling normalization method is widely used in RNA-seq data analysis. This method is known as RPKM (Reads per Kilobase of sequence range per Million mapped reads). RPKM adjusts for biases due to the higher probability of reads falling into longer regions. The intended meaning of RPKM is a measure of relative molar RNA concentration (rmc) and [335] showed that for each set of transcripts the average rmc is a constant, namely the inverse of the number of transcripts mapped. A correction of the RPKM normalisation called transcripts per million (TPM) respects the average invariance and eliminates statistical biases inherent in the RPKM measure. Several other normalization methods have been proposed for RNA-seq data analysis, as upper-quartile-normalized counts or Trimmed Mean of M values (TMM), and these methods have been compared in a comprehensive study [78]. The main conclusion of this study is that RPKM performed poorly compared to these other methods. Normalization issues might have a substantial impact on the results and therefore should be addressed carefully [78, 335].

More recently the CycliX consortium, and other groups tested a spike adjustment procedure (SAP) designed to allow comparison of occupancy levels for a set of loci of interest [245, 34]. It consists of adding a constant, low amount of a single batch of foreign chromatin (e.g., human) as an internal control to each sample of the chromatin of interest (e.g., mouse) before immunoprecipitation. This allows adjustment of the signals in each sample to the internal control. The CycliX consortium showed that unlike only scaling to the total amount of aligned sequence tags or quantile normalization, the SAP allowed the scoring of global and largely uniform changes when they result from biological differences [34]. Nevertheless, the behavior of spike-ins during library preparation is essential before applying any normalization method that makes use of them [279].

### This study

In this project, we used a quantile normalization on DHSs (+/-300 bp around DHS center, for DNase I and Pol II signal, and +/-1Kbp for H3K27ac). We applied the same procedure at TSSs (+/- 1Kbp). Finally, in the case of the micro-array data, we used a standard micro-array normalization procedure, namely the robust multi-array average (RMA) normalization [150].

## 4.4 Assessment of oscillatory signal

### State of the art

Identification of rhythmic features such as genes or regulatory elements is a decisive stage to identify clock-controlled pathways and processes. Such features are largely detected by searching periodic patterns in any types of data. For example, a lot of studies have been performed using microarrays, RNA-seq or ChIP-seq data [171, 333, 334, 143, 95, 355].

In our case we were interested to detect these oscillating features using DNase I-seq, histone 3 lysine 27 acetylated (H3K27ac) and polymerase II (Pol II) ChIP-seq. In order to detect these features several methods have been recently developed, such as COSOPT [143], ARSER [348], JTK-cycle [145], Fisher G-test [99], RAIN [318] and linear regression based methods. These methods are able to perform quite well on large data-sets with low sampling frequency and few or no replicates and high levels of noise. Several study compared these different methods and enhanced them in order to detect asymmetric patterns, non-sinusoidal periodic waveforms, or to cope with missing values [147, 318].

### This study

In this project we had a sampling of 7 time points with a 4 hour resolution on a single cycle (ZT02-ZT26) that did not provide a high statistical power. On the other hand, we had several marks probed on the same animals and 3 mice pooled per time points. Therefore we applied a harmonic regression on Log2 transformed signal at DHSs or at TSSs, and we did not consider asymmetric or non sinusoidal waveforms.

Equation 4.1 was used to perform least squared fit of temporal profiles  $y_t$ .

$$y_t = A_0 + A_1 \cos(\omega t) + A_2 \sin(\omega t) + \epsilon \quad (4.1)$$

where  $A_0$ ,  $A_1$ ,  $A_2$  are regression coefficients and  $\epsilon$  is white noise.

$\omega = \frac{2\pi}{\tau}$  is the frequency and we take a diurnal period  $\tau = 24h$ .

Once we compute the three coefficients using the standard linear regression (lm function in R), we can compute the phase  $\phi$ , amplitude  $A_{24}$  by writing:

$$A_1 \cos(\omega t) + A_2 \sin(\omega t) = A_{24} \cos(\omega t - \phi) \quad (4.2)$$

with

$$\begin{cases} A_{24} = 2\sqrt{A_1^2 + A_2^2} \\ \phi = \tan^{-1}\left(\frac{A_2}{A_1}\right) \end{cases} \quad (4.3)$$

We then computed the  $p$ -value of the fit using a  $F$ -test. The  $p$ -value associated with 24h rhythmic profiles was computed for each DHS and for each mark. The Fisher combined probability of Pol II, H3K27ac and DNase I was computed to select rhythmic DHSs. Specifically, we computed the  $p$ -values for each of those  $k = 3$  marks and combined them using the statistics, which assumes a Chi-squared distribution with  $2k$  degrees of freedom [177].

$$-2 \sum_{i=1}^k \ln(p_i) \sim \chi_{2k}^2 \quad (4.4)$$

The resulting combined  $p$ -value was used to estimate False Discovery Rates (FDR) via the linear step-up method using the Benjamini-Hochberg method in R [31].

### 4.5 Motif analysis in DNase I hypersensitive sites

#### State of the art

Motifs analyses are frequently done after ChIP-seq experiments [21]. In the usual pipeline there is a first mapping of the reads to a reference genome followed by a peak calling (see peak calling methods section 4.2). Then the sequences of these peak locations are retrieved and used to find enriched motifs or position weight matrix (PWM) [77]. There are two general philosophy of finding motifs. The first one uses *de novo* motif finding algorithms, such as RSAT or MEME [319, 23], followed by motif comparisons (TOMTOM) [121] with PWMs from databases, as Jaspar or Transfac, to identify known and unknown motifs [258, 217]. In the second one, we can directly scan the sequences with PWM database. PWMs scanning tools, such as FIMO, use a log-likelihood ratio score and a *p*-value threshold in order to find positive matches of PWMs in the given sequences [114].

Another common way to scan a sequence for a given set of motifs is to use a HMM as with the MCAST motif scanning tool from MEME Suite [115].

#### This study

In this project we tested the two options (*de novo* motif finding or motif scanning) but we finally decided to use only known motifs from databases and a straightforward motifs scan of our 65000 peak sequences of 600bp defined by our peak calling. We made this choice because these *de novo* motifs finding tools can produce a lot of false positive especially in DNase I-seq data, as accessible sites sequences are not only enriched for a single transcription factor but for all DNA binding proteins that are expressed in the considered sample. Moreover, it is quite difficult to use unknown motifs for any interpretation without experimental validation.

In order to determine the right threshold parameter for our motif scans we did a stratification of the FIMO DNA motif match *p*-value on 3306 sequences of DHS that are oscillating for DNase I signal in WT mice in figure 4.2. We tried several PWM related to the circadian rhythm as E-box (CLOCKBMAL), and D-box (VBP). As expected in these oscillating accessible regions, we had a high proportion of match for these motifs. We found that the number of match is increasing exponentially with the *p*-value threshold, and we concluded from this optimization that the threshold of FIMO *p*-value below  $10^{-4}$  gave a reasonable amount of match with a well-conserved motif and a visible raw footprint in DNase I signal in an average plot centered around our motif matches. In summary we used a stringent threshold in order to favor specificity over sensitivity and thus avoid false positives.

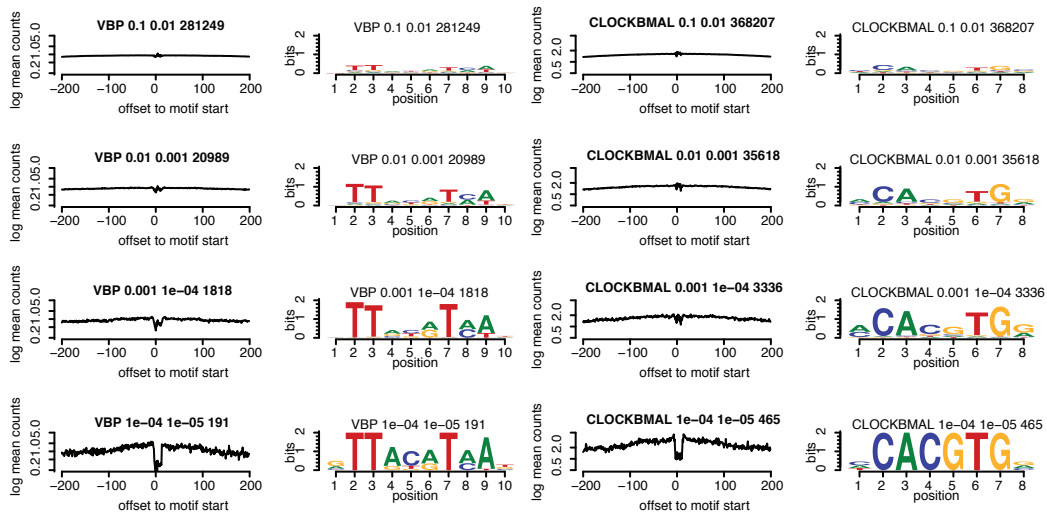


Figure 4.2 – FIMO threshold stratification for CLOCK/BMAL and VBP motifs in 3306 circadian DHSs

## 4.6 DNase I-seq footprint detection

Analysis of digital genomic footprinting data can be separated roughly into two categories, namely TFBS motif-centric and footprint-centric approaches. The first category attempt to quantify TF binding at defined genomic locations, whereas the second category focus on *de novo* detection and annotation of DNase I footprints [331]. Moreover, a pending question of the DNase I cleavage rate bias is still debated in the literature (see section 4.6.3). Following sub-sections will explain the different methods, as well as the one that we used in this project. Finally, I will highlight the current controversy about the DNase I cutting bias and briefly present the results of a comparative study of DNase I footprint detection methods.

### 4.6.1 TFBS motif-centric approach

#### State of the art

Genomic match to position weight matrices for hundreds of TFs can be generated readily with algorithms such as FIMO [114] in DNase I hypersensitives sites (see motif analysis section 4.5). TF binding at these recognition sites can be quantified under the assumption that occupied sequence specific TFBS have significant different cleavage pattern than expected. This strategy is quite efficient because using PWM motif match as a prior provide the location and the expected width of the of the TF-DNA interaction. Thanks to these information provided, a common strategy is to explicitly model the per-nucleotide cleavage pattern around PWM match for a given TF and use machine-learning approaches for classification such as in CENTIPEDE method. This method models bound and unbound cleavage pattern using multinomial distributions containing parameters for each nucleotide in the window surround-

ing the motif [257]. The learning strategy used by CENTIPEDE is representative of a broad class of unsupervised methods for DNase I footprint analysis that perform model parameter optimization and classification simultaneously and directly using the observed data. On the other side, supervised learning methods have been developed, such as FLR and BinDNase [350, 184]. They require positive and negative labeled training data (based on, for example, ChIP-seq peaks or other occupancy measures) to fit model parameters that are subsequently used for the discrimination of TF recognition sequences of unknown binding status. ChIP-seq can be used to determine the genomic binding of a defined TF, but it cannot discriminate direct DNA binding from indirect binding. The combination of motif match and ChIP-seq overlap might be the most performing strategy to highlight genomic footprinting data by resolving the assignment of specific TFs to footprints in cases where distinct TFs utilize highly similar binding sites. However, the high frequency of likely indirect binding observed with most TFs, combined with a propensity for artifactual enrichment at highly active loci and poor reflection of binding kinetics, precludes ChIP-seq from serving as a true gold standard for the evaluation of footprinting data [331]. An other important limitation of this approach is the fact that current PWM databases are not complete and a lot of TF consensus recognition sites have not been characterized.

### This study

In our study we developed a mixture model that was able to optimize the footprint width around the motif and subsequently predict bound or unbound sites (see Appendix A.1). For the unbound model, we expressed the likelihood of the measured cuts as a product of independent Poisson variables with a common mean. On the other side, for the bound model we made the simplifying assumption that the shape of a footprint can be approximated by a rectangular shape, showing on average less counts in the protected regions. Therefore, we assumed two distinct means for unprotected, and for protected sites, representing the average number of cuts outside, and inside the footprinted region. We formulated the mixture model by introducing a global probability (that had to be estimated) to be in the bound state. We could then marginalize over this global probability to obtain the likelihood of the whole data with respect to the boundaries. Finally, once we found the optimal boundaries we could estimate the optimal global probability and also assign posterior probabilities to each region for each of the two models. We used DNase I hypersensitive sites and overlapped by ChIP-seq peaks (of BMAL1 in this case) in figure 4.3 containing a well conserved TFBS (FIMO PWM match  $p$ -value  $< 10^{-4}$ ) of interest.

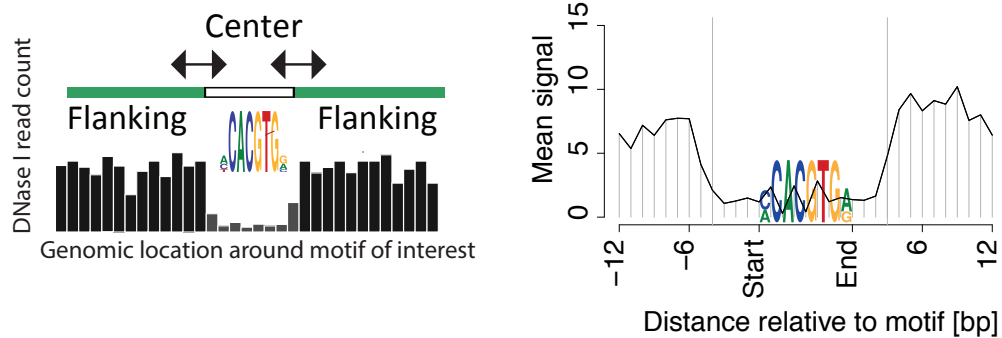


Figure 4.3 – Digital genomic footprinting methods: our method consists of analyzing DNase I tags around a motif of interest (overlapped by some ChIP-seq signal if available), in this case BMAL1/CLOCK E-box motif in DNase I hypersensitive regions. The basic idea is to find regions where read counts are depleted in the center region compared to the flanking regions. We developed a method that maximize the number of bound sites by adjusting the width of the core footprint using a mixture model.

This method was successfully used on several example in the result section 2.1.5. Moreover, we were able to detect a change in shape around double E-box motifs with a spacer of 6-7 bp overlapped by BMAL1 ChIP-seq that suggested a hetero-tetramer binding mode of BMAL1/CLOCK at these sites. This method can be used to study the kinetic of binding on DNA of other protein complexes in the future.

### 4.6.2 *De novo* detection of footprint from DNase I signal

#### State of the art

In the past four years, several algorithms for *de novo* detection of DNase I footprints have been published. A main difficulty of *de novo* footprint detection is that the basic parameters defining a TF-DNA interaction are unknown a priori and must be learned simultaneously with the footprint detection process [331]. Several strategies make use of a sliding window approach for comparing observed cleavage rates in a central region to those within the adjacent flanking sequences and frequently work with parameters, such as footprint and flanking-sequence widths, which are entirely specified during the detection process and parsimoniously selected. In figure 4.4, the Wellington algorithm is based on a similar windowing strategy, enhanced by incorporating information on DNA strands and scores the central window. Wellington perform a binomial test using the cleavage rate within the flanking windows as the expected cleavage rate [256]. Additional methods are now available that make use of probabilistic frameworks such as hidden Markov models (HINT,Boyle) [36, 123] and dynamic Bayesian networks (DBFP) [57] to model per-nucleotide cleavage data.

#### This study

In our DNase hypersensitive sites footprints detection we worked with the Wellington algorithm (pyDNase library) [256] with parameters: `-sh 20,36,5 -fdr 0.05` on all DNase samples concatenated. The DHS containing at least one footprint were subsequently used for our motif activity inference using a penalized linear model (see section 4.7).

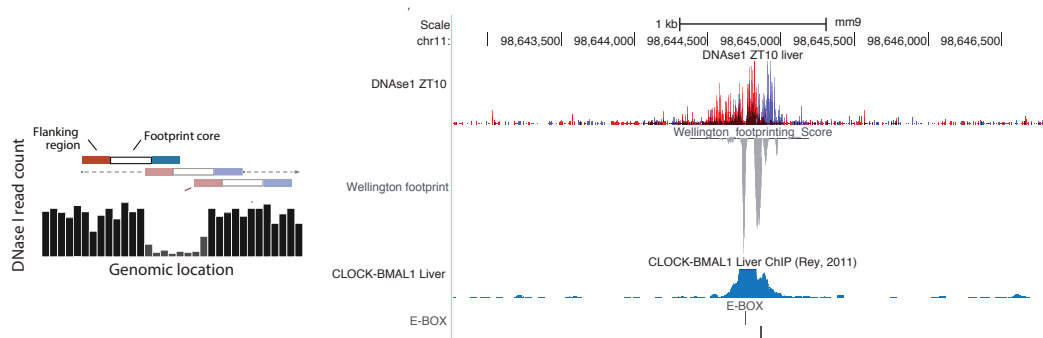


Figure 4.4 – Digital genomic footprinting signal centric method: the Wellington algorithm uses a triple sliding widows to find bound regions, taking into account the forward and the reverse strand reads asymmetry during the footprint detection [256]. The DNase I hypesensitive site is upstream *Nr1d1* gene and contains two E-box motifs overlapped with BMAL1 ChIP-seq. We observed that the Wellington score is maximal at these E-Box motifs.



### 4.6.3 DNase I cutting bias

Despite great potential of detecting the majority of TFBSs in one assay, DNA sequence specific biases, together with transcription factor dependent binding kinetics, have been recently recognized as major confounding factors in DNase-seq experiments [314, 132]. An example is the DNase I sequence cleavage bias, which is due to the different binding affinities of DNase I toward specific DNA sequences. These bias and artifacts are inherent to the experimental protocols used [350]. Another experimental aspect affecting the computational analysis of DNase-seq is the residence time of TF binding. Sung *et al.* showed that short-lived TFs have a lower DNase I cleavage-protection pattern in a TF-specific manner. These influencing factors were not considered by most of the previous methods for the analysis of chromatin accessibility data [212]. He *et al.* suggested that sequence cleavage bias around TFBSs impacts the performance of a computational footprinting method. There is room for improvement in current methodologies by making use of the sequence specificity of each enzyme/assay, including ATAC-seq, but there is no clear consensus in its importance for digital genomic footprinting [212]. However, A contradictory study claims that this cleavage bias of the DNase I doesn't affect much the detection of footprints [184].

In our data (figure 4.5 and table 4.3), we observed a cleavage bias at the 5' end of the sequenced reads with a maximal information content of 0.08 on a scale of 2 bits. the information content of the base frequencies at that position of a motif is computed with the following formula:

$$I_i = 2 + \sum_b f_{b,i} \log_2(f_{b,i}) \quad (4.5)$$

where  $f_{b,i}$  indicates the frequency of base  $b$  at position  $i$ . Positions that are perfectly conserved contain 2 bits of information, those where two of the four bases occur 50% of the time each contain 1 bit, and positions where all four bases occur equally often contain no information [77].

We saw that this bias is slightly different in each time points and that ZT2 and ZT26 cluster together. We can conclude that this bias is protocol and enzyme specific, and that in the case of a time series analysis this bias is partially time point dependent.

Finally a comparative analysis of Different digital genomic footprinting methods showed that the methods that takes this bias into account with a 6-mers correction of the signal around the cleavage position performed slightly better than the other methods (DNase2TE, HINT). Interestingly, the wellington algorithm was the best performing method that didn't applied a bias correction [122], which confirm the validity of our analysis.

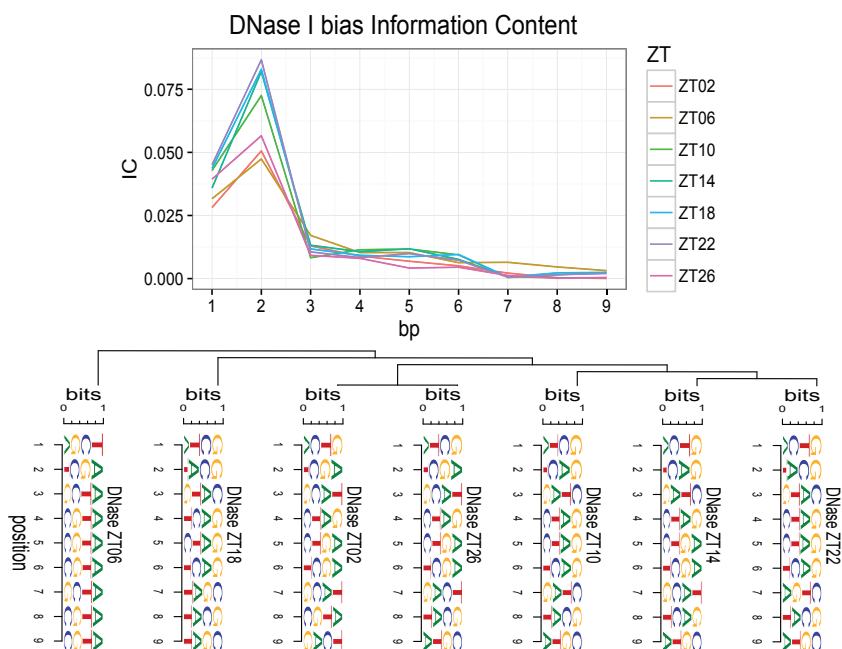


Figure 4.5 – Information Content per base pairs at the 5’end of our DNase I-seq libraries. This analysis was done using the FastQC report of our DNase I-seq libraries and SeqLogo

Table 4.3 – DNase I cutting bias in our DNase I libraries. The frequency of the nine fist positions of our reads are displayed in %.

Dnase I library	Base	Position 1	Position 2	Position 3	Position 4	Position 5	Position 6	Position 7	Position 8	Position 9
ZT02	A	16.85	30.30	27.14	27.38	28.36	27.90	25.93	25.74	24.95
ZT02	C	26.80	26.36	25.02	21.05	21.66	24.47	24.62	24.22	24.98
ZT02	G	28.26	28.93	19.66	27.77	25.82	25.54	22.95	24.88	24.72
ZT02	T	28.09	14.41	28.18	23.80	24.17	22.09	26.49	25.16	25.35
ZT06	A	16.73	31.15	29.01	28.72	28.84	28.99	27.52	27.43	26.69
ZT06	C	26.74	25.93	24.14	20.44	20.56	22.81	23.09	22.81	23.29
ZT06	G	26.05	28.02	19.21	25.33	24.77	23.84	22.23	23.26	23.41
ZT06	T	30.48	14.90	27.65	25.51	25.83	24.35	27.16	26.50	26.62
ZT10	A	16.12	29.09	25.84	26.93	27.34	26.77	24.59	24.27	23.42
ZT10	C	27.62	27.83	27.30	20.87	21.23	25.53	25.95	26.07	26.54
ZT10	G	32.41	30.74	20.55	28.94	28.86	27.40	24.37	26.11	26.24
ZT10	T	23.86	12.35	26.31	23.26	22.56	20.30	25.10	23.54	23.80
ZT14	A	15.98	29.68	26.80	27.62	28.07	27.54	25.32	25.13	24.31
ZT14	C	26.32	27.98	27.01	20.68	20.59	24.41	25.17	25.09	25.70
ZT14	G	29.62	30.73	19.30	28.01	27.97	26.95	23.66	25.60	25.38
ZT14	T	28.08	11.61	26.88	23.69	23.37	21.10	25.85	24.18	24.60
ZT18	A	14.94	28.34	26.49	26.41	26.76	26.49	24.66	24.44	23.82
ZT18	C	28.57	28.44	28.51	22.42	22.02	25.43	26.21	25.80	26.65
ZT18	G	28.69	31.62	20.01	28.97	28.55	27.76	24.80	26.73	26.23
ZT18	T	27.80	11.59	24.98	22.20	22.66	20.32	24.33	23.03	23.31
ZT22	A	14.88	28.44	25.98	26.72	26.57	26.30	24.30	24.11	23.45
ZT22	C	27.41	28.98	28.14	21.44	21.45	25.33	26.09	26.12	26.58
ZT22	G	30.01	31.30	20.11	28.29	28.94	27.44	24.56	26.16	26.02
ZT22	T	27.71	11.28	25.77	23.55	23.04	20.92	25.05	23.61	23.95
ZT26	A	15.72	30.40	26.40	26.52	27.09	26.98	24.93	24.92	24.10
ZT26	C	27.48	26.43	25.67	21.50	22.42	25.55	25.65	25.09	25.75
ZT26	G	30.65	29.34	20.35	28.36	26.52	25.71	23.36	25.44	25.33
ZT26	T	26.15	13.83	27.59	23.63	23.96	21.76	26.05	24.55	24.81

## 4.7 Linear model to infer transcription factor binding motif temporal activity

The predominant theory for circadian transcription regulation is the phase vector model, whereby a new phase results from the combinatorial synthesis of two or more transcriptional regulators or two clock-controlled DNA element [326]. In our circadian context, we aimed at inferring the phase of each TF binding on DHS using Pol II loadings at TSS of active genes and the motif content of DHS nearby (figure 4.6).

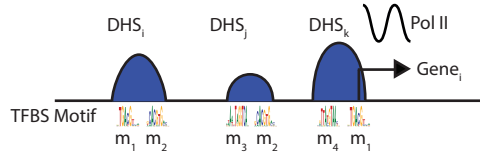


Figure 4.6 – The PolII signal at TSS is used with the motif content of nearby DHSs in order to compute the phase specific motif activity.

We fitted the following linear model to infer motif activities  $A_m$  :

$$Y_g = \sum_m N_{gm} A_m + \epsilon \quad (4.6)$$

Here  $Y_g$  is a complex amplitude summarizing the oscillation with a period of 24 h, i.e.  $Y_g = \sum_t e^{i\omega t} Y_t$ ,  $A_m$  is the corresponding complex motif activity,  $N_{gm}$  is a regression matrix defined below and  $\epsilon$  is a noise term. In practice we can represent the complex numbers  $Y_g$  and  $A_m$  as two-dimensional vectors.

For this regression a matrix  $N_{gm}$ , was built which contains motif  $m$  for every DHS  $d$  in a given vicinity of active gene  $g$ . More precisely, we simply sum the motif content of DHSs in a given proximity of a gene. To obtain this  $N_{gm}$  matrix, FIMO [114] was used to scan our DHSs sequences and different sets of position weight matrices (PWM), such as JASPAR [258], TRANSFAC [217], SELEX [156] and WANG [338] with a detection threshold  $p$ -value below  $10^{-4}$ . We used a set of approximately 1900 PWM.

The maximum likelihood estimates of unknown activities  $\hat{A}_m$  can be computed as follow using least squares method:

$$\hat{A}_m = \left( N_{gm}^T N_{gm} \right)^{-1} N_{gm}^T Y_g \quad (4.7)$$

However, to avoid overfitting and control for the redundant motifs, we used a penalized linear

regression model developed by T. Hastie [100]. This method is available as an R package called GLMNET and uses the elastic-net penalized regression. This penalized linear regression model is defined as follows:

$$\hat{\beta}_{\alpha,\lambda} = \arg \min_{\beta} \left[ \|y - X\beta\|_2^2 + \lambda \left( \alpha \|\beta\|_2^2 + (1 - \alpha) \|\beta\|_1 \right) \right] \quad (4.8)$$

where we have switch to the standard notation for linear regression, i.e. using  $X$  for the design matrix,  $y$  for the data, and  $\beta$  for the regression variable.

The elastic net regression is a method that allows regularization and variable selection simultaneously. It includes the Lasso penalty  $l_1 = \|\beta\|_1$  that ensure the sparsity of the solution. The caveat of Lasso penalty implies that the solution is bounded by the number of 'samples' and that it tends to select one variable from a group and ignore the others. To overcome these limitations the elastic-net includes a second part of the penalty. it's the quadratic term  $l_2 = \|\beta\|_2^2$  called the Ridge penalty.  $\alpha$  is for the elastic-net mixing parameter with range  $\alpha \in (0, 1)$ , where  $\alpha = 1$  is the lasso and  $\alpha = 0$  is the ridge.  $\lambda$  is a vector of values, which allows to progressively decrease the penalty in the "Pathwise Coordinate Descent algorithm" [100].

We tested several other motifs databases as well, like HOCOMOCO [181], Swiss Regulon [246] or JaspAr 2016 [215], and we observed that we had similar predictions for E-Box motifs, RRE-motifs or D-box. A major conclusion of this database comparison is that motifs quality of the database, and the redundancy affect slightly the analysis. A data cleaning effort improve drastically subsequent interpretation of the biological significance of the outcome of the model (figure 4.7). Briefly, we clustered the PWMs based on their matches on DHSs and we selected for each representative cluster the PWM with the highest circadian activity using our penalized linear model with 1900 PWMs (Table 4.4).

#### 4.7. Linear model to infer transcription factor binding motif temporal activity

Table 4.4 – Selected non-redundant PWMs from different sources

Original ID	Clean ID	Source
V_E4BP4_01	D_Box	Transfac
V_RORA2_01	RRE	Transfac
V_CREBATF_Q6	CREB	Transfac
BHLHE40	E_Box	Wang
V_AR_Q6_01	GR	Transfac
V_FOXO1_Q5	FOX	Transfac
SP1	SP1	Wang
V_ZF5_01	ZF5	Transfac
V_PPARG_01	PPAR	Transfac
V_AP2_Q3	AP2	Transfac
V_ZIC2_04	ZIC2	Transfac
V_PU1_Q4	PU1	Transfac
V_RXRA_03	RXR	Transfac
V_ERR1_Q2_01	ERE	Transfac
V_MITF_Q6	MITF	Transfac
HNF1A_M161	HNF1	Selex
V_NFY_01	NFY	Transfac
V_SREBP1_02	SREB	Transfac
V_GABPAGABPB_Q6	GABP	Transfac
Jdp2_M131	JDP2	Selex
V_SMAD_Q6	SMAD	Transfac
V_HNF4A_02	HNF4	Transfac
V_NKX62_Q2	NKX6	Transfac
V_EKLF_Q5	KLF	Transfac
V_GATA3_05	GATA3	Transfac
V_OCT1_01	OCT1	Transfac
V_LXRDR4_Q3	LXR	Transfac
V_MAF_Q6_01	MAF	Transfac
HSF1_M61	HSF	Selex
V_MEF2A_Q6	MEF2A	Transfac
V_FXR_Q2	FXR	Transfac
V_MTF1_02	MTF1	Transfac
EGR1_M3	EGR1	Selex
MA0080.2	ETS	Jaspar
V_CEBP_C	CEBP	Transfac
V_GFI1B_01	GFI1	Transfac
NFIA_M238	NFIA	Selex
V_E2F_02	E2F	Transfac
V_MYB_Q6	MYB	Transfac
V_YY1_01	YY1	Transfac
GLI2_M4	GLI2	Selex
NFAT5_M80	NFAT5	Selex
V_ACAAT_B	SOX	Transfac
V_TBP_01	TBP	Transfac
TEAD1_M113	TEAD	Selex
RFX5	RFX	Wang
SRF	SRF	Wang

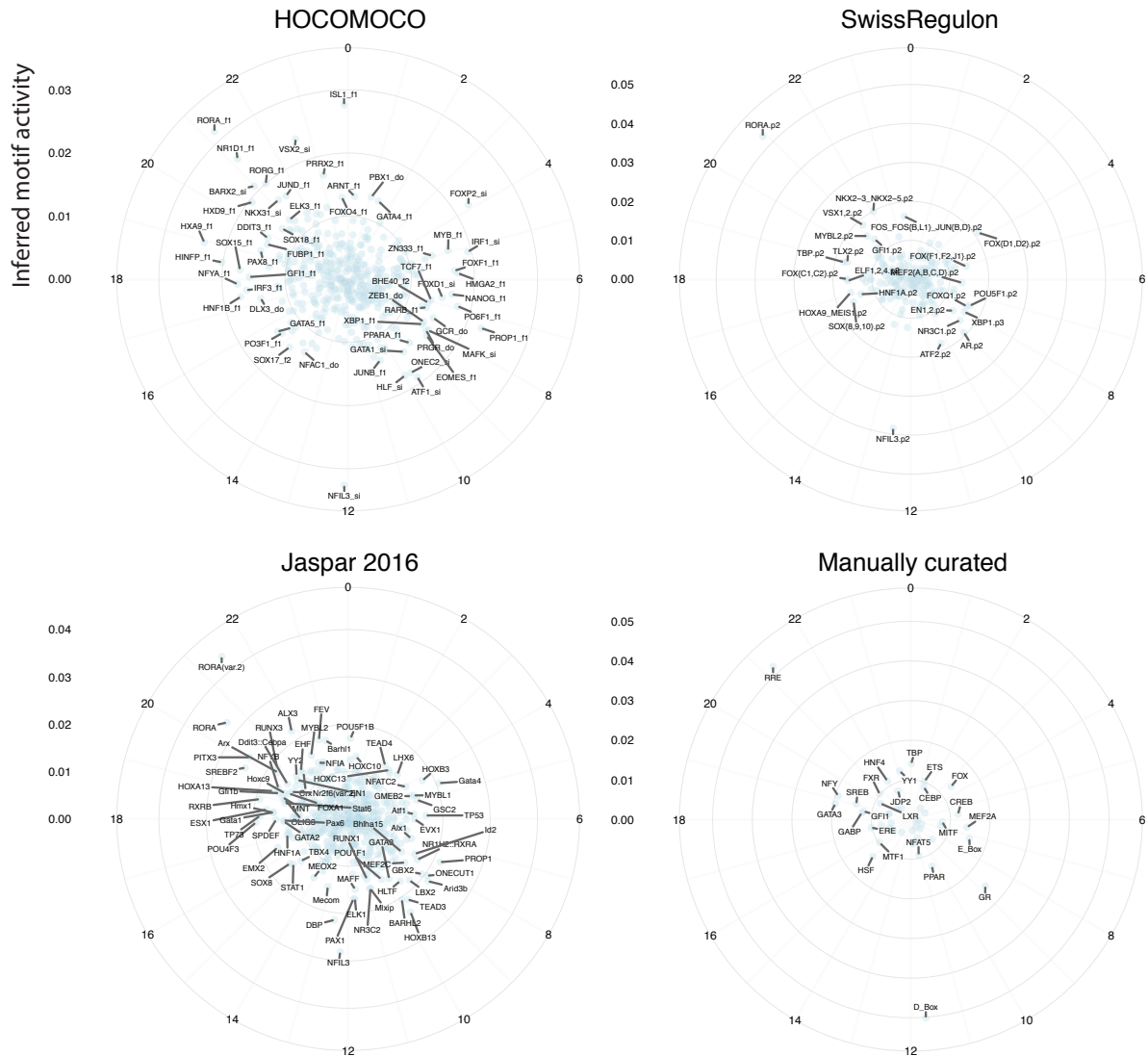


Figure 4.7 – Penalized linear model with four different motif databases. Pol II loading at TSS of active genes (in WT mice) was fitted with DHS motif content in a vicinity of 50 Kb. The parameters used for these penalized linear model were  $\lambda = 0.1$  and  $\alpha = 0$

We optimized our linear model to estimate the  $\hat{\beta}$  (in our notation the unknown activities estimates  $\hat{A}_m$ ), by doing a parameter space exploration in figure 4.8, and we found that an  $\alpha$  of 0.1 was performing well. We tested our model with the motif occurrence in the  $N_{gm}$  matrix, or a boolean representing the presence or the absence of the motif, and we observed that the motif occurrence had a higher deviance ratio (equivalent to the variance explained by the linear model). Moreover, the Glmnet package has a k-fold cross validation implemented. We used a 10-fold cross-validation for our final set of motifs that explain a part of the circadian oscillation in our data.

#### 4.7. Linear model to infer transcription factor binding motif temporal activity

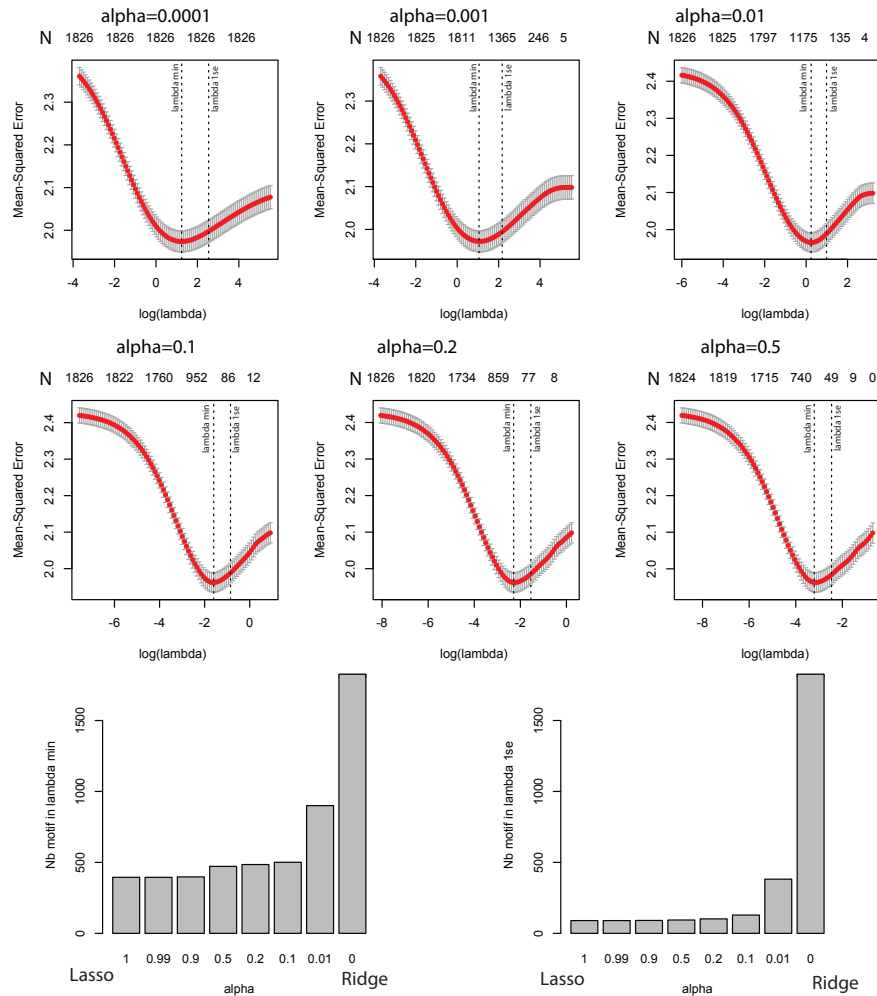


Figure 4.8 – The elastic-net has been performed with various  $\alpha$  and  $\lambda$  in order to determine the optimal parameters that will select a minimal set of motif with a minimum mean-squared error

In summary, our penalized linear model was accomplished using the 1900 PWM on a common selection active genes detected using a SVM (see Section 4.8 of DHS to active transcripts) using Pol II loading at TSS in the WT or *Bmal1*<sup>-/-</sup> context. Phase and amplitude of Pol II loadings at TSS was fitted with the motif matrix to infer motif activity in WT and *Bmal1*<sup>-/-</sup>. DHS motif content was included in  $N_{gm}$  matrix within a given proximity (distance threshold) with their closest active genes with one or more footprints, detected using Wellington algorithm from pyDNase Library [256].

### 4.8 Annotation of DHS to active transcripts and enrichment analysis

Using ChIP-seq data for Pol II, H3K4me3, H3K36me3 and H3K27ac from CyclIX [188] in the WT condition, a support vector machine classifier (SVM) was developed by Irina Krier to detect active transcripts among all Ensembl annotated transcript (version NCBIM37). The goal was to use a classifier that would recognize the shape of signals typical of active transcription [27] based on histone modifications, Pol II and DNase I. We selected regions of interests to be +/- 300 bp around the TSS for Pol II and H3K4me3, and also +/- 300 bp around the TES for DNase I, and the last 600 bp of each transcript for H3K36me3. Read counts on the same strand as the transcript annotation were counted per 10 bp and quantile-normalized across time. To correct for relative difference in signal between marks, correction coefficients were applied based on the sum of each signal. A set of active and inactive transcripts were extracted consisting in the top 10% and bottom 10% respectively, as determined by Pol II RPKM along each transcript. Our SVM was trained on these active versus inactive transcripts, and subsequently applied to all transcripts at each time point. Cross-validation indicated that the SVM had satisfactory False Positive and False Negative results for very high or very low Pol II signals (98% of test transcripts were correctly classified either active or inactive at ZT10). Classifier predictions for all time points were very similar, with more than 80% correlation. Transcripts shorter than 600bp were set to “active” if they had higher Pol II RPKM than the lower quartile of active transcripts. Transcripts were considered active when they were classified as active in at least one time point. This classification resulted in 51'947 out of the 97'214 transcripts to be classified as active (17'467 out of 37'583 genes). The active transcripts list (51'947 transcripts) was used to associate DHS with the closest active transcription start site (TSS) using the Chippeakanno R package [359] from Bioconductor. The annotation result provided 13'457 unique active genes linked with at least one DHS. These active genes were subsequently used in our gene centric analysis with our penalized linear regression. Basically a DHS could regulate all active genes that are in a certain proximity defined by our so called distance threshold (see section 4.7). Finally, we performed pathway enrichment analysis using gprofileR [269] on rhythmic mRNA in WT and/or in *Bmal1*<sup>-/-</sup> genotype. We then used amplitudes of genes annotated in a given significant pathway for downstream analysis that revealed important changes in the clock impaired genotype.

### 4.9 3D structure of the heterotetramer BMAL1/CLOCK and molecular dynamics

For the single BMAL1/CLOCK, the crystal structure of the heterodimeric BMAL1/CLOCK (pdb id: 4F3L) was used as an initial model [141]. In this structure there are 5 flexible loops lacking density. The residues in positions 129-134 (length 6 residues), 212-237 (26 residues), 257-275 (19 residues), 291-309 (19 residues) were missing from BMAL1, and the residues 224-247 (24 residues) were missing from CLOCK. These missing parts were computed by Rosetta's loop



#### 4.9. 3D structure of the heterotetramer BMAL1/CLOCK and molecular dynamics

---

modeling application (v3.5); an application that extensively remodels the backbone of the loops [261]. The loops were remodeled and refined by the CCD (Cyclic Coordinate Descent) algorithm [49]. The fragment files, used by CCD were made by Robetta Server [165]. The BMAL1/CLOCK structure, as a unique chain, was used as Rosetta input and from the output we selected the lowest energy loops for the single BMAL1/CLOCK model. In order to bind the single BMAL1/CLOCK model to the E-box, the complex crystal structure of BMAL1/CLOCK basic helix-loop-helix domains bonded on the E-box (CACGTG) (pdb id: 4H10) was used [339]. This structure was superimposed to the single BMAL1/CLOCK model with the UCSF Chimera visualization program (v1.5.3) [253]. In accordance to this super-position the single BMAL1/CLOCK model the N-terminal helices of CLOCK and BMAL1 was replaced by the helices in the 4H10 structure from the protein data bank. The base-pair geometry of the DNA in the 4H10 structure was analyzed by the 3DNA software (v2.0) [205]. Two double-strand DNA models, spacing 6 (sp6) and spacing 7 (sp7), with sequence 5'-CACGTGAAAAAA(A)CACGTG-3', were generated by 3DNA. The CACGTG parts were rebuilding based on the analysis of the DNA in the 4H10 structure. The interval base pairs of sp6 were building with the standard B-DNA backbone conformation for A-T pairs. For the final models two BMAL1/CLOCK models were bound to the DNA models with a spacer of 6 bp (sp6) or 7bp (sp7), by superimposing them with UCSF Chimera.

On the 2 heterotetramer models, molecular dynamics simulations were performed with the NAMD program (v2.8) [255] in order to explore the various structural conformations. Each of the models was placed in an explicit solvent box with 0.15 M NaCl concentration under periodic boundary conditions. The models were parameterized using the AMBER force field (ff99bsc0) and the TIP3P model was used for water molecules (LeaP program) [52]. Initially, geometry optimization via energy minimization to refine DNA backbone and base pairing geometry was performed for 12500 steps, followed by an NPT phase of 90ns and 130ns for the sp6 and sp7 model respectively. In all phases a time step of 1 fs was used, the covalent bonds involving hydrogen atoms were constrained by the RATTLE algorithm [11] and the Van der Waals interaction cutoff distances was set at 12 Å.



# A Appendix

## A.1 Mixture model for DNase I footprint detection

*This mixture model for DNase I data footprint detection around transcription factor binding sites has been developed by Felix Naef. I was implicated in the development of the visual output as well as the parameter optimization and the analysis of different TFBS, such as BMAL1, CREB, SREB, FOXA2 HSF1, and REVERB $\alpha$*

For this model, we assume to have DNase I cuts  $n_{ri} \in \mathbb{N}$  for regions  $r \in \{1, \dots, R\}$  at positions  $i \in \{1, \dots, L\}$ , where  $R$  is the number of regions and  $L$  the length of the regions (all assumed to be of the same length). Indices  $i_L$  and  $i_R$  refer to boundary positions (the left and right boundaries of a protected sequence motif) and we assume that the cuts are all aligned with respect to some anchor, for example the position of a sequence motif reflecting the specificity of a known protein-DNA interaction. We define  $J = \{i_L, \dots, i_R\}$  and  $I = \{1, \dots, L\} \setminus J$ .

The goal is to calibrate a mixture model from the data, such that we can learn which sites are bound (showing footprints), and which are not. While similar approaches have been proposed [257], we are interested in additionally learning the optimal boundary positions, such that we can detect if a footprint changes shape (in our case its width) in different conditions (for example at different time points). To make the model tractable, we make the simplifying assumption that the shape of a footprint can be approximated by a rectangular shape, showing on average less counts in the protected regions. While it was shown previously that the signal within the protected region can be nonuniform for some factors [237, 320, 338], the rectangular model is a simple generic model that captures the essential properties of DNase I signals around bound sites in many cases. Specifically, we express the probability (likelihood) of the measured cuts  $\vec{n} = (n_1, \dots, n_L)$  (here for a single region or one row in the matrix  $n_{ri}$ ) as a product of independent Poisson variables with a common mean  $\lambda$  when the site is not bound

$$P_1(\vec{n}|\vec{\lambda}) = \prod_{i=1}^L \text{Pois}(n_i|\lambda) = \mathcal{M}(\vec{n}|\vec{p}, N) \text{Pois}(N|\Lambda) \equiv Q_1(\vec{n}; \Lambda) \quad , \quad (\text{A.1})$$

## Appendix A. Appendix

---

where we used the property that products of independent Poisson variable can be factored into a multinomial ( $\mathcal{M}$ ) and one Poisson distribution, and defined  $N = \sum_i n_i$ ,  $\Lambda = \sum_i \lambda_i = L\lambda$ ,  $p_i = \frac{1}{L}$ . The notation  $Q_1$  emphasises that this is now a function of  $\Lambda$ .

For the second (bound) model we assume two distinct means  $\lambda_I$  for unprotected, and  $\lambda_J$  for protected sites, representing the average number of cuts outside ( $i \in I$ ), and inside ( $j \in J$ ) the footprinted region, respectively. This leads to

$$P_2(\vec{n}|\lambda_I, \lambda_J, i_L, i_R) = \prod_{i \in I} P(n_i|\lambda_I) \prod_{j \in J} P(n_j|\lambda_J) \quad (\text{A.2})$$

$$= \mathcal{M}(\vec{n}|\vec{q}, N) \text{Pois}(N|\Lambda) \equiv Q_2(\vec{n}; i_L, i_R, q_J, \Lambda) \quad , \quad (\text{A.3})$$

where here  $\Lambda = \sum_i \lambda_i = L_1\lambda_I + L_2\lambda_J$  with  $L_1 = |I|$ ,  $L_2 = |J|$ ,  $q_i = \lambda_I/\Lambda \equiv q_I$  for  $i \in I$  and  $q_i = \lambda_J/\Lambda \equiv q_J$  for  $i \in J$ . The notation  $Q_2$  shows the dependencies in the new variables.

We then marginalize the probabilities over the unknown  $\Lambda$  (using an improper flat prior, such that  $\int_0^\infty \text{Pois}(N|\Lambda) d\Lambda = P(N) = 1$ , and thus equivalent to making no assumption on the total number of cuts). After some straightforward algebra, this leads to

$$F_1(\vec{n}) = \int_0^\infty d\Lambda Q_1(\vec{n}; \Lambda) = \frac{N!}{\prod_i (n_i!)} L^{-N} \quad (\text{A.4})$$

where  $N = \sum_i n_i$ , and

$$F_2(\vec{n}|i_L, i_R) = L \int_0^{1/L} dq_J \int d\Lambda Q_2(\vec{n}; i_L, i_R, q_J, \Lambda) \quad (\text{A.5})$$

$$= \frac{1}{N+1} \frac{N_1!}{\prod_{i \in I} (n_i!)} \frac{N_2!}{\prod_{j \in J} (n_j!)} L_1^{-N_1} L_2^{-N_2} \frac{1}{r} I_r(N_2+1, N_1+1) \quad (\text{A.6})$$

where  $N_1 = \sum_{i \in I} n_i$ ,  $N_2 = \sum_{j \in J} n_j$ ,  $r = \frac{L_2}{L}$ , and  $I_r(\alpha, \beta)$  is the regularized incomplete Beta function ( $I_1(\alpha, \beta) = 1$ ) that comes from the  $q_J$  integral. Note that since the same improper (not normalized) prior on  $\Lambda$  is used for both models, this does not pose any difficulties. The upper integration bound on the  $q_J$  integral uses the assumption that  $q_J \leq q_I \leq 1/L$ , reflecting that the probability of cuts is reduced inside  $J$  due to protection from the bound protein.

We can now formulate the mixture model by introducing a global probability  $q$  (to be esti-

## A.1. Mixture model for DNase I footprint detection

---

mated) to be in the bound state, such that

$$P(\vec{n}|i_L, i_R, q) = (1 - q)F_1(\vec{n}) + qF_2(\vec{n}|i_L, i_R) \quad , \quad (\text{A.7})$$

or since  $q$  is assumed to be common to all regions

$$P(\{n_{ri}\}|i_L, i_R, q) = \prod_r P(n_{r\bullet}|i_L, i_R, q) \quad . \quad (\text{A.8})$$

Finally we can marginalize over  $q$  to obtain the likelihood of the whole data with respect to the indices  $I$ :

$$P(\{n_{ri}\}|i_L, i_R) = \int_0^1 dq P(\{n_{ri}\}|i_L, i_R, q) \quad , \quad (\text{A.9})$$

where we have assumed a uniform prior on  $q$ . The interesting aspect is that this likelihood calculation requires only doing a one-dimensional numerical integral, and one can then maximize with respect to the discrete indices  $i_L$  and  $i_R$  to find the optimal boundaries. Once these have been found, it is straightforward to estimate the optimal  $q$  using Eq. 8, and also to assign posterior probabilities to each region for each of the two models using Eqs. 4 and 6.

## **A.2 *Circadian life, a smartphone app to monitor our circadian activity***

### **A.2.1 Introduction**

Since the dawn of time, the earth rotation on its axis and around the sun imposes to all living creatures daily changes and seasonal variations. These changes imply light-dark cycles, temperatures changes, and food availability or predators presence. In an evolutionary scale, all living creatures needed to anticipate those changes in order to develop an efficient survival strategy. In humans, our internal clock controls several behaviors, such as sleep-wake cycle, food intake, hormonal cycle, and immunity. Despite this circadian control, humans often use alarm clocks and medication to align their sleep and wake times with social obligations (e.g., work and school schedules or other social events) [281]. Activity varies between work and free days due to a changing in sleep duration but also in sleep timing. In order to recover from sleep deficiency accumulated over the working period, people commonly oversleep on free days [281]. Current methods to monitor human activity are subjective (such as questionnaires), providing only a general overview of daytime activity [282]. The availability of smartphones represents an opportunity to monitor human activity objectively. Moreover, the wide-ranging adoption of smartphones across age, gender, and socioeconomic segments is critical to study the behaviors of free-living representative individuals at a population scale [336, 112]. Thus, it is essential to understand our rhythm on a daily basis for increasing our ability to live in a healthy state. Recent breakthroughs of chronobiology are significant in chronotherapy – e.g., when should one take a particular drug to cure a disease more efficiently [178] – and in chrononutrition, as when should one eat to stay in a healthy state [112].

### **A.2.2 Goal of the project**

We aimed at gathering data on daily activity of users and had access to various information such as their locations, ages, gender and more, thanks to our smartphone app, called *Circadian Life*. Using these data, we will be able to study different chronobiology-related features including jet-lag, chronotype, light entrainment (seasonal changes, effect of latitude) and social entrainment at the population level.

In collaboration with Jonathan Baeriswyl (a "civiliste" in the Naef lab), we developed the first version of *Circadian Life*, a smartphone app to record the activity of users at a population scale. The goal of the app is to record user activity through phone sensors (e.g., accelerometer, light and GPS) and possibly wearable gadgets, smart watch or wristband, in an automated manner. Users will have access to their personal data via a web interface and through the app interface (figure A.1). In addition, surveys will be designed in a progressive manner, starting with a few simple questions, the app will ask further questions as users are better engaged.

## A.2. *Circadian life*, a smartphone app

to monitor our circadian activity

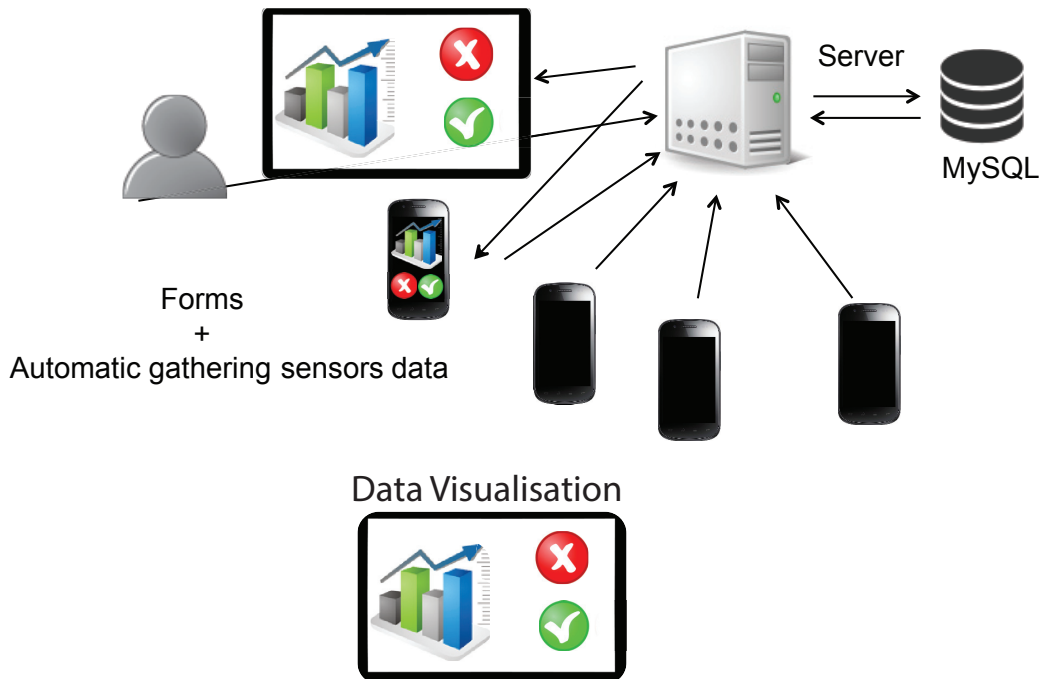


Figure A.1 – The app structure includes an SQL database and a server storing the results of surveys and sensor from smartphones. Users can have access to their data thanks to their phone or a website.

### A.2.3 Preliminary results

The *Circadian Life* app is based on the recording of sampled activity using the accelerometer of the phone. The measures are sampled each five minutes (one minute of recording), and an average of those measures is calculated for each hour. Data Cleansing starts with the subtraction of the gravity acceleration detected by the accelerometer, to all the measures recorded. Thus, the cleaned data provide us the activity of the user per hour.

In a first experiment, 36 users were recruited in November 2015, mostly students from Felix Naef class. Several participants were recorded for more than three months. This allowed us to explore several visualizations and statistics to analyze users' data.

A first visualization of the data using actograms revealed that users were well entrained to the daylight cycle (figure A.2), with an expected period of 24h computed with a Fourier analysis. Indeed, we did not observe a drift of the active period as reported in the FASPS subjects [260].

As it was reported, activity varies between work and free days due to a changing in sleep duration but also in sleep timing [281]. Therefore, we wanted to quantify these differences using the weekday and the weekends averaged data from a single user over the whole recording (figure A.3). With these averaged activity profiles, we quantified the difference between work days and free days. Interestingly, work day activity pattern reflected habits of the user. Thus,

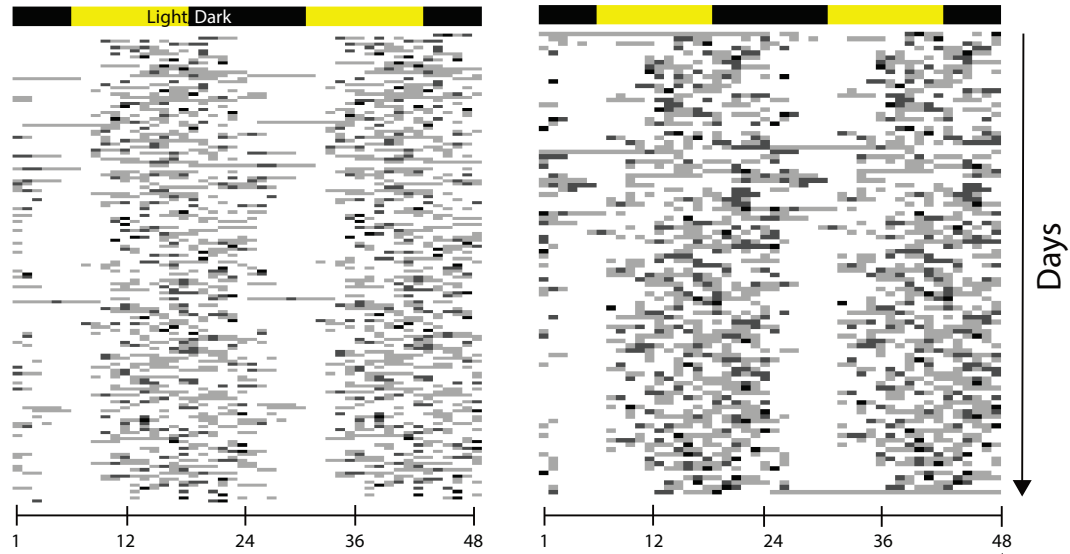


Figure A.2 – Double-plotted actograms of two users of "Circadian life" recorded for several months.

we observed three peaks of activity during the workdays (arrival/departure from work and lunch break).

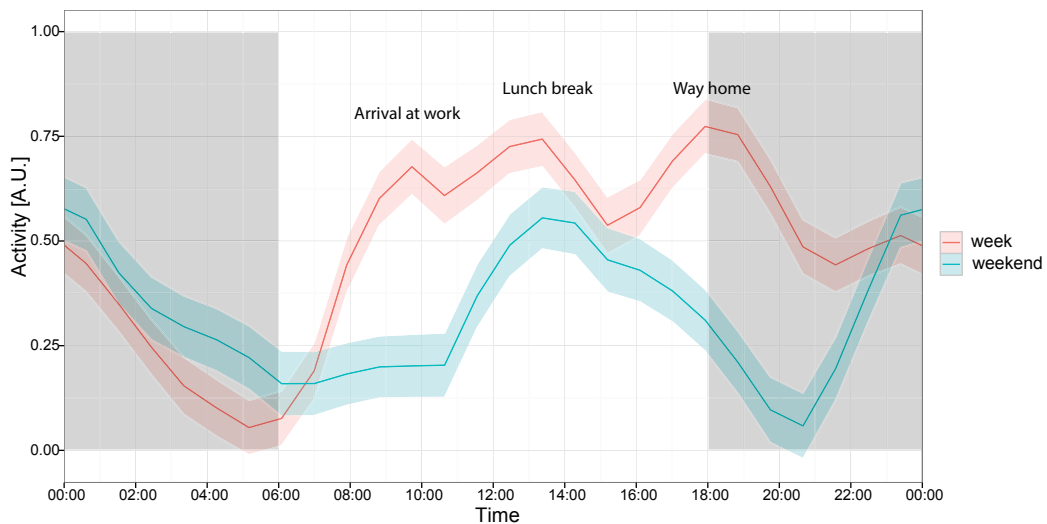


Figure A.3 – Average activity was computed for weeks and weekends separately for a single user of our smartphone app. Using this type of visualization, we can observe a pattern in the signal as the arrival and the departure from work as well as the lunch break. Moreover, the analysis of work days and free days allows the quantification of the social jet-lag.



## A.2. *Circadian life, a smartphone app* to monitor our circadian activity

---

### A.2.4 Sleep-wake cycle analysis

Sleep-wake cycle is critical for a healthy state of life. Therefore, we developed in collaboration with Etienne Dubois (a bachelor student who completed an internship in the Naef lab) a Hidden Markov Model (HMM) to extract, from the activity data, sleep-wake measurements as the sleep length, the activity length, the wake-up time and the bedtime (the time at which the user goes to bed).

The data was analyzed using a two-state HMM, that allowed finding a pattern of activity/inactivity (the hidden states) using the binarized recorded data (figure A.4 A). We performed the Viterbi algorithm to find the most probable path through a two state model (active/inactive) on the activity (figure A.4 B). We observed that our minimal model with two states could find most of the active/inactive periods. Thanks to our HMM, we noted that the distribution of wake-up time of the user had a single mode at 8 AM, and the distribution of bedtime was bimodal with one peak at 10 PM and another at 1 AM (figure A.4 C). On the other side, the distribution of sleep length had a mode of 10 hours of sleep. In addition, the distribution of activity length had a mode at 14h. The distribution of activity revealed that some sleep periods are missed by our HMM. Thus, we observed that 5% of actives periods are longer than 30 hours.

In summary, our app allowed a decent monitoring of the users that have participated in our study. Further explorations of the data are needed in order to evaluate the discrepancies between different users from a biological perspective as well as from a technical point of view. Furthermore, we have to assess differences due to various phone sensors and phone brand and we should compare our measurements to wearable in order to improve our data sampling and calibration, in addition, minimizing battery consumption.

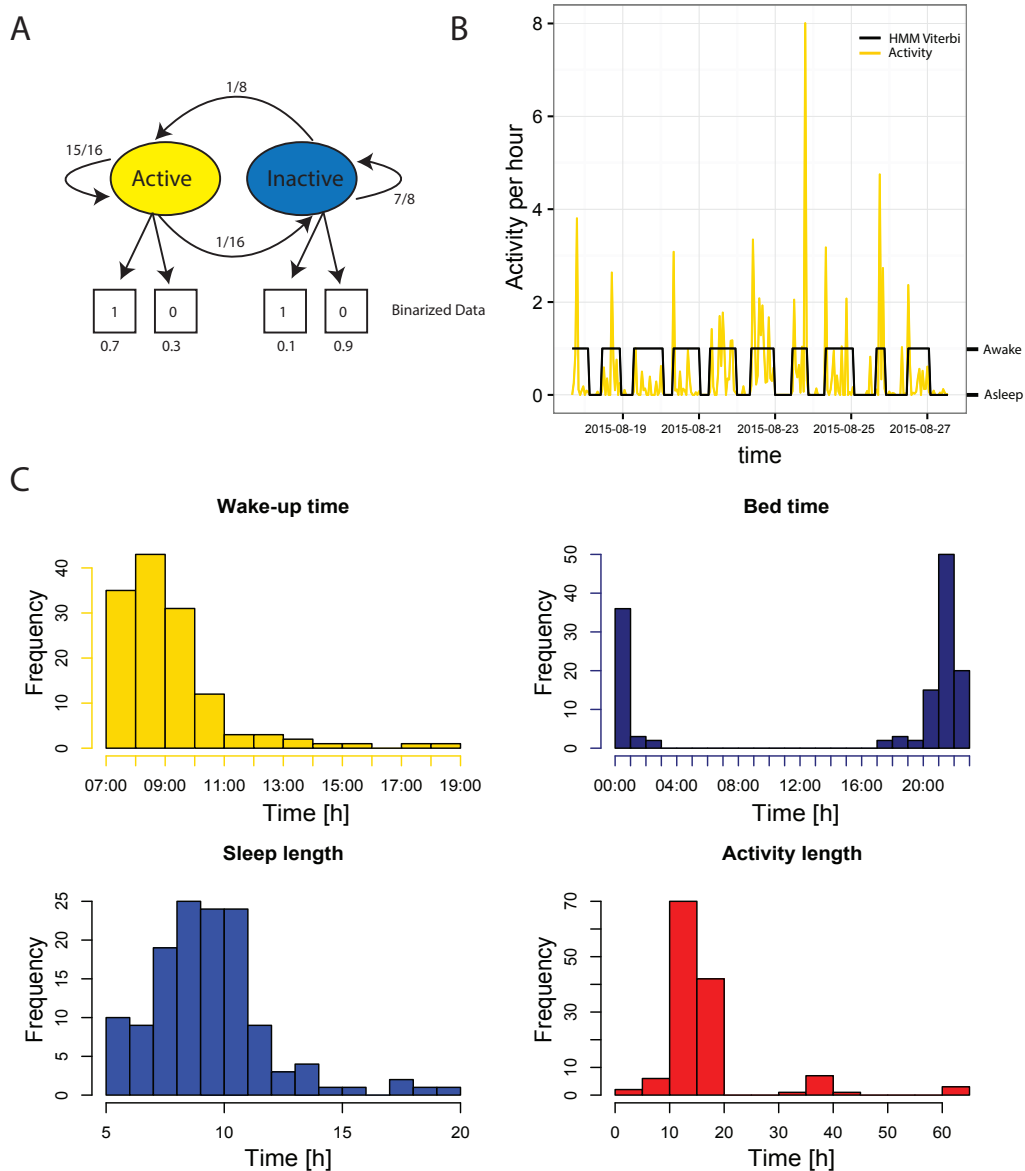


Figure A.4 – HMM for Smartphone app data analysis. **A** The two state model with the transition and the emission probabilities. **B** Viterbi path overlapped on activity data. **C** Distributions of sleep length, activity length, wake-up time and bedtime estimated from the Viterbi path of a single user.

## A.2. Circadian life, a smartphone app to monitor our circadian activity

### A.2.5 Perspectives

This project of smartphone app will continue with a new collaboration with the HES-SO Valais. Alexis Murciano (a student supervised by Dr. Pierre-Andre Mudry) is already working on the second version of the app. His goal is to produce a user-friendly and researcher-friendly interface that allows more flexibility for the app content (figure A.5). We will include some educational material about the circadian rhythm, such as information on mechanisms, hormonal regulation, structure of the clock and more. We will have as well a better control on survey inside of the app, and we will add new functionalities. A second round of experiments on students will start in October 2016.

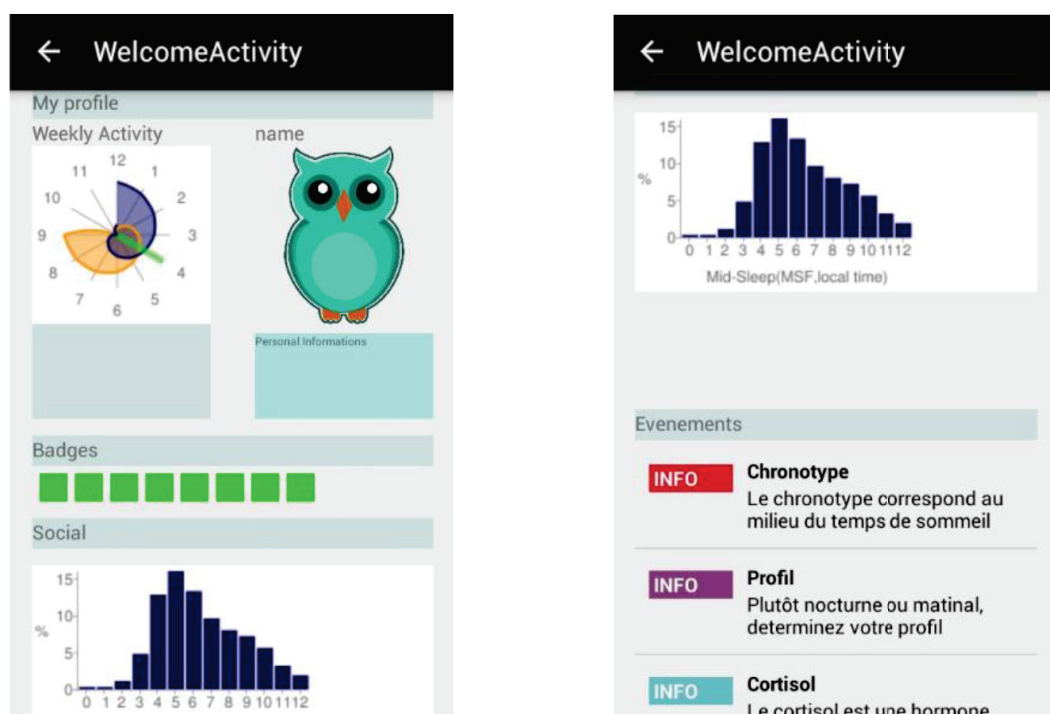


Figure A.5 – User-friendly app interface developed in collaboration with Alexis Murciano from the HESSO Valais under the supervision of Dr Pierre-Andre Mudry.

This app might serve later on for chrono-therapeutical studies on drugs (side effects and pharmaco-vigilance) or chrono-nutrition studies depending on the add-on that we will be able to develop. We might be able in the future to use these activity data integrated with other medical data. For instance, heartbeats and blood glucose level, or even biological data such as genome sequences, epigenetic marks, chromatin accessibility, protein and mRNA levels from users could be integrated in order to improve diagnosis and prognosis and anticipate diseases or seizure. The app market is growing exponentially, and health-related applications are necessary to improve the well-being of individuals.



## Bibliography

- [1] Y. Adamovich, L. Rousso-Noori, Z. Zwihaft, A. Neufeld-Cohen, M. Golik, J. Kraut-Cohen, M. Wang, X. Han, and G. Asher. Circadian clocks and feeding time regulate the oscillations and levels of hepatic triglycerides. *Cell Metabolism*, 19(2):319–330, 2014.
- [2] E. Afgan, D. Baker, M. van den Beek, D. Blankenberg, D. Bouvier, M. Čech, J. Chilton, D. Clements, N. Coraor, C. Eberhard, B. Grüning, A. Guerler, J. Hillman-Jackson, G. V. Kuster, E. Rasche, N. Soranzo, N. Turaga, J. Taylor, A. Nekrutenko, and J. Goecks. The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res*, page gkw343, may 2016.
- [3] L. Aguilar-Arnal and P. Sassone-Corsi. Chromatin landscape and circadian dynamics: Spatial and temporal organization of clock transcription. *Proceedings of the National Academy of Sciences*, 112(22):6863–6870, nov 2014.
- [4] M. Akashi and T. Takumi. The orphan nuclear receptor ROR regulates circadian transcription of the mammalian core-clock *bmal1*. *Nat Struct Mol Biol*, 12(5):441–448, apr 2005.
- [5] R. A. Akhtar, A. B. Reddy, E. S. Maywood, J. D. Clayton, V. M. King, A. G. Smith, T. W. Gant, M. H. Hastings, and C. P. Kyriacou. Circadian cycling of the mouse liver transcriptome, as revealed by cDNA microarray, is driven by the suprachiasmatic nucleus. *Curr Biol*, 12(7):540–550, 2002.
- [6] J. H. Albrecht, R. Y. Poon, C. L. Ahonen, B. M. Rieland, C. Deng, and G. S. Crary. Involvement of p21 and p27 in the regulation of cdk activity and cell cycle progression in the regenerating liver. *Oncogene*, 1998.
- [7] B. Alipanahi, A. Delong, M. T. Weirauch, and B. J. Frey. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat Biotechnol*, 33(8):831–838, jul 2015.
- [8] M. M. Altarejos JY. Creb and the crtc co-activators: sensors for hormonal and metabolic signals. *Nat Rev Mol Cell Biol*, 2011.
- [9] A. F. M. Altelaar, J. Munoz, and A. J. R. Heck. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet*, 14(1):35–48, dec 2012.

## Bibliography

---

- [10] D. R. Ambrosini G. Bucher P. Principles of ChIP-seq Data Analysis Illustrated with Examples. In *Proceedings of the 2nd International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO)*, 2014.
- [11] H. C. Andersen. Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics*, 52(1):24–34, oct 1983.
- [12] J. L. Andrews, X. Zhang, J. J. McCarthy, E. L. McDearmon, T. A. Hornberger, B. Russell, K. S. Campbell, S. Arbogast, M. B. Reid, J. R. Walker, J. B. Hogenesch, J. S. Takahashi, and K. A. Esser. CLOCK and BMAL1 regulate MyoD and are necessary for maintenance of skeletal muscle phenotype and function. *Proceedings of the National Academy of Sciences*, 107(44):19090–19095, oct 2010.
- [13] C. Arizmendi, S. Liu, C. Croniger, V. Poli, and J. E. Friedman. The transcription factor CCAAT/enhancer-binding protein regulates gluconeogenesis and phosphoenolpyruvate carboxykinase (GTP) gene transcription during diabetes. *Journal of Biological Chemistry*, 274(19):13033–13040, may 1999.
- [14] J. V. Aschoff and R. Wever. Spontanperiodik des menschen bei ausschluß aller zeitgeber. *Naturwissenschaften*, 49(15):337–342, aug 1962.
- [15] G. Asher, D. Gatfield, M. Stratmann, H. Reinke, C. Dibner, F. Kreppel, R. Mostoslavsky, E. W. Alt, and U. Schibler. SIRT1 regulates circadian clock gene expression through PER2 deacetylation. *Cell*, 134(2):317–328, 2008.
- [16] G. Asher and U. Schibler. Crosstalk between components of circadian and metabolic cycles in mammals. *Cell Metabolism*, 13(2):125–137, feb 2011.
- [17] F. Atger, C. Gobet, J. Marquis, E. Martin, J. Wang, B. Weger, G. Lefebvre, P. Descombes, F. Naef, and F. Gachon. Circadian and feeding rhythms differentially affect rhythmic mRNA transcription and translation in mouse liver. *Proceedings of the National Academy of Sciences*, 112(47):E6579–E6588, nov 2015.
- [18] A. Auton, G. R. Abecasis, D. M. Altshuler, R. M. Durbin, G. R. Abecasis, D. R. Bentley, A. Chakravarti, A. G. Clark, P. Donnelly, E. E. Eichler, P. Flück, S. B. Gabriel, R. A. Gibbs, E. D. Green, M. E. Hurles, B. M. Knoppers, J. O. Korbel, E. S. Lander, C. Lee, H. Lehrach, E. R. Mardis, G. T. Marth, G. A. McVean, D. A. Nickerson, J. P. Schmidt, S. T. Sherry, J. Wang, R. K. Wilson, R. A. Gibbs, E. Boerwinkle, H. Doddapaneni, Y. Han, V. Korchina, C. Kovar, S. Lee, D. Muzny, J. G. Reid, Y. Zhu, J. Wang, Y. Chang, Q. Feng, X. Fang, X. Guo, M. Jian, H. Jiang, X. Jin, T. Lan, G. Li, J. Li, Y. Li, S. Liu, X. Liu, Y. Lu, X. Ma, M. Tang, B. Wang, G. Wang, H. Wu, R. Wu, X. Xu, Y. Yin, D. Zhang, W. Zhang, J. Zhao, M. Zhao, X. Zheng, E. S. Lander, D. M. Altshuler, S. B. Gabriel, N. Gupta, N. Gharani, L. H. Toji, N. P. Gerry, A. M. Resch, P. Flück, J. Barker, L. Clarke, L. Gil, S. E. Hunt, G. Kelman, E. Kulesha, R. Leinonen, W. M. McLaren, R. Radhakrishnan, A. Roa, D. Smirnov, R. E. Smith, I. Streeter, A. Thormann, I. Toneva, B. Vaughan, X. Zheng-Bradley, D. R. Bentley, R. Grocock, S. Humphray, T. James, Z. Kingsbury, H. Lehrach, R. Sudbrak, M. W. Albrecht, V. S.

Amstislavskiy, T. A. Borodina, M. Lienhard, F. Mertes, M. Sultan, B. Timmermann, M.-L. Yaspo, E. R. Mardis, R. K. Wilson, L. Fulton, R. Fulton, S. T. Sherry, V. Ananiev, Z. Belaia, D. Beloslyudtsev, N. Bouk, C. Chen, D. Church, R. Cohen, C. Cook, J. Garner, T. Hefferon, M. Kimelman, C. Liu, J. Lopez, P. Meric, C. O'Sullivan, Y. Ostapchuk, L. Phan, S. Ponomarov, V. Schneider, E. Shekhtman, K. Sirotkin, D. Slotta, H. Zhang, G. A. McVean, R. M. Durbin, S. Balasubramaniam, J. Burton, P. Danecek, T. M. Keane, A. Kolb-Kokocinski, S. McCarthy, J. Stalker, M. Quail, J. P. Schmidt, C. J. Davies, J. Gollub, T. Webster, B. Wong, Y. Zhan, A. Auton, C. L. Campbell, Y. Kong, A. Marcketta, R. A. Gibbs, F. Yu, L. Antunes, M. Bainbridge, D. Muzny, A. Sabo, Z. Huang, J. Wang, L. J. M. Coin, L. Fang, X. Guo, X. Jin, G. Li, Q. Li, Y. Li, Z. Li, H. Lin, B. Liu, R. Luo, H. Shao, Y. Xie, C. Ye, C. Yu, F. Zhang, H. Zheng, H. Zhu, C. Alkan, E. Dal, F. Kahveci, G. T. Marth, E. P. Garrison, D. Kural, W.-P. Lee, W. F. Leong, M. Stromberg, A. N. Ward, J. Wu, M. Zhang, M. J. Daly, M. A. DePristo, R. E. Handsaker, D. M. Altshuler, E. Banks, G. Bhatia, G. del Angel, S. B. Gabriel, G. Genovese, N. Gupta, H. Li, S. Kashin, E. S. Lander, S. A. McCarroll, J. C. Nemes, R. E. Poplin, S. C. Yoon, J. Lihm, V. Makarov, A. G. Clark, S. Gottipati, A. Keinan, J. L. Rodriguez-Flores, J. O. Korbel, T. Rausch, M. H. Fritz, A. M. Stütz, P. Flicek, K. Beal, L. Clarke, A. Datta, J. Herrero, W. M. McLaren, G. R. S. Ritchie, R. E. Smith, D. Zerbino, X. Zheng-Bradley, P. C. Sabeti, I. Shlyakhter, S. F. Schaffner, J. Vitti, D. N. Cooper, E. V. Ball, P. D. Stenson, D. R. Bentley, B. Barnes, M. Bauer, R. K. Cheetham, A. Cox, M. Eberle, S. Humphray, S. Kahn, L. Murray, J. Peden, R. Shaw, E. E. Kenny, M. A. Batzer, M. K. Konkel, J. A. Walker, D. G. MacArthur, M. Lek, R. Sudbrak, V. S. Amstislavskiy, R. Herwig, E. R. Mardis, L. Ding, D. C. Koboldt, D. Larson, K. Ye, S. Gravel, A. Swaroop, E. Chew, T. Lappalainen, Y. Erlich, M. Gymrek, T. F. Willems, J. T. Simpson, M. D. Shriver, J. A. Rosenfeld, C. D. Bustamante, S. B. Montgomery, F. M. D. L. Vega, J. K. Byrnes, A. W. Carroll, M. K. DeGorter, P. Lacroite, B. K. Maples, A. R. Martin, A. Moreno-Estrada, S. S. Shringarpure, F. Zakharia, E. Halperin, Y. Baran, C. Lee, E. Cerveira, J. Hwang, A. Malhotra, D. Plewczynski, K. Radew, M. Romanovitch, C. Zhang, F. C. L. Hyland, D. W. Craig, A. Christoforides, N. Homer, T. Izatt, A. A. Kurdoglu, S. A. Sinari, K. Squire, S. T. Sherry, C. Xiao, J. Sebat, D. Antaki, M. Gujral, A. Noor, K. Ye, E. G. Burchard, R. D. Hernandez, C. R. Gignoux, D. Haussler, S. J. Katzman, W. J. Kent, B. Howie, A. Ruiz-Linares, E. T. Dermitzakis, S. E. Devine, G. R. Abecasis, H. M. Kang, J. M. Kidd, T. Blackwell, S. Caron, W. Chen, S. Emery, L. Fritsche, C. Fuchsberger, G. Jun, B. Li, R. Lyons, C. Scheller, C. Sidore, S. Song, E. Sliwerska, D. Taliun, A. Tan, R. Welch, M. K. Wing, X. Zhan, P. Awadalla, A. Hodgkinson, Y. Li, X. Shi, A. Quitadamo, G. Lunter, G. A. McVean, J. L. Marchini, S. Myers, C. Churchhouse, O. Delaneau, A. Gupta-Hinch, W. Kretschmar, Z. Iqbal, I. Mathieson, A. Menelaou, A. Rimmer, D. K. Xifara, T. K. Oleksyk, Y. Fu, X. Liu, M. Xiong, L. Jorde, D. Witherspoon, J. Xing, E. E. Eichler, B. L. Browning, S. R. Browning, F. Hormozdiari, P. H. Sudmant, E. Khurana, R. M. Durbin, M. E. Hurles, C. Tyler-Smith, C. A. Albers, Q. Ayub, S. Balasubramaniam, Y. Chen, V. Colonna, P. Danecek, L. Jostins, T. M. Keane, S. McCarthy, K. Walter, Y. Xue, M. B. Gerstein, A. Abyzov, S. Balasubramanian, J. Chen, D. Clarke, Y. Fu, A. O. Harman, M. Jin, D. Lee, J. Liu, X. J. Mu, J. Zhang, Y. Zhang, Y. Li, R. Luo, H. Zhu, C. Alkan, E. Dal, F. Kahveci, G. T. Marth, E. P. Garrison, D. Kural, W.-P. Lee, A. N. Ward, J. Wu, M. Zhang, S. A. McCarroll, R. E.

## Bibliography

---

Handsaker, D. M. Altshuler, E. Banks, G. del Angel, G. Genovese, C. Hartl, H. Li, S. Kashin, J. C. Nemesh, K. Shakir, S. C. Yoon, J. Lihm, V. Makarov, J. Degenhardt, J. O. Korbel, M. H. Fritz, S. Meiers, B. Raeder, T. Rausch, A. M. Stütz, P. Flicek, F. P. Casale, L. Clarke, R. E. Smith, O. Stegle, X. Zheng-Bradley, D. R. Bentley, B. Barnes, R. K. Cheetham, M. Eberle, S. Humphray, S. Kahn, L. Murray, R. Shaw, E.-W. Lameijer, M. A. Batzer, M. K. Konkel, J. A. Walker, L. Ding, I. Hall, K. Ye, P. Lacroute, C. Lee, E. Cerveira, A. Malhotra, J. Hwang, D. Plewczynski, K. Radew, M. Romanovitch, C. Zhang, D. W. Craig, N. Homer, D. Church, C. Xiao, J. Sebat, D. Antaki, V. Bafna, J. Michaelson, K. Ye, S. E. Devine, E. J. Gardner, G. R. Abecasis, J. M. Kidd, R. E. Mills, G. Dayama, S. Emery, G. Jun, X. Shi, A. Quitadamo, G. Lunter, G. A. McVean, K. Chen, X. Fan, Z. Chong, T. Chen, D. Witherspoon, J. Xing, E. E. Eichler, M. J. Chaisson, F. Hormozdiari, J. Huddleston, M. Malig, B. J. Nelson, P. H. Sudmant, N. F. Parrish, E. Khurana, M. E. Hurles, B. Blackburne, S. J. Lindsay, Z. Ning, K. Walter, Y. Zhang, M. B. Gerstein, A. Abyzov, J. Chen, D. Clarke, H. Lam, X. J. Mu, C. Sisu, J. Zhang, Y. Zhang, R. A. Gibbs, F. Yu, M. Bainbridge, D. Challis, U. S. Evani, C. Kovar, J. Lu, D. Muzny, U. Nagaswamy, J. G. Reid, A. Sabo, J. Yu, X. Guo, W. Li, Y. Li, R. Wu, G. T. Marth, E. P. Garrison, W. F. Leong, A. N. Ward, G. del Angel, M. A. DePristo, S. B. Gabriel, N. Gupta, C. Hartl, R. E. Poplin, A. G. Clark, J. L. Rodriguez-Flores, P. Flicek, L. Clarke, R. E. Smith, X. Zheng-Bradley, D. G. MacArthur, E. R. Mardis, R. Fulton, D. C. Koboldt, S. Gravel, C. D. Bustamante, D. W. Craig, A. Christoforides, N. Homer, T. Izatt, S. T. Sherry, C. Xiao, E. T. Dermitzakis, G. R. Abecasis, H. M. Kang, G. A. McVean, M. B. Gerstein, S. Balasubramanian, L. Habegger, H. Yu, P. Flicek, L. Clarke, F. Cunningham, I. Dunham, D. Zerbino, X. Zheng-Bradley, K. Lage, J. B. Jaspersen, H. Horn, S. B. Montgomery, M. K. DeGorter, E. Khurana, C. Tyler-Smith, Y. Chen, V. Colonna, Y. Xue, M. B. Gerstein, S. Balasubramanian, Y. Fu, D. Kim, A. Auton, A. Marcketta, R. Desalle, A. Narechania, M. A. W. Sayres, E. P. Garrison, R. E. Handsaker, S. Kashin, S. A. McCarroll, J. L. Rodriguez-Flores, P. Flicek, L. Clarke, X. Zheng-Bradley, Y. Erlich, M. Gymrek, T. F. Willems, C. D. Bustamante, F. L. Mendez, G. D. Poznik, P. A. Underhill, C. Lee, E. Cerveira, A. Malhotra, M. Romanovitch, C. Zhang, G. R. Abecasis, L. Coin, H. Shao, D. Mittelman, C. Tyler-Smith, Q. Ayub, R. Banerjee, M. Cerezo, Y. Chen, T. W. Fitzgerald, S. Louzada, A. Massaia, S. McCarthy, G. R. Ritchie, Y. Xue, F. Yang, R. A. Gibbs, C. Kovar, D. Kalra, W. Hale, D. Muzny, J. G. Reid, J. Wang, X. Dan, X. Guo, G. Li, Y. Li, C. Ye, X. Zheng, D. M. Altshuler, P. Flicek, L. Clarke, X. Zheng-Bradley, D. R. Bentley, A. Cox, S. Humphray, S. Kahn, R. Sudbrak, M. W. Albrecht, M. Lienhard, D. Larson, D. W. Craig, T. Izatt, A. A. Kurdoglu, S. T. Sherry, C. Xiao, D. Haussler, G. R. Abecasis, G. A. McVean, R. M. Durbin, S. Balasubramanian, T. M. Keane, S. McCarthy, J. Stalker, A. Chakravarti, B. M. Knoppers, G. R. Abecasis, K. C. Barnes, C. Beiswanger, E. G. Burchard, C. D. Bustamante, H. Cai, H. Cao, R. M. Durbin, N. P. Gerry, N. Gharani, R. A. Gibbs, C. R. Gignoux, S. Gravel, B. Henn, D. Jones, L. Jorde, J. S. Kaye, A. Keinan, A. Kent, A. Kerasidou, Y. Li, R. Mathias, G. A. McVean, A. Moreno-Estrada, P. N. Ossorio, M. Parker, A. M. Resch, C. N. Rotimi, C. D. Royal, K. Sandoval, Y. Su, R. Sudbrak, Z. Tian, S. Tishkoff, L. H. Toji, C. Tyler-Smith, M. Via, Y. Wang, H. Yang, L. Yang, J. Zhu, W. Bodmer, G. Bedoya, A. Ruiz-Linares, Z. Cai, Y. Gao, J. Chu, L. Peltonen, A. Garcia-Montero, A. Orfao, J. Dutil, J. C. Martinez-Cruzado, T. K. Oleksyk, K. C. Barnes,



- R. A. Mathias, A. Hennis, H. Watson, C. McKenzie, F. Qadri, R. LaRocque, P. C. Sabeti, J. Zhu, X. Deng, P. C. Sabeti, D. Asogun, O. Folarin, C. Happi, O. Omoniwa, M. Strelau, R. Tariyal, M. Jallow, F. S. Joof, T. Corrah, K. Rockett, D. Kwiatkowski, J. Kooner, T. T. Hiên, S. J. Dunstan, N. T. Hang, R. Fonnies, R. Garry, L. Kanneh, L. Moses, P. C. Sabeti, J. Schieffelin, D. S. Grant, C. Gallo, G. Poletti, D. Saleheen, A. Rasheed, L. D. Brooks, A. L. Felsenfeld, J. E. McEwen, Y. Vaydylevich, E. D. Green, A. Duncanson, M. Dunn, J. A. Schloss, J. Wang, H. Yang, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, and G. R. Abecasis. A global reference for human genetic variation. *Nature*, 526(7571):68–74, sep 2015.
- [19] R. Aviram, G. Manella, N. Kopelman, A. Neufeld-Cohen, Z. Zwihaft, M. Elimelech, Y. Adamovich, M. Golik, C. Wang, X. Han, and G. Asher. Lipidomics analyses reveal temporal and spatial lipid organization and uncover daily oscillations in intracellular organelles. *Molecular Cell*, 62(4):636–648, may 2016.
- [20] S. Baek, M. H. Sung, and G. L. Hager. Quantitative analysis of genome-wide chromatin remodeling. *Methods Mol Biol*, 833:433–441, 2012.
- [21] T. Bailey, P. Krajewski, I. Ladunga, C. Lefebvre, Q. Li, T. Liu, P. Madrigal, C. Taslim, and J. Zhang. Practical guidelines for the comprehensive analysis of ChIP-seq data. *PLoS Comput Biol*, 9(11):e1003326, nov 2013.
- [22] T. L. Bailey and P. Machanick. Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Research*, 40(17):e128–e128, may 2012.
- [23] T. L. Bailey, N. Williams, C. Mislé, and W. W. Li. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic acids research*, 34(Web Server issue):W369–73, jul 2006.
- [24] A. Balsalobre, S. A. Brown, L. Marcacci, F. Tronche, C. Kellendonk, H. M. Reichardt, G. Schutz, and U. Schibler. Resetting of circadian time in peripheral tissues by glucocorticoid signaling. *Science*, 289(5488):2344–2347, 2000.
- [25] A. Balsalobre, F. Damiola, and U. Schibler. A serum shock induces circadian gene expression in mammalian tissue culture cells. *Cell*, 93(6):929–937, 1998.
- [26] P. J. Balwierz, M. Pachkov, P. Arnold, A. J. Gruber, M. Zavolan, and E. van Nimwegen. ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. *Genome Res*, 24(5):869–884, 2014.
- [27] A. Barski, S. Cuddapah, K. Cui, T.-Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao. High-resolution profiling of histone methylations in the human genome. *Cell*, 129(4):823–837, may 2007.
- [28] J. Bass and J. S. Takahashi. Circadian integration of metabolism and energetics. *Science*, 330(6009):1349–1354, dec 2010.

## Bibliography

---

- [29] S. Becker-Weimann, J. Wolf, H. Herzelt, and A. Kramer. Modeling feedback loops of the Mammalian circadian oscillator. *Biophys J*, 87(5):3023–3034, 2004.
- [30] I. Beling. Uber das zeitgedachtnis der bienen. *Z. Vgl. Physiol*, 1929.
- [31] Y. Benjamini. Discovering the false discovery rate. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):405–416, aug 2010.
- [32] C. Bertolucci, N. Cavallari, I. Colognesi, J. Aguzzi, Z. Chen, P. Caruso, A. Foa, G. Tosini, F. Bernardi, and M. Pinotti. Evidence for an overlapping role of CLOCK and NPAS2 transcription factors in liver circadian oscillators. *Mol Cell Biol*, 28(9):3070–3075, 2008.
- [33] C. Bodenstern, I. Heiland, and S. Schuster. Temperature compensation and entrainment in circadian rhythms. *Phys. Biol.*, 9(3):036011, jun 2012.
- [34] N. Bonhoure, G. Bounova, D. Bernasconi, V. Praz, F. Lammers, D. Canella, I. M. Willis, W. Herr, N. Hernandez, M. Delorenzi, N. Hernandez, M. Delorenzi, B. Deplancke, B. Desvergne, N. Guex, W. Herr, F. Naef, J. Rougemont, U. Schibler, T. Andersin, P. Cousin, F. Gilardi, P. Gos, F. Lammers, S. Raghav, D. Villeneuve, R. Fabbretti, V. Vlegel, I. Xenarios, E. Migliavacca, V. Praz, F. David, Y. Jarosz, D. Kuznetsov, R. Liechti, O. Martin, J. Delafontaine, J. Cajan, K. Gustafson, I. Krier, M. Leleu, N. Molina, A. Naldi, L. Rib, L. Symul, and G. Bounova. Quantifying ChIP-seq data: a spiking method providing an internal reference for sample-to-sample normalization. *Genome Research*, 24(7):1157–1168, apr 2014.
- [35] S. Bonn, R. P. Zinzen, C. Girardot, E. H. Gustafson, A. Perez-Gonzalez, N. Delhomme, Y. Ghavi-Helm, B. Wilczynski, A. Riddell, and E. E. Furlong. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat Genet*, 44(2):148–156, 2012.
- [36] A. P. Boyle, S. Davis, H. P. Shulha, P. Meltzer, E. H. Margulies, Z. Weng, T. S. Furey, and G. E. Crawford. High-resolution mapping and characterization of open chromatin across the genome. *Cell*, 132(2):311–322, 2008.
- [37] A. P. Boyle, L. Song, B. K. Lee, D. London, D. Keefe, E. Birney, V. R. Iyer, G. E. Crawford, and T. S. Furey. High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. *Genome Res*, 21(3):456–464, 2011.
- [38] K. Bozek, A. Relogio, S. M. Kielbasa, M. Heine, C. Dame, A. Kramer, and H. Herzelt. Regulation of clock-controlled genes in mammals. *PLoS One*, 4(3):e4882, 2009.
- [39] K. Bozek, A. L. Rosahl, S. Gaub, S. Lorenzen, and H. Herzelt. Circadian transcription in liver. *Biosystems*, 102(1):61–69, 2010.
- [40] M. S. Bray and M. E. Young. The role of cell-specific circadian clocks in metabolism and disease. *Obesity Reviews*, 10:6–13, nov 2009.

- 
- [41] S. A. Brown, E. Kowalska, and R. Dallmann. (Re)inventing the circadian feedback loop. *Dev Cell*, 22(3):477–487, 2012.
- [42] J. D. Buenrostro, B. Wu, U. M. Litzenburger, D. Ruff, M. L. Gonzales, M. P. Snyder, H. Y. Chang, and W. J. Greenleaf. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*, 523(7561):486–490, jun 2015.
- [43] A. Bugge, D. Feng, L. J. Everett, E. R. Briggs, S. E. Mullican, F. Wang, J. Jager, and M. A. Lazar. Rev-erbalpha and Rev-erbbeta coordinately protect the circadian clock and normal metabolic function. *Genes Dev*, 26(7):657–667, 2012.
- [44] H. J. Bussemaker, H. Li, and E. D. Siggia. Regulatory element detection using correlation with expression. *Nature Genetics*, 27(2):167–174, feb 2001.
- [45] Bünning. Zur kenntnis der erblichen ttagesperiodizita bei den pimarblattern von phaseolus multiflorus. *Jahrb wiss Botan*, 1935.
- [46] L. C and L. JD. Pgc-1 coactivators in the control of energy metabolism. *Acta Biochim Biophys Sin (Shanghai)*, 2011.
- [47] P. C, D. R, and G. J. Role of the liver in the control of carbohydrate and lipid homeostasis. *Diabetes Metab.*, 2004.
- [48] F. R. Cagampang and K. D. Bruce. The role of the circadian clock system in nutrition and metabolism. *Br J Nutr*, 108(3):381–392, 2012.
- [49] A. A. Canutescu and R. L. Dunbrack Jr. Cyclic coordinate descent: A robotics algorithm for protein loop closure. *Protein Sci*, 12(5):963–972, 2003.
- [50] J. Cao, Z. Luo, Q. Cheng, Q. Xu, Y. Zhang, F. Wang, Y. Wu, and X. Song. Three-dimensional regulation of transcription. *Protein Cell*, 6(4):241–253, feb 2015.
- [51] B. T. Carneiro and J. F. Araujo. Food entrainment: major and recent findings. *Front Behav Neurosci*, 6:83, 2012.
- [52] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods. The amber biomolecular simulation programs. *J. Comput. Chem.*, 26(16):1668–1688, 2005.
- [53] A. Castello, R. Horos, C. Strein, B. Fischer, K. Eichelbaum, L. M. Steinmetz, J. Krijgsveld, and M. W. Hentze. System-wide identification of RNA-binding proteins by interactome capture. *Nat Protoc*, 8(3):491–500, feb 2013.
- [54] N. Cermakian. Altered behavioral rhythms and clock gene expression in mice with a targeted mutation in the period1 gene. *The EMBO Journal*, 20(15):3967–3974, aug 2001.
- [55] A. Chaix, A. Zarrinpar, P. Miu, and S. Panda. Article Time-Restricted Feeding Is a Preventative and Therapeutic Intervention against Diverse Nutritional Challenges. *Cell Metabolism*, 20(6):991–1005, 2014.

## Bibliography

---

- [56] R. Chavan, C. Feillet, S. S. F. Costa, J. E. Delorme, T. Okabe, J. A. Ripperger, and U. Albrecht. Liver-derived ketone bodies are necessary for food anticipation. *Nature Communications*, 7:10580, feb 2016.
- [57] X. Chen, M. M. Hoffman, J. a. Bilmes, J. R. Hesselberth, and W. S. Noble. A dynamic Bayesian network for identifying protein-binding footprints from single molecule-based sequencing data. *Bioinformatics (Oxford, England)*, 26(12):i334–42, jun 2010.
- [58] Y. Chen, M. Jorgensen, R. Kolde, X. Zhao, B. Parker, E. Valen, J. Wen, and A. Sandelin. Prediction of RNA Polymerase II recruitment, elongation and stalling from histone modification data. *BMC Genomics*, 12:544, 2011.
- [59] S. Cheon, N. Park, S. Cho, and K. Kim. Glucocorticoid-mediated Period2 induction delays the phase of circadian rhythm. *Nucleic Acids Res*, 41(12):6161–6174, 2013.
- [60] H. Cho, X. Zhao, M. Hatori, R. T. Yu, G. D. Barish, M. T. Lam, L. W. Chong, L. DiTacchio, A. R. Atkins, C. K. Glass, C. Liddle, J. Auwerx, M. Downes, S. Panda, and R. M. Evans. Regulation of circadian behaviour and metabolism by REV-ERB-alpha and REV-ERB-beta. *Nature*, 485(7396):123–127, 2012.
- [61] L. A. Cirillo, F. R. Lin, I. Cuesta, D. Friedman, M. Jarnik, and K. S. Zaret. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Molecular Cell*, 9(2):279–289, feb 2002.
- [62] C. R. Clapier and B. R. Cairns. The biology of chromatin remodeling complexes. *Annu Rev Biochem*, 78:273–304, 2009.
- [63] M. Claussnitzer, S. N. Dankel, K.-H. Kim, G. Quon, W. Meuleman, C. Haugen, V. Glunk, I. S. Sousa, J. L. Beaudry, V. Puvionandran, N. A. Abdennur, J. Liu, P.-A. Svensson, Y.-H. Hsu, D. J. Drucker, G. Mellgren, C.-C. Hui, H. Hauner, and M. Kellis. FTO obesity variant circuitry and adipocyte browning in humans. *New England Journal of Medicine*, 373(10):895–907, sep 2015.
- [64] G. E. Crawford, I. E. Holt, J. Whittle, B. D. Webb, D. Tai, S. Davis, E. H. Margulies, Y. Chen, J. a. Bernat, D. Ginsburg, D. Zhou, S. Luo, T. J. Vasicek, M. J. Daly, T. G. Wolfsberg, and F. S. Collins. Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome research*, 16(1):123–31, jan 2006.
- [65] M. P. Creighton, A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young, and R. Jaenisch. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A*, 107(50):21931–21936, 2010.
- [66] C. Crumbley, Y. Wang, D. J. Kojetin, and T. P. Burris. Characterization of the core mammalian clock component, NPAS2, as a REV-ERBalpha/RORalpha target gene. *J Biol Chem*, 285(46):35386–35392, 2010.

- [67] S. Daan. *The Circadian Clock*, chapter A History of Chronobiological Concepts, pages 1–35. Springer New York, 2010.
- [68] H. Daitoku and A. Fukamizu. FOXO transcription factors in the regulatory networks of longevity. *J Biochem*, 141(6):769–774, 2007.
- [69] R. Dallmann, S. A. Brown, and F. Gachon. Chronopharmacology: New insights and therapeutic implications. *Annu. Rev. Pharmacol. Toxicol.*, 54(1):339–361, jan 2014.
- [70] F. Damiola, N. Le Minh, N. Preitner, B. Kornmann, F. Fleury-Olela, and U. Schibler. Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. *Genes Dev*, 14(23):2950–2961, 2000.
- [71] G. S. Daria Shlyueva and A. Stark. Transcriptional enhancers: from properties to genome-wide predictions. *NATURE REVIEWS GENETICS*, 2014.
- [72] F. P. David, J. Delafontaine, S. Carat, F. J. Ross, G. Lefebvre, Y. Jarosz, L. Sinclair, D. Noordermeer, J. Rougemont, and M. Leleu. HTSstation: a web application and open-access libraries for high-throughput sequencing data analysis. *PLoS One*, 9(1):e85879, 2014.
- [73] P. J. DeCoursey, J. C. Dunlap, and J. J. Loros. *Chronobiology*. Sinauer Associates Inc, 2003.
- [74] J. J. DeMairan. Observation botanique. *L'histoire de l'Academie Royale Scientifique*, 1729.
- [75] R. Dentin, J. Girard, and C. Postic. Carbohydrate responsive element binding protein (ChREBP) and sterol regulatory element binding protein-1c (SREBP-1c): two key regulators of glucose metabolism and lipid synthesis in liver. *Biochimie*, 87(1):81–86, jan 2005.
- [76] L. Desbarats, S. Gaubatz, and M. Eilers. Discrimination between different E-box-binding proteins at an endogenous target gene of c-myc. *Genes Dev*, 10(4):447–460, 1996.
- [77] P. D'haeseleer. What are DNA sequence motifs? *Nat Biotechnol*, 24(4):423–425, apr 2006.
- [78] M.-A. Dillies, A. Rau, J. Aubert, C. Hennequet-Antier, M. Jeanmougin, N. Servant, C. Keime, G. Marot, D. Castel, J. Estelle, G. Guernec, B. Jagla, L. Jouneau, D. Laloe, C. L. Gall, B. Schaeffer, S. L. Crom, M. Guedj, and F. Jaffrezic. A comprehensive evaluation of normalization methods for illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics*, 14(6):671–683, sep 2012.
- [79] L. DiTacchio, H. D. Le, C. Vollmers, M. Hatori, M. Witcher, J. Secombe, and S. Panda. Histone lysine demethylase JARID1a activates CLOCK-BMAL1 and influences the circadian clock. *Science*, 333(6051):1881–1885, sep 2011.
- [80] M. Doi, J. Hirayama, and P. Sassone-Corsi. Circadian regulator CLOCK is a histone acetyltransferase. *Cell*, 125(3):497–508, 2006.

## Bibliography

---

- [81] A. Dominguez-Rodriguez, P. Abreu-Gonzalez, J. J. Sanchez-Sanchez, J. C. Kaski, and R. J. Reiter. Melatonin and circadian biology in human cardiovascular disease. *Journal of Pineal Research*, pages no–no, may 2010.
- [82] X. C. Dong, K. D. Copps, S. Guo, Y. Li, R. Kollipara, R. A. DePinho, and M. F. White. Inactivation of hepatic Foxo1 by insulin signaling is required for adaptive nutrient homeostasis and endocrine growth regulation. *Cell Metab*, 8(1):65–76, 2008.
- [83] N.-H. Du, A. B. Arpat, M. D. Matos, and D. Gatfield. Micrnas shape circadian hepatic gene expression on a transcriptome-wide scale. *eLife*, 2014.
- [84] J. J. L. Dunlap, Jay C. and P. J. DeCoursey. *Chronobiology: Biological timekeeping*. Sinauer Associates, 2004.
- [85] H. A. Duong and C. J. Weitz. Temporal orchestration of repressive chromatin modifiers by circadian clock period complexes. *Nat Struct Mol Biol*, 21(2):126–132, jan 2014.
- [86] J. Dupuis, C. Langenberg, I. Prokopenko, R. Saxena, N. Soranzo, A. U. Jackson, E. Wheeler, N. L. Glazer, N. Bouatia-Naji, A. L. Gloyn, C. M. Lindgren, R. Mägi, A. P. Morris, J. Randall, T. Johnson, P. Elliott, D. Rybin, G. Thorleifsson, V. Steinthorsdottir, P. Henneman, H. Grallert, A. Dehghan, J. J. Hottenga, C. S. Franklin, P. Navarro, K. Song, A. Goel, J. R. B. Perry, J. M. Egan, T. Lajunen, N. Grarup, T. Sparsø, A. Doney, B. F. Voight, H. M. Stringham, M. Li, S. Kanoni, P. Shrader, C. Cavalcanti-Proença, M. Kumari, L. Qi, N. J. Timpson, C. Gieger, C. Zabena, G. Rocheleau, E. Ingelsson, P. An, J. O'Connell, J. Luan, A. Elliott, S. A. McCarroll, F. Payne, R. M. Ruccasecca, F. Pattou, P. Sethupathy, K. Ardlie, Y. Ariyurek, B. Balkau, P. Barter, J. P. Beilby, Y. Ben-Shlomo, R. Benediktsson, A. J. Bennett, S. Bergmann, M. Bochud, E. Boerwinkle, A. Bonnefond, L. L. Bonnycastle, K. Borch-Johnsen, Y. Böttcher, E. Brunner, S. J. Bumpstead, G. Charpentier, Y.-D. I. Chen, P. Chines, R. Clarke, L. J. M. Coin, M. N. Cooper, M. Cornelis, G. Crawford, L. Crisponi, I. N. M. Day, E. J. C. de Geus, J. Delplanque, C. Dina, M. R. Erdos, A. C. Fedson, A. Fischer-Rosinsky, N. G. Forouhi, C. S. Fox, R. Frants, M. G. Franzosi, P. Galan, M. O. Goodarzi, J. Graessler, C. J. Groves, S. Grundy, R. Gwilliam, U. Gyllensten, S. Hadjadj, G. Hallmans, N. Hammond, X. Han, A.-L. Hartikainen, N. Hassanali, C. Hayward, S. C. Heath, S. Hercberg, C. Herder, A. A. Hicks, D. R. Hillman, A. D. Hingorani, A. Hofman, J. Hui, J. Hung, B. Iso-maa, P. R. V. Johnson, T. Jørgensen, A. Jula, M. Kaakinen, J. Kaprio, Y. A. Kesaniemi, M. Kivimaki, B. Knight, S. Koskinen, P. Kovacs, K. O. Kyvik, G. M. Lathrop, D. A. Lawlor, O. L. Bacquer, C. Lecoeur, Y. Li, V. Lyssenko, R. Mahley, M. Mangino, A. K. Manning, M. T. Martínez-Larrad, J. B. McAteer, L. J. McCulloch, R. McPherson, C. Meisinger, D. Melzer, D. Meyre, B. D. Mitchell, M. A. Morcken, S. Mukherjee, S. Naitza, N. Narisu, M. J. Neville, B. A. Oostra, M. Orrù, R. Pakyz, C. N. A. Palmer, G. Paolisso, C. Pattaro, D. Pearson, J. F. Peden, N. L. Pedersen, M. Perola, A. F. H. Pfeiffer, I. Pichler, O. Polasek, D. Posthuma, S. C. Potter, A. Pouta, M. A. Province, B. M. Psaty, W. Rathmann, N. W. Rayner, K. Rice, S. Ripatti, F. Rivadeneira, M. Roden, O. Rolandsson, A. Sandbaek, M. Sandhu, S. Sanna, A. A. Sayer, P. Scheet, L. J. Scott, U. Seedorf, S. J. Sharp, B. Shields, G. Sigurdsson, E. J. G.



- Sijbrands, A. Silveira, L. Simpson, A. Singleton, N. L. Smith, U. Sovio, A. Swift, H. Syddall, A.-C. Syvänen, T. Tanaka, B. Thorand, J. Tichet, A. Tönjes, T. Tuomi, A. G. Uitterlinden, K. W. van Dijk, M. van Hoek, D. Varma, S. Visvikis-Siest, V. Vitart, N. Vogelzangs, G. Waeber, P. J. Wagner, A. Walley, G. B. Walters, K. L. Ward, H. Watkins, M. N. Weedon, S. H. Wild, G. Willemsen, J. C. M. Witteman, J. W. G. Yarnell, E. Zeggini, D. Zelenika, B. Zethelius, G. Zhai, J. H. Zhao, M. C. Zillikens, I. B. Borecki, R. J. F. Loos, P. Meneton, P. K. E. Magnusson, D. M. Nathan, G. H. Williams, A. T. Hattersley, K. Silander, V. Salomaa, G. D. Smith, S. R. Bornstein, P. Schwarz, J. Spranger, F. Karpe, A. R. Shuldiner, C. Cooper, G. V. Dedoussis, M. Serrano-Ríos, A. D. Morris, L. Lind, L. J. Palmer, F. B. Hu, P. W. Franks, S. Ebrahim, M. Marmot, W. H. L. Kao, J. S. Pankow, M. J. Sampson, J. Kuusisto, M. Laakso, T. Hansen, O. Pedersen, P. P. Pramstaller, H. E. Wichmann, T. Illig, I. Rudan, A. F. Wright, M. Stumvoll, H. Campbell, J. F. Wilson, R. N. Bergman, T. A. Buchanan, F. S. Collins, K. L. Mohlke, J. Tuomilehto, T. T. Valle, D. Altshuler, J. I. Rotter, D. S. Siscovick, B. W. J. H. Penninx, D. I. Boomsma, P. Deloukas, T. D. Spector, T. M. Frayling, L. Ferrucci, A. Kong, U. Thorsteinsdottir, K. Stefansson, C. M. van Duijn, Y. S. Aulchenko, A. Cao, A. Scuteri, D. Schlessinger, M. Uda, A. Ruokonen, M.-R. Jarvelin, D. M. Waterworth, P. Vollenweider, L. Peltonen, V. Mooser, G. R. Abecasis, N. J. Wareham, R. Sladek, P. Froguel, R. M. Watanabe, J. B. Meigs, L. Groop, M. Boehnke, M. I. McCarthy, J. C. Florez, and I. Barroso. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nature Genetics*, 42(2):105–116, jan 2010.
- [87] G. Dürnberger, T. Bürckstümmer, K. Huber, R. Giambruno, T. Doerks, E. Karayel, T. R. Burkard, I. Kaupe, A. C. Müller, A. Schönegger, G. F. Ecker, H. Lohninger, P. Bork, K. L. Bennett, G. Superti-Furga, and J. Colinge. Experimental characterization of the human non-sequence-specific nucleic acid interactome. *Genome Biol*, 14(7):R81, 2013.
- [88] R. Eberly and H. Feldman. Obesity and shift work in the general population. *Internet Journal of Allied Health Sciences and Practice*, 2010.
- [89] K. Eckel-Mahan and P. Sassone-Corsi. Metabolism and the circadian clock converge. *Physiological reviews*, 93(1):107–35, jan 2013.
- [90] K. L. Eckel-Mahan, V. R. Patel, R. P. Mohny, K. S. Vignola, P. Baldi, and P. Sassone-Corsi. Coordination of the transcriptome and metabolome by the circadian clock. *PNAS*, 2012.
- [91] R. S. Edgar, E. W. Green, Y. Zhao, G. van Ooijen, M. Olmedo, X. Qin, Y. Xu, M. Pan, U. K. Valekunja, K. A. Feeney, E. S. Maywood, M. H. Hastings, N. S. Baliga, M. Meroow, A. J. Millar, C. H. Johnson, C. P. Kyriacou, J. S. O’Neill, and A. B. Reddy. Peroxiredoxins are conserved markers of circadian rhythms. *Nature*, 2012.
- [92] J. Ernst. Mapping enhancer and promoter interactions. *Cell Res*, 22(5):789–790, mar 2012.
- [93] L. J. Everett, J. Lay, S. Lukovac, D. Bernstein, D. J. Steger, M. A. Lazar, and K. H. Kaestner. Integrative genomic analysis of CREB defines a critical role for transcription factor networks in mediating the fed/fasted switch in liver. *BMC Genomics*, 14(1):337, 2013.

## Bibliography

---

- [94] L. J. Everett, J. Le Lay, S. Lukovac, D. Bernstein, D. J. Steger, M. A. Lazar, and K. H. Kaestner. Integrative genomic analysis of CREB defines a critical role for transcription factor networks in mediating the fed/fasted switch in liver. *BMC Genomics*, 14:337, 2013.
- [95] B. Fang, L. J. Everett, J. Jager, E. Briggs, S. M. Armour, D. Feng, A. Roy, Z. Gerhart-Hines, Z. Sun, and M. A. Lazar. Circadian enhancers coordinate multiple phases of rhythmic gene transcription in vivo. *Cell*, 159(5):1140–1152, 2014.
- [96] A. J. Faure, D. Schmidt, S. Watt, P. C. Schwalie, M. D. Wilson, H. Xu, R. G. Ramsay, D. T. Odom, and P. Flicek. Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules. *Genome Res*, 22(11):2163–2175, 2012.
- [97] H. M. Feldman JE. Isolation of circadian clock mutants of neurospora crassa. *Genetics*, 1973.
- [98] D. Feng, T. Liu, Z. Sun, A. Bugge, S. E. Mullican, T. Alenghat, X. S. Liu, and M. A. Lazar. A circadian rhythm orchestrated by histone deacetylase 3 controls hepatic lipid metabolism. *Science*, 331(6022):1315–1319, 2011.
- [99] R. A. Fisher. Tests of significance in harmonic analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 125(796):54–59, aug 1929.
- [100] J. Friedman, T. Hastie, and R. Tibshirani. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*, 33(1):1–22, 2010.
- [101] R. Frijters, W. Fleuren, E. J. Toonen, J. P. Tuckermann, H. M. Reichardt, H. van der Maaden, A. van Elsas, M.-J. van Lierop, W. Dokter, J. de Vlieg, and W. Alkema. Prednisolone-induced differential gene expression in mouse liver carrying wild type or a dimerization-defective glucocorticoid receptor. *BMC Genomics*, 11(1):359, 2010.
- [102] O. Froy. Circadian rhythms, aging, and life span in mammals. *Physiology*, 26(4):225–235, aug 2011.
- [103] F. Gachon, F. F. Olela, O. Schaad, P. Descombes, and U. Schibler. The circadian PAR-domain basic leucine zipper transcription factors DBP, TEF, and HLF modulate basal and inducible xenobiotic detoxification. *Cell metabolism*, 4(1):25–36, jul 2006.
- [104] M. Garaulet, M. D. Corbalán, J. A. Madrid, E. Morales, J. C. Baraza, Y. C. Lee, and J. M. Ordovas. CLOCK gene is implicated in weight reduction in obese patients participating in a dietary programme based on the mediterranean diet. *Int J Obes Relat Metab Disord*, 34(3):516–523, jan 2010.
- [105] M. Garaulet, C. Sánchez-Moreno, C. E. Smith, Y.-C. Lee, F. Nicolás, and J. M. Ordovás. Ghrelin, sleep reduction and evening preference: Relationships to CLOCK 3111 T/c SNP and weight loss. *PLoS ONE*, 6(2):e17435, feb 2011.



- [106] D. Gatfield, G. L. Martelot, C. E. Vejnar, D. Gerlach, O. Schaad, F. Fleury-Olela, A.-L. Ruskeepaa, M. Oresic, C. C. Esau, E. M. Zdobnov, and U. Schibler. Integration of microRNA miR-122 in hepatic circadian gene expression. *Genes & Development*, 23(11):1313–1326, jun 2009.
- [107] D. Gau, T. Lemberger, C. von Gall, O. Kretz, N. Le Minh, P. Gass, W. Schmid, U. Schibler, H. W. Korf, and G. Schutz. Phosphorylation of CREB Ser142 regulates light-induced phase shifts of the circadian clock. *Neuron*, 34(2):245–253, 2002.
- [108] N. Gekakis. Role of the CLOCK protein in the mammalian circadian mechanism. *Science*, 280(5369):1564–1569, jun 1998.
- [109] A. Gerber, C. Esnault, G. Aubert, R. Treisman, F. Pralong, and U. Schibler. Blood-borne circadian signal stimulates daily oscillations in actin dynamics and SRF activity. *Cell*, 152(3):492–503, jan 2013.
- [110] M. B. Gerstein, A. Kundaje, M. Hariharan, S. G. Landt, K.-K. Yan, C. Cheng, X. J. Mu, E. Khurana, J. Rozowsky, R. Alexander, R. Min, P. Alves, A. Abyzov, N. Addleman, N. Bhardwaj, A. P. Boyle, P. Cayting, A. Charos, D. Z. Chen, Y. Cheng, D. Clarke, C. Eastman, G. Euskirchen, S. Frietze, Y. Fu, J. Gertz, F. Grubert, A. Harmanci, P. Jain, M. Kasowski, P. Lacroute, J. Leng, J. Lian, H. Monahan, H. O’Geen, Z. Ouyang, E. C. Partridge, D. Patacsil, F. Pauli, D. Raha, L. Ramirez, T. E. Reddy, B. Reed, M. Shi, T. Slifer, J. Wang, L. Wu, X. Yang, K. Y. Yip, G. Zilberman-Schapira, S. Batzoglou, A. Sidow, P. J. Farnham, R. M. Myers, S. M. Weissman, and M. Snyder. Architecture of the human regulatory network derived from ENCODE data. *Nature*, 489(7414):91–100, sep 2012.
- [111] F. Gilardi, E. Migliavacca, A. Naldi, M. Baruchet, D. Canella, G. Le Martelot, N. Guex, B. Desvergne, and X. C. Cyli. Genome-wide analysis of SREBP1 activity around the clock reveals its combined dependency on nutrient and circadian signals. *PLoS Genet*, 10(3):e1004155, 2014.
- [112] S. Gill and S. Panda. A smartphone app reveals erratic diurnal eating patterns in humans that can be modulated for health benefits. *Cell Metabolism*, 22(5):789–798, nov 2015.
- [113] D. A. Golombek and R. E. Rosenstein. Physiology of circadian entrainment. *Physiological Reviews*, 90(3):1063–1102, jul 2010.
- [114] C. E. Grant, T. L. Bailey, and W. S. Noble. FIMO: scanning for occurrences of a given motif. *Bioinformatics*, 27(7):1017–1018, 2011.
- [115] C. E. Grant, J. Johnson, T. L. Bailey, and W. S. Noble. MCAST: scanning for cis-regulatory motif clusters. *Bioinformatics*, 32(8):1217–1219, dec 2015.
- [116] B. M. Gross DN, van den Heuvel AP. The role of foxo in the regulation of metabolism. *Oncogene*, 2008.

## Bibliography

---

- [117] R. Gualdi, P. Bossard, M. Zheng, Y. Hamada, J. R. Coleman, and K. S. Zaret. Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. *Genes & Development*, 10(13):1670–1682, jul 1996.
- [118] C. Gubelmann, P. C. Schwalie, S. K. Raghav, E. Röder, T. Delessa, E. Kiehlmann, S. M. Waszak, A. Corsinotti, G. Udin, W. Holcombe, G. Rudofsky, D. Trono, C. Wolfrum, and B. Deplancke. Identification of the transcription factor ZEB1 as a central component of the adipogenic gene regulatory network. *eLife*, 3, aug 2014.
- [119] M. G. Guenther, S. S. Levine, L. A. Boyer, R. Jaenisch, and R. A. Young. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell*, 130(1):77–88, 2007.
- [120] F. Guillaumond. Differential control of *bmal1* Circadian transcription by REV-ERB and ROR nuclear receptors. *Journal of Biological Rhythms*, 20(5):391–403, oct 2005.
- [121] S. Gupta, J. A. Stamatoyannopoulos, T. L. Bailey, and W. Noble. Quantifying similarity between motifs. *Genome Biol*, 8(2):R24, 2007.
- [122] E. G. Gusmao, M. Allhoff, M. Zenke, and I. G. Costa. Analysis of computational footprinting methods for DNase sequencing experiments. *Nature Methods*, 13(4):303–309, feb 2016.
- [123] E. G. Gusmao, C. Dieterich, M. Zenke, and I. G. Costa. Detection of active transcription factor binding sites with the combination of DNase hypersensitivity and histone modifications. *Bioinformatics*, 30(22):3143–3151, aug 2014.
- [124] O. Hakim, M. H. Sung, T. C. Voss, E. Splinter, S. John, P. J. Sabo, R. E. Thurman, J. A. Stamatoyannopoulos, W. de Laat, and G. L. Hager. Diverse gene reprogramming events occur in the same spatial clusters of distal regulatory elements. *Genome Res*, 21(5):697–706, 2011.
- [125] T. Håndstad, M. B. Rye, F. Drabløs, and P. Sætrom. A ChIP-seq benchmark shows that sequence conservation mainly improves detection of strong transcription factor binding sites. *PLoS ONE*, 6(4):e18430, apr 2011.
- [126] R. Hardeland, D. P. Cardinali, V. Srinivasan, D. W. Spence, G. M. Brown, and S. R. Pandi-Perumal. Melatonin—a pleiotropic, orchestrating regulator molecule. *Progress in Neurobiology*, 93(3):350–384, mar 2011.
- [127] J. W. Hastings and B. M. Sweeney. A persistent diurnal rhythm of luminescence in *Gonyaulax polyedra*. *Biological Bulletin*, 115(3):440, dec 1958.
- [128] M. H. Hastings, E. S. Maywood, and A. B. Reddy. Two decades of circadian time. *J Neuroendocrinol*, 20(6):812–819, 2008.

- [129] M. H. Hastings, A. B. Reddy, D. G. McMahon, and E. S. Maywood. Analysis of circadian mechanisms in the suprachiasmatic nucleus by transgenesis and biolistic transfection. In *Methods in Enzymology*, pages 579–592. Elsevier BV, 2005.
- [130] M. Hatori, C. Vollmers, A. Zarrinpar, L. DiTacchio, E. A. Bushong, S. Gill, M. Leblanc, A. Chaix, M. Joens, J. A. Fitzpatrick, M. H. Ellisman, and S. Panda. Time-restricted feeding without reducing caloric intake prevents metabolic diseases in mice fed a high-fat diet. *Cell Metabolism*, 15(6):848–860, jun 2012.
- [131] H. H. He, C. A. Meyer, M. W. Chen, V. C. Jordan, M. Brown, and X. S. Liu. Differential DNase I hypersensitivity reveals factor-dependent chromatin dynamics. *Genome Res*, 22(6):1015–1025, 2012.
- [132] H. H. He, C. A. Meyer, S. S. Hu, M. W. Chen, C. Zang, Y. Liu, P. K. Rao, T. Fei, H. Xu, H. Long, X. S. Liu, and M. Brown. Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. *Nat Methods*, 11(1):73–78, 2014.
- [133] W. He, Y. Barak, A. Hevener, O. P. L. D., and L. J. et al. Adipose-specific peroxisome proliferator-activated receptor gamma knockout causes insulin resistance in fat and liver but not in muscle. *Proc Natl Acad Sci USA.*, 2003.
- [134] N. D. Heintzman, G. C. Hon, R. D. Hawkins, P. Kheradpour, A. Stark, L. F. Harp, Z. Ye, L. K. Lee, R. K. Stuart, C. W. Ching, K. A. Ching, J. E. Antosiewicz-Bourget, H. Liu, X. Zhang, R. D. Green, V. V. Lobanenko, R. Stewart, J. A. Thomson, G. E. Crawford, M. Kellis, and B. Ren. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, 459(7243):108–112, 2009.
- [135] S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, and C. K. Glass. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and b cell identities. *Molecular Cell*, 38(4):576–589, may 2010.
- [136] D. V. Herrmann, HC. The feedback regulation of angiotensinogen production by component of the renin-angiotensin system. *Circ Res*, 1983.
- [137] S. Herzig, F. Long, U. S. Jhala, S. Hedrick, R. Quinn, A. Bauer, D. Rudolph, G. Schutz, C. Yoon, P. Puigserver, B. Spiegelman, and M. Montminy. Creb regulates hepatic gluconeogenesis through the coactivator pgc-1. *Nature*, 413(6852):179–183, sep 2001.
- [138] J. R. Hesselberth, X. Chen, Z. Zhang, P. J. Sabo, R. Sandstrom, A. P. Reynolds, R. E. Thurman, S. Neph, M. S. Kuehn, W. S. Noble, S. Fields, and J. A. Stamatoyannopoulos. Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat Methods*, 6(4):283–289, 2009.
- [139] J. B. Hogenesch, W. K. Chan, V. H. Jackiw, R. C. Brown, Y.-Z. Gu, M. Pray-Grant, G. H. Perdew, and C. A. Bradfield. Characterization of a subset of the basic-helix-loop-helix-

## Bibliography

---

- PAS superfamily that interacts with components of the dioxin signaling pathway. *Journal of Biological Chemistry*, 272(13):8581–8593, mar 1997.
- [140] Y. Hu, A. Shmygelska, D. Tran, N. Eriksson, J. Y. Tung, and D. 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.
- [141] N. Huang, Y. Chelliah, Y. Shan, C. A. Taylor, S. H. Yoo, C. Partch, C. B. Green, H. Zhang, and J. S. Takahashi. Crystal structure of the heterodimeric CLOCK:BMAL1 transcriptional activator complex. *Science*, 337(6091):189–194, 2012.
- [142] J. R. Hughes, N. Roberts, S. McGowan, D. Hay, E. Giannoulatou, M. Lynch, M. D. Gobbi, S. Taylor, R. Gibbons, and D. R. Higgs. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nature Genetics*, 46(2):205–212, jan 2014.
- [143] M. Hughes, L. Deharo, S. R. Pulivarthy, J. Gu, K. Hayes, S. Panda, and J. B. Hogenesch. High-resolution time course analysis of gene expression from pituitary. *Cold Spring Harb Symp Quant Biol*, 72:381–386, 2007.
- [144] M. E. Hughes, L. DiTacchio, K. R. Hayes, C. Vollmers, S. Pulivarthy, J. E. Baggs, S. Panda, and J. B. Hogenesch. Harmonics of circadian gene transcription in mammals. *PLoS Genetics*, 2009.
- [145] M. E. Hughes, J. B. Hogenesch, and K. Kornacker. JTK\_CYCLE: An efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *Journal of Biological Rhythms*, 25(5):372–380, sep 2010.
- [146] M. E. Hughes, H. K. Hong, J. L. Chong, A. a. Indacochea, S. S. Lee, M. Han, J. S. Takahashi, and J. B. Hogenesch. Brain-specific rescue of Clock reveals system-driven transcriptional rhythms in peripheral tissue. *PLoS Genetics*, 8(7), 2012.
- [147] A. L. Hutchison, M. Maienschein-Cline, A. H. Chiang, S. M. A. Tabei, H. Gudjonson, N. Bahroos, R. Allada, and A. R. Dinner. Improved Statistical Methods Enable Greater Sensitivity in Rhythm Detection for Genome-Wide Data. *PLOS Computational Biology*, 11(3):e1004094, 2015.
- [148] K. Iizuka, R. K. Bruick, G. Liang, J. D. Horton, and K. Uyeda. From the cover: Deficiency of carbohydrate response element-binding protein (ChREBP) reduces lipogenesis as well as glycolysis. *Proceedings of the National Academy of Sciences*, 101(19):7281–7286, apr 2004.
- [149] I. Inoue, Y. Shinoda, M. Ikeda, K. Hayashi, K. Kanazawa, M. Nomura, T. Matsunaga, H. Y. Xu, S. Kawai, T. Awata, T. Komoda, and S. Katayama. Clock/bmal1 is involved in lipid metabolism via transactivation of the peroxisome proliferator-activated receptor (ppar) response element. *Journal of Atherosclerosis and Thrombosis*, 2005.

- [150] R. A. Irizarry, B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis, U. Scherf, and T. P. Speed. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4(2):249–264, 2003.
- [151] M. M. Jepson, J. M. Pell, P. C. Bates, and D. J. Millward. The effects of endotoxaemia on protein metabolism in skeletal muscle and liver of fed and fasted rats. *Biochemical Journal*, 1986.
- [152] H. Ji, H. Jiang, W. Ma, D. S. Johnson, R. M. Myers, and W. H. Wong. An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat Biotechnol*, 26(11):1293–1300, nov 2008.
- [153] C. Jin, C. Zang, G. Wei, K. Cui, W. Peng, K. Zhao, and G. Felsenfeld. H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat Genet*, 41(8):941–945, 2009.
- [154] S. Jitrapakdee. Transcription factors and coactivators controlling nutrient and hormonal regulation of hepatic gluconeogenesis. *Int J Biochem Cell Biol*, 44(1):33–45, 2012.
- [155] S. John, P. J. Sabo, R. E. Thurman, M.-H. Sung, S. C. Biddie, T. A. Johnson, G. L. Hager, and J. A. Stamatoyannopoulos. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nature Genetics*, 43(3):264–268, jan 2011.
- [156] A. Jolma, J. Yan, T. Whittington, J. Toivonen, K. R. Nitta, P. Rastas, E. Morgunova, M. Enge, M. Taipale, G. Wei, K. Palin, J. M. Vaquerizas, R. Vincentelli, N. M. Luscombe, T. R. Hughes, P. Lemaire, E. Ukkonen, T. Kivioja, and J. Taipale. DNA-binding specificities of human transcription factors. *Cell*, 152(1-2):327–339, 2013.
- [157] C. Jouffe, G. Cretenet, L. Symul, E. Martin, F. Atger, F. Naef, and F. Gachon. The circadian clock coordinates ribosome biogenesis. *PLoS Biol*, 11(1):e1001455, 2013.
- [158] L. KA, P. SJ, Y. RT, B. GD, U. NH, and J. J. et al. Cryptochromes mediate rhythmic repression of the glucocorticoid receptor. *Nature.*, 2011.
- [159] L. KA, S. UM, D. L, W. EC, A. JG, E. DE, V. DS, J. H, P. S, S. RJ, T. CB, and E. RM. Ampk regulates the circadian clock by cryptochrome phosphorylation and degradation. *Science.*, 2009.
- [160] R. Kageyama, S. Yoshiura, Y. Masamizu, and Y. Niwa. Ultradian oscillators in somite segmentation and other biological events. *Cold Spring Harbor Symposia on Quantitative Biology*, 72(1):451–457, jan 2007.
- [161] M. A. Kallio, J. T. Tuimala, T. Hupponen, P. Klemelä, M. Gentile, I. Scheinin, M. Koski, J. Käki, and E. I. Korpelainen. Chipster: user-friendly analysis software for microarray and other high-throughput data. *BMC Genomics*, 12(1):507, 2011.
- [162] A. Kalsbeek, S. la Fleur, and E. Fliers. Circadian control of glucose metabolism. *Molecular Metabolism*, 3(4):372–383, jul 2014.

## Bibliography

---

- [163] S. Katada and P. Sassone-Corsi. The histone methyltransferase MLL1 permits the oscillation of circadian gene expression. *Nat Struct Mol Biol*, 17(12):1414–1421, nov 2010.
- [164] P. V. Kharchenko, M. Y. Tolstorukov, and P. J. Park. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol*, 26(12):1351–1359, nov 2008.
- [165] D. E. Kim, D. Chivian, and D. Baker. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res*, 32(Web Server issue):W526–31, 2004.
- [166] D. H. Kim, G. Perdomo, T. Zhang, S. Slusher, S. Lee, B. E. Phillips, Y. Fan, N. Giannoukakis, R. Gramignoli, S. Strom, S. Ringquist, and H. H. Dong. FoxO6 integrates insulin signaling with gluconeogenesis in the liver. *Diabetes*, 60(11):2763–2774, 2011.
- [167] D. P. King, Y. Zhao, A. M. Sangoram, L. D. Wilsbacher, M. Tanaka, M. P. Antoch, T. D. Steeves, M. H. Vitaterna, J. M. Kornhauser, P. L. Lowrey, F. W. Turek, and J. S. Takahashi. Positional cloning of the mouse circadian clock gene. *Cell*, 89(4):641–653, may 1997.
- [168] A. Klarsfeld. At the dawn of chronobiology. Technical report, ESPCI ParisTech, 2013.
- [169] M. L. Ko, L. Shi, J.-Y. Tsai, M. E. Young, N. Neuendorff, D. J. Earnest, and G. Y.-P. Ko. Cardiac-specific mutation of clock alters the quantitative measurements of physical activities without changing behavioral circadian rhythms. *Journal of Biological Rhythms*, 26(5):412–422, sep 2011.
- [170] F. Koch, R. Fenouil, M. Gut, P. Cauchy, T. K. Albert, J. Zacarias-Cabeza, S. Spicuglia, A. L. de la Chapelle, M. Heidemann, C. Hintermair, D. Eick, I. Gut, P. Ferrier, and J. C. Andrau. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat Struct Mol Biol*, 18(8):956–963, 2011.
- [171] N. Koike, S. H. Yoo, H. C. Huang, V. Kumar, C. Lee, T. K. Kim, and J. S. Takahashi. Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science*, 338(6105):349–354, 2012.
- [172] R. J. Konopka and S. Benzer. Clock mutants of drosophila melanogaster. *Proc Natl Acad Sci*, 1971.
- [173] K. S. Konturek PC, Brzozowski T. Gut clock: implication of circadian rhythms in the gastrointestinal tract. *J Physiol Pharmacol. Apr*;62(2):139-50., 2011.
- [174] A. Korkmaz, T. Topal, D.-X. Tan, and R. J. Reiter. Role of melatonin in metabolic regulation. *Rev Endocr Metab Disord*, 10(4):261–270, nov 2009.
- [175] B. Kornmann, O. Schaad, H. Bujard, J. S. Takahashi, and U. Schibler. System-driven and oscillator-dependent circadian transcription in mice with a conditionally active liver clock. *PLoS Biol*, 5(2):e34, 2007.



- [176] B. Kornmann, O. Schaad, H. Reinke, C. Saini, and U. Schibler. Regulation of circadian gene expression in liver by systemic signals and hepatocyte oscillators. *Cold Spring Harb Symp Quant Biol*, 72:319–330, 2007.
- [177] J. T. Kost and M. P. McDermott. Combining dependent p-values. *Statistics & Probability Letters*, 60(2):183–190, nov 2002.
- [178] M. Koziróg, A. R. Poliwczak, P. Duchnowicz, M. Koter-Michalak, J. Sikora, and M. Broncel. Melatonin treatment improves blood pressure, lipid profile, and parameters of oxidative stress in patients with metabolic syndrome. *Journal of Pineal Research*, 50(3):261–266, dec 2010.
- [179] G. Kramer. Experiments on bird orientation. *Naturwissen- schaften*, 99(2):196–227, apr 1952.
- [180] S. J. Kuhlman, S. R. Mackey, and J. F. Duffy. Biological rhythms workshop i: Introduction to chronobiology. *Cold Spring Harbor Symposia on Quantitative Biology*, 72(1):1–6, jan 2007.
- [181] I. V. Kulakovskiy, Y. A. Medvedeva, U. Schaefer, A. S. Kasianov, I. E. Vorontsov, V. B. Bajic, and V. J. Makeev. HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. *Nucleic Acids Research*, 41(D1):D195–D202, nov 2012.
- [182] K. Kume, M. J. Zylka, S. Sriram, L. P. Shearman, D. R. Weaver, X. Jin, E. S. Maywood, M. H. Hastings, and S. M. Reppert. mCRY1 and mCRY2 Are essential components of the negative limb of the circadian clock feedback loop. *Cell*, 98(2):193–205, jul 1999.
- [183] A. Kusumi, T. A. Tsunoyama, K. M. Hirose, R. S. Kasai, and T. K. Fujiwara. Tracking single molecules at work in living cells. *Nature Chemical Biology*, 10(7):524–532, jun 2014.
- [184] J. Kähärä and H. Lähdesmäki. BinDNase: a discriminatory approach for transcription factor binding prediction using DNase i hypersensitivity data. *Bioinformatics*, 31(17):2852–2859, may 2015.
- [185] K. A. Lamia, K. F. Storch, and C. J. Weitz. Physiological significance of a peripheral tissue circadian clock. *Proc Natl Acad Sci U S A*, 105(39):15172–15177, 2008.
- [186] B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 10(3):R25, 2009.
- [187] J. Le Lay and K. H. Kaestner. The Fox genes in the liver: from organogenesis to functional integration. *Physiol Rev*, 90(1):1–22, 2010.
- [188] G. Le Martelot, D. Canella, L. Symul, E. Migliavacca, F. Gilardi, R. Liechti, O. Martin, K. Harshman, M. Delorenzi, B. Desvergne, W. Herr, B. Deplancke, U. Schibler, J. Rouge-mont, N. Guex, N. Hernandez, F. Naef, and X. C. Cyli. Genome-wide RNA polymerase

## Bibliography

---

- II profiles and RNA accumulation reveal kinetics of transcription and associated epigenetic changes during diurnal cycles. *PLoS Biol*, 10(11):e1001442, 2012.
- [189] G. Le Martelot, T. Claudel, D. Gatfield, O. Schaad, B. Kornmann, G. Lo Sasso, A. Moschetta, and U. Schibler. REV-ERB $\alpha$  participates in circadian SREBP signaling and bile acid homeostasis. *PLoS Biol*, 7(9):e1000181, 2009.
- [190] N. Le Minh, F. Damiola, F. Tronche, G. Schutz, and U. Schibler. Glucocorticoid hormones inhibit food-induced phase-shifting of peripheral circadian oscillators. *EMBO J*, 20(24):7128–7136, 2001.
- [191] C. C. Lee, B. Zheng, D. W. Larkin, U. Albrecht, Z. S. Sun, M. Sage, G. Eichele, and A. Bradley. The *mper2* gene encodes a functional component of the mammalian circadian clock. *Nature*, 400(6740):169–173, jul 1999.
- [192] D. Lee, J. Le Lay, and K. H. Kaestner. The transcription factor CREB has no non-redundant functions in hepatic glucose metabolism in mice. *Diabetologia*, 57(6):1242–1248, 2014.
- [193] M. Leleu, G. Lefebvre, and J. Rougemont. Processing and analyzing ChIP-seq data: from short reads to regulatory interactions. *Brief Funct Genomics*, 9(5-6):466–476, 2010.
- [194] J. C. Leloup and A. Goldbeter. A model for circadian rhythms in *Drosophila* incorporating the formation of a complex between the PER and TIM proteins. *J Biol Rhythms*, 13(1):70–87, 1998.
- [195] Q. Li, J. B. Brown, H. Huang, and P. J. Bickel. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.*, 5(3):1752–1779, sep 2011.
- [196] S. Lichtsteiner, J. Wuarin, and U. Schibler. The interplay of DNA-binding proteins on the promoter of the mouse albumin gene. *Cell*, 51(6):963–973, 1987.
- [197] E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293, oct 2009.
- [198] E. R. Lindahl. *Molecular Dynamics Simulations*. Springer, Humana Press, 2008.
- [199] G. Ling, A. Sugathan, T. Mazor, E. Fraenkel, and D. J. Waxman. Unbiased, genome-wide in vivo mapping of transcriptional regulatory elements reveals sex differences in chromatin structure associated with sex-specific liver gene expression. *Mol Cell Biol*, 30(23):5531–5544, 2010.
- [200] C. Liu, S. Li, T. Liu, J. Borjigin, and J. D. Lin. Transcriptional coactivator PGC-1 integrates the mammalian clock and energy metabolism. *Nature*, 447(7143):477–481, may 2007.



- [201] Y. Liu, W. Hu, Y. Murakawa, J. Yin, G. Wang, M. Landthaler, and J. Yan. Cold-induced RNA-binding proteins regulate circadian gene expression by controlling alternative polyadenylation. *Sci. Rep.*, 3, jun 2013.
- [202] Y.-H. E. Loh and L. Shen. Analysis and visualization of ChIP-seq and RNA-seq sequence alignments using ngs.plot. In *Methods in Molecular Biology*, pages 371–383. Springer Science Business Media, 2016.
- [203] P. L. Lowrey, K. Shimomura, M. P. Antoch, S. Yamazaki, P. D. Zemenides, M. R. Ralph, M. Menaker, and J. S. Takahashi. Positional syntenic cloning and functional characterization of the mammalian circadian mutation tau. *Science*, 288(5465):483–492, 2000.
- [204] P. L. Lowrey and J. S. Takahashi. Genetics of the mammalian circadian system: Photoc entrainment, circadian pacemaker mechanisms, and posttranslational regulation. *Annu Rev Genet*, 34:533–562, 2000.
- [205] X. J. Lu and W. K. Olson. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res*, 31(17):5108–5121, 2003.
- [206] M. Lupien, J. Eeckhoutte, C. A. Meyer, Q. Wang, Y. Zhang, W. Li, J. S. Carroll, X. S. Liu, and M. Brown. FoxA1 Translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell*, 132(6):958–970, mar 2008.
- [207] A. Lusser and J. T. Kadonaga. Chromatin remodeling by ATP-dependent molecular machines. *Bioessays*, 25(12):1192–1200, nov 2003.
- [208] V. Lyssenko, C. L. F. Nagorny, M. R. Erdos, N. Wierup, A. Jonsson, P. Spégel, M. Bugliani, R. Saxena, M. Fex, N. Pulizzi, B. Isomaa, T. Tuomi, P. Nilsson, J. Kuusisto, J. Tuomilehto, M. Boehnke, D. Altshuler, F. Sundler, J. G. Eriksson, A. U. Jackson, M. Laakso, P. Marchetti, R. M. Watanabe, H. Mulder, and L. Groop. Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. *Nature Genetics*, 41(1):82–88, dec 2008.
- [209] S. Lück, K. Thurley, P. F. Thaben, and P. O. Westermarck. Rhythmic degradation explains and unifies circadian transcriptome and proteome data. *Cell Reports*, 9(2):741–751, oct 2014.
- [210] L. M-D, L. C-M, and W. Z. The role of circadian clocks in metabolic disease. *The Yale Journal of Biology and Medicine.*, 2012.
- [211] K. D. MacIsaac, K. A. Lo, W. Gordon, S. Motola, T. Mazor, and E. Fraenkel. A quantitative model of transcriptional regulation reveals the influence of binding location on expression. *PLoS Comput Biol*, 6(4):e1000773, 2010.
- [212] P. Madrigal and P. Krajewski. Current bioinformatic approaches to identify DNase I hypersensitive sites and genomic footprints from DNase-seq data. *Frontiers in genetics*, 3(October):230, jan 2012.

## Bibliography

---

- [213] L. Magraner-Pardo, V. Pelechano, M. D. Coloma, and V. Tordera. Dynamic remodeling of histone modifications in response to osmotic stress in *Saccharomyces cerevisiae*. *BMC Genomics*, 15:247, 2014.
- [214] B. Marcheva, K. M. Ramsey, E. D. Buhr, Y. Kobayashi, H. Su, C. H. Ko, G. Ivanova, C. Omura, S. Mo, M. H. Vitaterna, J. P. Lopez, L. H. Philipson, C. A. Bradfield, S. D. Crosby, L. JeBailey, X. Wang, J. S. Takahashi, and J. Bass. Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes. *Nature*, 466(7306):627–631, 2010.
- [215] A. Mathelier, O. Fornes, D. J. Arenillas, C. Yu Chen, G. Denay, J. Lee, W. Shi, C. Shyr, G. Tan, R. Worsley-Hunt, A. W. Zhang, F. Parcy, B. Lenhard, A. Sandelin, and W. W. Wasserman. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*, 44(D1):D110–D115, nov 2015.
- [216] A. Mathelier and W. W. Wasserman. The next generation of transcription factor binding site prediction. *PLoS Comput Biol*, 9(9):e1003214, sep 2013.
- [217] V. Matys, O. V. Kel-Margoulis, E. Fricke, I. Liebich, S. Land, A. Barre-Dirrie, I. Reuter, D. Chekmenev, M. Krull, K. Hornischer, N. Voss, P. Stegmaier, B. Lewicki-Potapov, H. Saxel, A. E. Kel, and E. Wingender. TRANSFAC and its module TRANSCOMP: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res*, 34(Database issue):D108–10, 2006.
- [218] D. Mauvoisin, J. Wang, C. Jouffe, E. Martin, F. Atger, P. Waridel, M. Quadroni, F. Gachon, and F. Naef. Circadian clock-dependent and -independent rhythmic proteomes implement distinct diurnal functions in mouse liver. *Proc Natl Acad Sci U S A*, 111(1):167–172, 2014.
- [219] N. J. McGlincy, A. Valomon, J. E. Chesham, E. S. Maywood, M. H. Hastings, and J. Ule. Regulation of alternative splicing by the circadian clock and food related cues. *Genome Biol*, 13(6):R54, 2012.
- [220] A. C. Meireles-Filho, A. F. Bardet, J. O. Yáñez-Cuna, G. Stampfel, and A. Stark. cis-regulatory requirements for tissue-specific programs of the circadian clock. *Current Biology*, 24(1):1–10, jan 2014.
- [221] M.-A. Mendoza-Parra, M. Nowicka, W. V. Gool, and H. Gronemeyer. Characterising ChIP-seq binding patterns by model-based peak shape deconvolution. *BMC Genomics*, 14(1):834, 2013.
- [222] J. S. Menet, S. Pescatore, and M. Rosbash. CLOCK:BMAL1 is a pioneer-like transcription factor. *Genes Dev*, 28(1):8–13, 2014.
- [223] J. S. Menet, J. Rodriguez, K. C. Abruzzi, and M. Rosbash. Nascent-Seq reveals novel features of mouse circadian transcriptional regulation. *Elife*, 1:e00011, 2012.

- [224] S. MH, B. S, and H. GL. Genome-wide footprinting: ready for prime time? *nature methods*, 2016.
- [225] V. MH, K. DP, and e. a. Chang A-M. Mutagenesis and mapping of a mouse gene, clock, essential for circadian behavior. *Science (New York, NY)*, 1994.
- [226] B. Mifsud, F. Tavares-Cadete, A. N. Young, R. Sugar, S. Schoenfelder, L. Ferreira, S. W. Wingett, S. Andrews, W. Grey, P. A. Ewels, B. Herman, S. Happe, A. Higgs, E. LeProust, G. A. Follows, P. Fraser, N. M. Luscombe, and C. S. Osborne. Mapping long-range promoter contacts in human cells with high-resolution capture hi-c. *Nature Genetics*, 47(6):598–606, may 2015.
- [227] B. H. Miller, E. L. McDearmon, S. Panda, K. R. Hayes, J. Zhang, J. L. Andrews, M. P. Antoch, J. R. Walker, K. A. Esser, J. B. Hogenesch, and J. S. Takahashi. Circadian and CLOCK-controlled regulation of the mouse transcriptome and cell proliferation. *Proc Natl Acad Sci U S A*, 104(9):3342–3347, 2007.
- [228] A. Mitsui, S. Kumazawa, A. Takahashi, H. Ikemoto, S. Cao, and T. Arai. Strategy by which nitrogen-fixing unicellular cyanobacteria grow photoautotrophically. *Nature*, 1986.
- [229] V. Mooser and C. Currat. The lausanne institutional biobank: A new resource to catalyse research in personalised medicine and pharmaceutical sciences. *Swiss Med Wkly*, 2014.
- [230] J. Morf, G. Rey, K. Schneider, M. Stratmann, J. Fujita, F. Naef, and U. Schibler. Cold-inducible RNA-binding protein modulates circadian gene expression posttranscriptionally. *Science*, 338(6105):379–383, aug 2012.
- [231] F. Mosteller and R. A. Fisher. Questions and answers. *The American Statistician*, 2(5):30, oct 1948.
- [232] A. Mouralidarane, J. Soeda, D. Sugden, A. Bocianowska, R. Carter, S. Ray, R. Saraswati, P. Cordero, M. Novelli, G. Fusai, M. Vinciguerra, L. Poston, P. D. Taylor, and J. A. Oben. Maternal obesity programs offspring non-alcoholic fatty liver disease through disruption of 24-h rhythms in mice. *Int J Obes Relat Metab Disord*, 39(9):1339–1348, may 2015.
- [233] N. U. Nair, A. D. Sahu, P. Bucher, and B. M. E. Moret. ChIPnorm: a statistical method for normalizing and identifying differential regions in histone modification ChIP-seq libraries. *PloS one*, 7(8):e39573, jan 2012.
- [234] Y. Nakahata, M. Yoshida, A. Takano, H. Soma, T. Yamamoto, A. Yasuda, T. Nakatsu, and T. Takumi. A direct repeat of E-box-like elements is required for cell-autonomous circadian rhythm of clock genes. *BMC Mol Biol*, 9:1, 2008.
- [235] M. Nakajima. Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro. *Science*, 308(5720):414–415, apr 2005.

## Bibliography

---

- [236] A. Natarajan, G. G. Yardimci, N. C. Sheffield, G. E. Crawford, and U. Ohler. Predicting cell-type-specific gene expression from regions of open chromatin. *Genome Res*, 22(9):1711–1722, 2012.
- [237] S. Neph, J. Vierstra, A. B. Stergachis, A. P. Reynolds, E. Haugen, B. Vernot, R. E. Thurman, S. John, R. Sandstrom, A. K. Johnson, M. T. Maurano, R. Humbert, E. Rynes, H. Wang, S. Vong, K. Lee, D. Bates, M. Diegel, V. Roach, D. Dunn, J. Neri, A. Schafer, R. S. Hansen, T. Kutuyavin, E. Giste, M. Weaver, T. Canfield, P. Sabo, M. Zhang, G. Balasundaram, R. Byron, M. J. MacCoss, J. M. Akey, M. A. Bender, M. Groudine, R. Kaul, and J. A. Stamatoyannopoulos. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature*, 489(7414):83–90, 2012.
- [238] J. C. Newman and E. Verdin. Ketone bodies as signaling metabolites. *Trends in Endocrinology & Metabolism*, 25(1):42–52, jan 2014.
- [239] H. O’Geen, C. M. Nicolet, K. Blahnik, R. Green, and P. J. Farnham. Comparison of sample preparation methods for ChIP-chip assays. *Biotechniques*, 41(5):577–580, 2006.
- [240] K.-J. Oh, H.-S. Han, M.-J. Kim, and S.-H. Koo. Creb and foxo1: two transcription factors for the regulation of hepatic gluconeogenesis. *BMB Reports*, 46(12):567–574, dec 2013.
- [241] K.-J. Oh, J. Park, S. S. Kim, H. Oh, C. S. Choi, and S.-H. Koo. TCF7l2 modulates glucose homeostasis by regulating CREB- and FoxO1-Dependent transcriptional pathway in the liver. *PLoS Genetics*, 8(9):e1002986, sep 2012.
- [242] K. Oishi, N. Amagai, H. Shirai, K. Kadota, N. Ohkura, and N. Ishida. Genome-wide expression analysis reveals 100 adrenal gland-dependent circadian genes in the mouse liver. *DNA Res*, 12(3):191–202, 2005.
- [243] J. S. O’Neill, E. S. Maywood, J. E. Chesham, J. S. Takahashi, and M. H. Hastings. cAMP-dependent signaling as a core component of the mammalian circadian pacemaker. *Science*, 320(5878):949–953, may 2008.
- [244] J. S. O’Neill and A. B. Reddy. Circadian clocks in human red blood cells. *Nature*, 469(7331):498–503, jan 2011.
- [245] D. A. Orlando, M. W. Chen, V. E. Brown, S. Solanki, Y. J. Choi, E. R. Olson, C. C. Fritz, J. E. Bradner, and M. G. Guenther. Quantitative ChIP-seq normalization reveals global modulation of the epigenome. *Cell Reports*, 9(3):1163–1170, nov 2014.
- [246] M. Pachkov, P. J. Balwierz, P. Arnold, E. Ozonov, and E. van Nimwegen. SwissRegulon, a database of genome-wide annotations of regulatory sites: recent updates. *Nucleic Acids Research*, 41(D1):D214–D220, nov 2012.
- [247] K. Padmanabhan, M. S. Robles, T. Westerling, and C. J. Weitz. Feedback regulation of transcriptional termination by the mammalian circadian clock PERIOD complex. *Science*, 337(6094):599–602, jul 2012.

- [248] S. Panda, M. P. Antoch, B. H. Miller, A. I. Su, A. B. Schook, M. Straume, P. G. Schultz, S. A. Kay, J. S. Takahashi, and J. B. Hogenesch. Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell*, 109(3):307–320, 2002.
- [249] E. R. Paquet, G. Rey, and F. Naef. Modeling an evolutionary conserved circadian cis-element. *PLoS Comput Biol*, 4(2):e38, 2008.
- [250] M. E. Patti, A. J. Butte, S. Crunkhorn, K. Cusi, R. Berria, S. Kashyap, Y. Miyazaki, I. Kohane, M. Costello, R. Saccone, E. J. Landaker, A. B. Goldfine, E. Mun, R. DeFronzo, J. Finlayson, C. R. Kahn, and L. J. Mandarino. Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1. *Proceedings of the National Academy of Sciences*, 100(14):8466–8471, jun 2003.
- [251] J. S. Pendergast, W. Nakamura, R. C. Friday, F. Hatanaka, T. Takumi, and S. Yamazaki. Robust food anticipatory activity in BMAL1-deficient mice. *PLoS One*, 4(3):e4860, 2009.
- [252] M. Petrascheck, D. Escher, T. Mahmoudi, C. P. Verrijzer, W. Schaffner, and A. Barberis. DNA looping induced by a transcriptional enhancer in vivo. *Nucleic Acids Res*, 33(12):3743–3750, 2005.
- [253] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*, 25(13):1605–1612, 2004.
- [254] P. Pevet and E. Challet. Melatonin: Both master clock output and internal time-giver in the circadian clocks network. *Journal of Physiology-Paris*, 105(4-6):170–182, dec 2011.
- [255] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. Scalable molecular dynamics with NAMD. *J Comput Chem*, 26(16):1781–1802, 2005.
- [256] J. Piper, M. C. Elze, P. Cauchy, P. N. Cockerill, C. Bonifer, and S. Ott. Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data. *Nucleic Acids Res*, 41(21):e201, 2013.
- [257] R. Pique-Regi, J. F. Degner, A. A. Pai, D. J. Gaffney, Y. Gilad, and J. K. Pritchard. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res*, 21(3):447–455, 2011.
- [258] E. Portales-Casamar, S. Thongjuea, A. T. Kwon, D. Arenillas, X. Zhao, E. Valen, D. Yusuf, B. Lenhard, W. W. Wasserman, and A. Sandelin. JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res*, 38(Database issue):D105–10, 2010.
- [259] N. Preitner, F. Damiola, L. Lopez-Molina, J. Zakany, D. Duboule, U. Albrecht, and U. Schibler. The orphan nuclear receptor REV-ERB $\alpha$  controls circadian transcription within the positive limb of the mammalian circadian oscillator. *Cell*, 110(2):251–260, 2002.

## Bibliography

---

- [260] L. J. Ptček, C. R. Jones, S. S. Campbell, S. E. Zone, F. Cooper, A. DeSano, P. J. Murphy, B. Jones, and L. Czajkowski. Familial advanced sleep-phase syndrome: A short-period circadian rhythm variant in humans. *Nature Medicine*, 5(9):1062–1065, sep 1999.
- [261] B. Qian, S. Raman, R. Das, P. Bradley, A. J. McCoy, R. J. Read, and D. Baker. High-resolution structure prediction and the crystallographic phase problem. *Nature*, 450(7167):259–264, 2007.
- [262] M. Ralph and M. Menaker. A mutation of the circadian system in golden hamsters. *Science*, 241(4870):1225–1227, sep 1988.
- [263] S. S. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. L. Aiden. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*, 159(7):1665–1680, 2014.
- [264] T. Raveh-Sadka, M. Levo, and E. Segal. Incorporating nucleosomes into thermodynamic models of transcription regulation. *Genome Res*, 19(8):1480–1496, 2009.
- [265] A. B. Reddy, N. A. Karp, E. S. Maywood, E. A. Sage, M. Deery, J. S. O'Neill, G. K. Wong, J. Chesham, M. Odell, K. S. Lilley, C. P. Kyriacou, and M. H. Hastings. Circadian orchestration of the hepatic proteome. *Current Biology*, 16(11):1107–1115, jun 2006.
- [266] A. B. Reddy, E. S. Maywood, N. A. Karp, V. M. King, Y. Inoue, F. J. Gonzalez, K. S. Lilley, C. P. Kyriacou, and M. H. Hastings. Glucocorticoid signaling synchronizes the liver circadian transcriptome. *Hepatology*, 45(6):1478–1488, 2007.
- [267] T. E. Reddy, F. Pauli, R. O. Sprouse, N. F. Neff, K. M. Newberry, M. J. Garabedian, and R. M. Myers. Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. *Genome Res*, 19(12):2163–2171, 2009.
- [268] B. D. Reed, A. E. Charos, A. M. Szekely, S. M. Weissman, and M. Snyder. Genome-wide occupancy of SREBP1 and its partners NFY and SP1 reveals novel functional roles and combinatorial regulation of distinct classes of genes. *PLoS Genetics*, 4(7), 2008.
- [269] J. Reimand, T. Arak, P. Adler, L. Kolberg, S. Reisberg, H. Peterson, and J. Vilo. g:profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res*, page gkw199, apr 2016.
- [270] H. Reinke, C. Saini, F. Fleury-Olela, C. Dibner, I. J. Benjamin, and U. Schibler. Differential display of DNA-binding proteins reveals heat-shock factor 1 as a circadian transcription factor. *Genes Dev*, 22(3):331–345, 2008.
- [271] B. Ren. E2F integrates cell cycle progression with DNA repair, replication, and g2/m checkpoints. *Genes & Development*, 16(2):245–256, jan 2002.
- [272] S. M. Reppert and D. R. Weaver. Coordination of circadian timing in mammals. *Nature*, 418(6901):935–941, 2002.



- 
- [273] G. Rey, F. Cesbron, J. Rougemont, H. Reinke, M. Brunner, and F. Naef. Genome-wide and phase-specific DNA-binding rhythms of BMAL1 control circadian output functions in mouse liver. *PLoS Biol*, 9(2):e1000595, 2011.
- [274] G. Rey and A. B. Reddy. Connecting cellular metabolism to circadian clocks. *Trends in Cell Biology*, 23(5):234–241, may 2013.
- [275] C. M. J. Rice G. A. Kane C. M. Contacts between mammalian RNA polymerase II and the template DNA in a ternary elongation complex. *Nucl. Acids Res.*, 21:113–118, 1992.
- [276] C. Richter. *A Behavioristic Study of the Activity of the Rat*. Comparative psychology monographs, 1922.
- [277] J. A. Ripperger and U. Schibler. Rhythmic CLOCK-BMAL1 binding to multiple E-box motifs drives circadian Dbp transcription and chromatin transitions. *Nat Genet*, 38(3):369–374, 2006.
- [278] J. A. Ripperger, L. P. Shearman, S. M. Reppert, and U. Schibler. CLOCK, an essential pacemaker component, controls expression of the circadian transcription factor DBP. *Genes Dev*, 14(6):679–689, 2000.
- [279] D. Risso, J. Ngai, T. P. Speed, and S. Dudoit. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol*, 32(9):896–902, aug 2014.
- [280] S. A. Rivkees, D. R. Weaver, and S. M. Reppert. Circadian and developmental regulation of Oct-2 gene expression in the suprachiasmatic nuclei. *Brain Res*, 598(1-2):332–336, 1992.
- [281] T. Roenneberg, K. V. Allebrandt, M. Meroow, and C. Vetter. Social jetlag and obesity. *Current Biology*, 22(10):939–943, 2012.
- [282] T. Roenneberg, T. Kuehnle, M. Juda, T. Kantermann, K. Allebrandt, M. Gordijn, and M. Meroow. Epidemiology of the human circadian clock. *Sleep Medicine Reviews*, 11(6):429–438, 2007.
- [283] T. Roenneberg, C. J. Kumar, and M. Meroow. The human circadian clock entrains to sun time. *Current Biology*, 17(2):R44–R45, jan 2007.
- [284] T. Roenneberg, A. Wirz-Justice, and M. Meroow. Life between clocks: Daily temporal patterns of human chronotypes. *Journal of Biological Rhythms*, 18(1):80–90, feb 2003.
- [285] J. Rougemont and F. Naef. Computational analysis of protein–DNA interactions from ChIP-seq data. In *Methods in Molecular Biology*, pages 263–273. Springer Science Business Media, aug 2011.
- [286] R. D. Rudic, P. McNamara, A. M. Curtis, R. C. Boston, S. Panda, J. B. Hogenesch, and G. A. Fitzgerald. BMAL1 and CLOCK, two essential components of the circadian clock, are involved in glucose homeostasis. *PLoS Biol*, 2(11):e377, 2004.

## Bibliography

---

- [287] M. B. Rye, P. Saetrom, and F. Drablos. A manually curated ChIP-seq benchmark demonstrates room for improvement in current peak-finder programs. *Nucleic Acids Research*, 39(4):e25–e25, nov 2010.
- [288] T. Rönn, J. Wen, Z. Yang, B. Lu, Y. Du, L. Groop, R. Hu, and C. Ling. A common variant in MTNR1B, encoding melatonin receptor 1B, is associated with type 2 diabetes and fasting plasma glucose in han chinese individuals. *Diabetologia*, 52(5):830–833, feb 2009.
- [289] E. Sabath, A. Báez-Ruiz, and R. M. Buijs. Non-alcoholic fatty liver disease as a consequence of autonomic imbalance and circadian desynchronization. *Obes Rev*, 16(10):871–882, jul 2015.
- [290] C. Saini, A. Liani, T. Curie, P. Gos, F. Kreppel, Y. Emmenegger, L. Bonacina, J. P. Wolf, Y. A. Poget, P. Franken, and U. Schibler. Real-time recording of circadian liver gene expression in freely moving mice reveals the phase-setting behavior of hepatocyte clocks. *Genes Dev*, 27(13):1526–1536, 2013.
- [291] C. Saini, J. Morf, M. Stratmann, P. Gos, and U. Schibler. Simulated body temperature rhythms reveal the phase-shifting behavior and plasticity of mammalian circadian oscillators. *Genes Dev*, 26(6):567–580, 2012.
- [292] A. Sanyal, B. R. Lajoie, G. Jain, and J. Dekker. The long-range interaction landscape of gene promoters. *Nature*, 489(7414):109–13, sep 2012.
- [293] R. C. Scarpulla, R. B. Vega, and D. P. Kelly. Transcriptional integration of mitochondrial biogenesis. *Trends in Endocrinology & Metabolism*, 23(9):459–466, sep 2012.
- [294] I. Schmutz, J. A. Ripperger, S. Baeriswyl-Aebischer, and U. Albrecht. The mammalian clock component period2 coordinates circadian output by interaction with nuclear receptors. *Genes Dev*, 2010.
- [295] S. Schoenfelder, M. Furlan-Magaril, B. Mifsud, F. Tavares-Cadete, R. Sugar, B.-M. Javierre, T. Nagano, Y. Katsman, M. Sakthidevi, S. W. Wingett, E. Dimitrova, A. Dimond, L. B. Edelman, S. Elderkin, K. Tabbada, E. Darbo, S. Andrews, B. Herman, A. Higgs, E. LeProust, C. S. Osborne, J. A. Mitchell, N. M. Luscombe, and P. Fraser. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Research*, 25(4):582–597, mar 2015.
- [296] H. Schrem. Liver-enriched transcription factors in liver function and development. part II: the c/EBPs and d site-binding protein in cell cycle control, carcinogenesis, circadian gene regulation, liver regeneration, apoptosis, and liver-specific gene regulation. *Pharmacological Reviews*, 56(2):291–330, jun 2004.
- [297] E. M. Scott, A. M. Carter, and P. J. Grant. Association between polymorphisms in the clock gene, obesity and the metabolic syndrome in man. *Int J Obes Relat Metab Disord*, 32(4):658–662, dec 2007.



- [298] S. Seok, T. Fu, S. E. Choi, Y. Li, R. Zhu, S. Kumar, X. Sun, G. Yoon, Y. Kang, W. Zhong, J. Ma, B. Kemper, and J. K. Kemper. Transcriptional regulation of autophagy by an FXR-CREB axis. *Nature*, 516(7529):108–111, 2014.
- [299] J. Shandilya and S. G. Roberts. The transcription cycle in eukaryotes: From productive initiation to RNA polymerase II recycling. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1819(5):391–400, may 2012.
- [300] S. Shimba, T. Ogawa, S. Hitosugi, Y. Ichihashi, Y. Nakadaira, M. Kobayashi, M. Tezuka, Y. Kosuge, K. Ishige, Y. Ito, K. Komiyama, Y. Okamatsu-Ogura, K. Kimura, and M. Saito. Deficient of a clock gene, brain and muscle Arnt-like protein-1 (BMAL1), induces dyslipidemia and ectopic fat formation. *PLoS One*, 6(9):e25231, 2011.
- [301] K. Shimomura, V. Kumar, N. Koike, T. K. Kim, J. Chong, E. D. Buhr, A. R. Whiteley, S. S. Low, C. Omura, D. Fenner, J. R. Owens, M. Richards, S. H. Yoo, H. K. Hong, M. H. Vitaterna, J. Bass, M. T. Pletcher, T. Wiltshire, J. Hogenesch, P. L. Lowrey, and J. S. Takahashi. Usf1, a suppressor of the circadian Clock mutant, reveals the nature of the DNA-binding of the CLOCK:BMAL1 complex in mice. *Elife*, 2:e00426, 2013.
- [302] R. Siersbæk, R. Nielsen, S. John, M.-H. Sung, S. Baek, A. Loft, G. L. Hager, and S. Mandrup. Extensive chromatin remodelling and establishment of transcription factor 'hotspots' during early adipogenesis. *The EMBO journal*, 30(8):1459–72, apr 2011.
- [303] S. Simpson and J. J. Galbraith. IV.—observations on the normal temperature of the monkey and its diurnal variation, and on the effect of changes in the daily routine on this variation. *Trans. R. Soc. Edinb.*, 45(01):65–104, jan 1906.
- [304] C. L. Smith and C. L. Peterson. ATP-dependent chromatin remodeling. In *Current Topics in Developmental Biology*, pages 115–148. Elsevier BV, 2004.
- [305] P. Smolen, D. A. Baxter, and J. H. Byrne. A reduced model clarifies the role of feedback loops and time delays in the drosophila circadian oscillator. *Biophysical Journal*, 83(5):2349–2359, nov 2002.
- [306] D. E. Somers. The physiology and molecular bases of the plant circadian clock. *Plant Physiology*, 1999.
- [307] W. W. Soon, M. Hariharan, and M. P. Snyder. High-throughput sequencing for biology and medicine. *Molecular Systems Biology*, 9(1):640–640, apr 2014.
- [308] F. Spitz and E. E. M. Furlong. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet*, 13(9):613–626, aug 2012.
- [309] C. Spyrou, R. Stark, A. G. Lynch, and S. Tavaré. BayesPeak: Bayesian analysis of ChIP-seq data. *BMC Bioinformatics*, 10(1):299, 2009.
- [310] L. Stadler Engel, Dolan., Gouridine, Weintraub. Tissue-Specific DNA Cleavages in the Globin Chromatin Domain Introduced by DNAase1. *Cell*, 20:451–460, 1980.

## Bibliography

---

- [311] J. A. Stamatoyannopoulos, M. Snyder, R. Hardison, B. Ren, T. Gingeras, D. M. Gilbert, M. Groudine, M. Bender, R. Kaul, T. Canfield, E. Giste, A. Johnson, M. Zhang, G. Balasundaram, R. Byron, V. Roach, P. J. Sabo, R. Sandstrom, A. S. Stehling, R. E. Thurman, S. M. Weissman, P. Cayting, M. Hariharan, J. Lian, Y. Cheng, S. G. Landt, Z. Ma, B. J. Wold, J. Dekker, G. E. Crawford, C. A. Keller, W. Wu, C. Morrissey, S. A. Kumar, T. Mishra, D. Jain, M. Byrska-Bishop, D. Blankenberg, B. R. Lajoie, G. Jain, A. Sanyal, K. B. Chen, O. Denas, J. Taylor, G. A. Blobel, M. J. Weiss, M. Pimkin, W. Deng, G. K. Marinov, B. A. Williams, K. I. Fisher-Aylor, G. Desalvo, A. Kiralusha, D. Trout, H. Amrhein, A. Mortazavi, L. Edsall, D. McCleary, S. Kuan, Y. Shen, F. Yue, Z. Ye, C. A. Davis, C. Zaleski, S. Jha, C. Xue, A. Dobin, W. Lin, M. Fastuca, H. Wang, R. Guigo, S. Djebali, J. Lagarde, T. Ryba, T. Sasaki, V. S. Malladi, M. S. Cline, V. M. Kirkup, K. Learned, K. R. Rosenbloom, W. J. Kent, E. A. Feingold, P. J. Good, M. Pazin, R. F. Lowdon, and L. B. Adams. An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol*, 13(8):418, 2012.
- [312] K. F. Storch, O. Lipan, I. Leykin, N. Viswanathan, F. C. Davis, W. H. Wong, and C. J. Weitz. Extensive and divergent circadian gene expression in liver and heart. *Nature*, 417(6884):78–83, 2002.
- [313] M. Stratmann, D. M. Suter, N. Molina, F. Naef, and U. Schibler. Circadian dbp transcription relies on highly dynamic BMAL1-CLOCK interaction with e boxes and requires the proteasome. *Molecular Cell*, 48(2):277–287, oct 2012.
- [314] M. H. Sung, M. J. Guertin, S. Baek, and G. L. Hager. DNase footprint signatures are dictated by factor dynamics and DNA sequence. *Mol Cell*, 56(2):275–285, 2014.
- [315] A. M. Szalkowski and C. D. Schmid. Rapid innovation in ChIP-seq peak-calling algorithms is outdistancing benchmarking efforts. *Briefings in Bioinformatics*, 12(6):626–633, nov 2010.
- [316] Y. Takeda, R. Jothi, V. Birault, and A. M. Jetten. RORgamma directly regulates the circadian expression of clock genes and downstream targets in vivo. *Nucleic Acids Res*, 40(17):8519–8535, 2012.
- [317] C. Taslim, J. Wu, P. Yan, G. Singer, J. Parvin, T. Huang, S. Lin, and K. Huang. Comparative study on ChIP-seq data: normalization and binding pattern characterization. *Bioinformatics*, 25(18):2334–2340, jun 2009.
- [318] P. F. Thaben and P. O. Westermark. Detecting rhythms in time series with RAIN. *Journal of Biological Rhythms*, 29(6):391–400, oct 2014.
- [319] M. Thomas-Chollier, O. Sand, J.-V. Turatsinze, R. Janky, M. Defrance, E. Vervisch, S. Brohee, and J. van Helden. RSAT: regulatory sequence analysis tools. *Nucleic Acids Research*, 36(Web Server):W119–W127, may 2008.
- [320] R. E. Thurman, E. Rynes, R. Humbert, J. Vierstra, M. T. Maurano, E. Haugen, N. C. Sheffield, A. B. Stergachis, H. Wang, B. Vernot, K. Garg, S. John, R. Sandstrom, D. Bates,

- L. Boatman, T. K. Canfield, M. Diegel, D. Dunn, A. K. Ebersol, T. Frum, E. Giste, A. K. Johnson, E. M. Johnson, T. Kutuyavin, B. Lajoie, B. K. Lee, K. Lee, D. London, D. Lotakis, S. Neph, F. Neri, E. D. Nguyen, H. Qu, A. P. Reynolds, V. Roach, A. Safi, M. E. Sanchez, A. Sanyal, A. Shafer, J. M. Simon, L. Song, S. Vong, M. Weaver, Y. Yan, Z. Zhang, Z. Zhang, B. Lenhard, M. Tewari, M. O. Dorschner, R. S. Hansen, P. A. Navas, G. Stamatoyannopoulos, V. R. Iyer, J. D. Lieb, S. R. Sunyaev, J. M. Akey, P. J. Sabo, R. Kaul, T. S. Furey, J. Dekker, G. E. Crawford, and J. A. Stamatoyannopoulos. The accessible chromatin landscape of the human genome. *Nature*, 489(7414):75–82, 2012.
- [321] J. M. Tian and U. Schibler. Tissue-specific expression of the gene encoding hepatocyte nuclear factor 1 may involve hepatocyte nuclear factor 4. *Genes Dev*, 5(12A):2225–2234, 1991.
- [322] G. Tosini, K. Baba, C. K. Hwang, and P. M. Iuvone. Melatonin: An underappreciated player in retinal physiology and pathophysiology. *Experimental Eye Research*, 103:82–89, oct 2012.
- [323] J.-Y. Tsai, P. C. Kienesberger, T. Pulinilkunnil, M. H. Sailors, D. J. Durgan, C. Villegas-Montoya, A. Jahoor, R. Gonzalez, M. E. Garvey, B. Boland, Z. Blasier, T. A. McElfresh, V. Nannegari, C.-W. Chow, W. C. Heird, M. P. Chandler, J. R. B. Dyck, M. S. Bray, and M. E. Young. Direct regulation of myocardial triglyceride metabolism by the cardiomyocyte circadian clock. *Journal of Biological Chemistry*, 285(5):2918–2929, nov 2009.
- [324] F. W. Turek. Obesity and metabolic syndrome in circadian clock mutant mice. *Science*, 308(5724):1043–1045, may 2005.
- [325] M. Ukai-Tadenuma, T. Kasukawa, and H. R. Ueda. Proof-by-synthesis of the transcriptional logic of mammalian circadian clocks. *Nat Cell Biol*, 10(10):1154–1163, 2008.
- [326] M. Ukai-Tadenuma, R. G. Yamada, H. Xu, J. A. Ripperger, A. C. Liu, and H. R. Ueda. Delay in feedback repression by cryptochrome 1 is required for circadian clock function. *Cell*, 144(2):268–281, 2011.
- [327] U. K. Valekunja, R. S. Edgar, M. Oklejewicz, G. T. J. van der Horst, J. S. O’Neill, F. Tamanini, D. J. Turner, and A. B. Reddy. Histone methyltransferase MLL3 contributes to genome-scale circadian transcription. *Proceedings of the National Academy of Sciences*, 110(4):1554–1559, jan 2013.
- [328] A. Valouev, D. S. Johnson, A. Sundquist, C. Medina, E. Anton, S. Batzoglou, R. M. Myers, and A. Sidow. Genome-wide analysis of transcription factor binding sites based on ChIP-seq data. *Nature Methods*, 5(9):829–834, aug 2008.
- [329] A. E. Vercesi, R. F. Castilho, A. J. Kowaltowski, and H. C. F. Oliveira. Mitochondrial energy metabolism and redox state in dyslipidemias. *IUBMB Life*, 59(4):263–268, 2007.
- [330] J. Vierstra, E. Rynes, R. Sandstrom, M. Zhang, T. Canfield, R. S. Hansen, S. Stehling-Sun, P. J. Sabo, R. Byron, R. Humbert, R. E. Thurman, A. K. Johnson, S. Vong, K. Lee, D. Bates,

## Bibliography

---

- F Neri, M. Diegel, E. Giste, E. Haugen, D. Dunn, M. S. Wilken, S. Josefowicz, R. Samstein, K. H. Chang, E. E. Eichler, M. De Bruijn, T. A. Reh, A. Skoultchi, A. Rudensky, S. H. Orkin, T. Papayannopoulou, P. M. Treuting, L. Selleri, R. Kaul, M. Groudine, M. A. Bender, and J. A. Stamatoyannopoulos. Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science*, 346(6212):1007–1012, 2014.
- [331] J. Vierstra and J. A. Stamatoyannopoulos. Genomic footprinting. *Nature Methods*, 13(3):213–221, feb 2016.
- [332] D. M. Virshup, E. J. Eide, D. B. Forger, M. Gallego, and E. V. Harnish. Reversible protein phosphorylation regulates circadian rhythms. *Cold Spring Harb Symp Quant Biol*, 72:413–420, 2007.
- [333] C. Vollmers, S. Gill, L. DiTacchio, S. R. Pulivarthy, H. D. Le, and S. Panda. Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression. *Proc Natl Acad Sci U S A*, 106(50):21453–21458, 2009.
- [334] C. Vollmers, R. J. Schmitz, J. Nathanson, G. Yeo, J. R. Ecker, and S. Panda. Circadian oscillations of protein-coding and regulatory RNAs in a highly dynamic mammalian liver epigenome. *Cell Metab*, 16(6):833–845, 2012.
- [335] G. P. Wagner, K. Kin, and V. J. Lynch. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.*, 131(4):281–285, aug 2012.
- [336] O. J. Walch, A. Cochran, and D. B. Forger. A global quantification of "normal" sleep schedules using smartphone data. *Science Advances*, 2(5):e1501705–e1501705, may 2016.
- [337] G.-Z. Wang, S. L. Hickey, L. Shi, H.-C. Huang, P. Nakashe, N. Koike, B. P. Tu, J. S. Takahashi, and G. Konopka. Cycling transcriptional networks optimize energy utilization on a genome scale. *Cell Reports*, 13(9):1868–1880, dec 2015.
- [338] J. Wang, J. Zhuang, S. Iyer, X. Lin, T. W. Whitfield, M. C. Greven, B. G. Pierce, X. Dong, A. Kundaje, Y. Cheng, O. J. Rando, E. Birney, R. M. Myers, W. S. Noble, M. Snyder, and Z. Weng. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res*, 22(9):1798–1812, 2012.
- [339] Z. Wang, Y. Wu, L. Li, and X. D. Su. Intermolecular recognition revealed by the complex structure of human CLOCK-BMAL1 basic helix-loop-helix domains with E-box DNA. *Cell Res*, 23(2):213–224, 2013.
- [340] W. W. Wasserman and A. Sandelin. Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet*, 5(4):276–287, 2004.
- [341] P. O. Westermark and H. Herzog. Mechanism for 12 hr rhythm generation by the circadian clock. *Cell Rep*, 3(4):1228–1238, 2013.

- [342] C. L. Woodcock and R. P. Ghosh. Chromatin higher-order structure and dynamics. *Cold Spring Harbor Perspectives in Biology*, 2(5):a000596–a000596, apr 2010.
- [343] J. L. Workman. Nucleosome displacement in transcription. *Genes Dev*, 20(15):2009–2017, 2006.
- [344] A. R. Wu, N. F. Neff, T. Kalisky, P. Dalerba, B. Treutlein, M. E. Rothenberg, F. M. Mburu, G. L. Mantalas, S. Sim, M. F. Clarke, and S. R. Quake. Quantitative assessment of single-cell RNA-sequencing methods. *Nature Methods*, 11(1):41–46, oct 2013.
- [345] H. Xu, L. Handoko, X. Wei, C. Ye, J. Sheng, C. L. Wei, F. Lin, and W. K. Sung. A signal-noise model for significance analysis of ChIP-seq with negative control. *Bioinformatics*, 26(9):1199–1204, apr 2010.
- [346] N. Y. S. S, A. G. K. M, and S.-C. P. Circadian control of the nad<sup>+</sup> salvage pathway by clock-sirt1. *Science*, 2009.
- [347] J. Yang and S. A. Ramsey. A DNA shape-based regulatory score improves position-weight matrix-based recognition of transcription factor binding sites. *Bioinformatics*, 31(21):3445–3450, jun 2015.
- [348] R. Yang and Z. Su. Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation. *Bioinformatics*, 26(12):168–174, 2010.
- [349] X. Yang, M. Downes, R. T. Yu, A. L. Bookout, W. He, M. Straume, D. J. Mangelsdorf, and R. M. Evans. Nuclear receptor expression links the circadian clock to metabolism. *Cell*, 126(4):801–810, 2006.
- [350] G. G. Yardimci, C. L. Frank, G. E. Crawford, and U. Ohler. Explicit DNase sequence bias modeling enables high-resolution transcription factor footprint detection. *Nucleic Acids Res*, 42(19):11865–11878, 2014.
- [351] S.-H. Yoo, S. Yamazaki, P. L. Lowrey, K. Shimomura, C. H. Ko, E. D. Buhr, S. M. Siepkha, H.-K. Hong, W. J. Oh, O. J. Yoo, M. Menaker, and J. S. Takahashi. PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. *Proceedings of the National Academy of Sciences*, 101(15):5339–5346, feb 2004.
- [352] F. Yue, Y. Cheng, A. Breschi, J. Vierstra, W. Wu, T. Ryba, R. Sandstrom, Z. Ma, C. Davis, B. D. Pope, Y. Shen, D. D. Pervouchine, S. Djebali, R. E. Thurman, R. Kaul, E. Rynes, A. Kirilusha, G. K. Marinov, B. A. Williams, D. Trout, H. Amrhein, K. Fisher-Aylor, I. Antoshechkin, G. DeSalvo, L.-H. See, M. Fastuca, J. Drenkow, C. Zaleski, A. Dobin, P. Prieto, J. Lagarde, G. Bussotti, A. Tanzer, O. Denas, K. Li, M. A. Bender, M. Zhang, R. Byron, M. T. Groudine, D. McCleary, L. Pham, Z. Ye, S. Kuan, L. Edsall, Y.-C. Wu, M. D. Rasmussen, M. S. Bansal, M. Kellis, C. A. Keller, C. S. Morrissey, T. Mishra, D. Jain, N. Dogan, R. S. Harris, P. Cayting, T. Kawli, A. P. Boyle, G. Euskirchen, A. Kundaje, S. Lin, Y. Lin, C. Jansen, V. S. Malladi,

## Bibliography

---

- M. S. Cline, D. T. Erickson, V. M. Kirkup, K. Learned, C. A. Sloan, K. R. Rosenbloom, B. L. de Sousa, K. Beal, M. Pignatelli, P. Flicek, J. Lian, T. Kahveci, D. Lee, W. J. Kent, M. R. Santos, J. Herrero, C. Notredame, A. Johnson, S. Vong, K. Lee, D. Bates, F. Neri, M. Diegel, T. Canfield, P. J. Sabo, M. S. Wilken, T. A. Reh, E. Giste, A. Shafer, T. Kuttyavin, E. Haugen, D. Dunn, A. P. Reynolds, S. Neph, R. Humbert, R. S. Hansen, M. D. Bruijn, L. Selleri, A. Rudensky, S. Josefowicz, R. Samstein, E. E. Eichler, S. H. Orkin, D. Levasseur, T. Papayannopoulou, K.-H. Chang, A. Skoultschi, S. Gosh, C. Disteché, P. Treuting, Y. Wang, M. J. Weiss, G. A. Blobel, X. Cao, S. Zhong, T. Wang, P. J. Good, R. F. Lowdon, L. B. Adams, X.-Q. Zhou, M. J. Pazin, E. A. Feingold, B. Wold, J. Taylor, A. Mortazavi, S. M. Weissman, J. A. Stamatoyannopoulos, M. P. Snyder, R. Guigo, T. R. Gingeras, D. M. Gilbert, R. C. Hardison, M. A. Beer, and B. Ren. A comparative encyclopedia of DNA elements in the mouse genome. *Nature*, 515(7527):355–364, nov 2014.
- [353] E. E. Zhang, A. C. Liu, T. Hirota, L. J. Miraglia, G. Welch, P. Y. Pongsawakul, X. Liu, A. Atwood, J. W. Huss, J. Janes, A. I. Su, J. B. Hogenesch, and S. A. Kay. A genome-wide RNAi screen for modifiers of the circadian clock in human cells. *Cell*, 139(1):199–210, oct 2009.
- [354] E. E. Zhang, Y. Liu, R. Dentin, P. Y. Pongsawakul, A. C. Liu, T. Hirota, D. A. Nusinow, X. Sun, S. Landais, Y. Kodama, D. A. Brenner, M. Montminy, and S. A. Kay. Cryptochrome mediates circadian regulation of cAMP signaling and hepatic gluconeogenesis. *Nat Med*, 16(10):1152–1156, 2010.
- [355] R. Zhang, N. F. Lahens, H. I. Ballance, M. E. Hughes, and J. B. Hogenesch. A circadian gene expression atlas in mammals: Implications for biology and medicine. *Proceedings of the National Academy of Sciences*, 111(45):16219–16224, 2014.
- [356] X. Zhang, G. Robertson, M. Krzywinski, K. Ning, A. Droit, S. Jones, and R. Gottardo. PICS: Probabilistic inference for ChIP-seq. *Biometrics*, 67(1):151–163, jun 2010.
- [357] Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, and X. S. Liu. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*, 9(9):R137, 2008.
- [358] V. W. Zhou, A. Goren, and B. E. Bernstein. Charting histone modifications and the functional organization of mammalian genomes. *Nature reviews. Genetics*, 12(1):7–18, jan 2011.
- [359] L. J. Zhu, C. Gazin, N. D. Lawson, H. Pages, S. M. Lin, D. S. Lapointe, and M. R. Green. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics*, 11:237, 2010.



# Jonathan Sobel

*PhD in computational biology*

Montchoisi 18  
1006 Lausanne  
Switzerland

+41 78 890 58 30  
✉ jsobel83@hotmail.com



## Education

- 2012–2016 **PhD**, *EPFL*, Lausanne, Computational systems biology.  
Phase specific transcriptional regulation of circadian clock and metabolism in mouse liver, under the supervision of Prof. Felix Naef. Next generation sequencing data analysis, circadian regulatory network, bio-informatics and bio-statistics
- 2010–2012 **Master**, *UNIGE*, Geneva, Bioinformatics and proteomics.  
Endometriosis biomarkers, under the supervision of Dr. Geraldine Canny
- 2007–2009 **Bachelor**, *HES-SO Valais*, Sion, Life Science Technology, Biotechnology Major.  
Expression vector BRET2 Gateway compatible

## Experience

### Vocational

- 2011–2012 **Research assistant**, *Mucosal immunity lab in CHUV*, Lausanne.  
Proteomic study on endometriosis biomarkers
- 2010 **Research assistant**, *Department of physiology at UNIL*, Lausanne.  
Development of a Biosensor for Caspase 3, Production of Plasmids, Transfections, Cell Culture and Microscopy
- 2009 **Research Internship**, *IRO*, Sion.  
Bachelor's diploma work in molecular biology  
Production of expression vector BRET2 Gateway compatible

### Miscellaneous

- 2013–2016 **President**, *ISCB-RSG-Switzerland*, Lausanne.  
International Society for Computational Biology (ISCB) Regional Section Group of the Student Council from Switzerland
- 2013–2016 **Board Member**, *Bio-Science Network*, Lausanne.  
Organization of Lausanne Life Sciences Career Day in 2014 & 2015
- 2015–2016 **Board Member**, *Hackuarium*, Lausanne.  
An open and community-driven citizen biology lab
- 2014–2015 **Writer**, *Blog Bioinfo-fr*.  
An open blog about bioinformatics (in french)

## Conference Presentations and Posters

- May 2016 **Conference of The Society for Research on Biological Rhythms (SRBR)**, *Palm Harbor, Florida*, Talk and Poster.  
Transcriptional regulatory logic of the diurnal cycle in the mouse liver

- May 2015 **Salon des technologies et de l'innovation de Lausanne (STIL)**, *Lausanne*, Talk.  
La bioinformatique en perspective: de la science fondamentale à l'innovation
- October 2014 **Conference of theoretical chrono-biology**, *Berlin*, Talk.  
Chromatin accessibility around the clock
- June 2014 **CIG Symposium, Rhythms in Biology**, *Lausanne*, Poster.  
Chromatin accessibility around the clock
- May 2013 & 2014 **Swiss Institute of Bioinformatics (SIB) research day**, *Bienne*, Poster.  
Phase specific DNase I hypersensitive site in the mouse liver during the circadian cycle
- September 2013 & 2014 **Institute of Bioengineering (IBI) research day**, *Lausanne*, Poster.  
Phase specific DNase I hypersensitive site in the mouse liver during the circadian cycle
- July 2013 **Annual International Conference on Intelligent Systems for Molecular Biology (ISMB/ECCB)**, *Berlin*, Poster.  
Understanding the fine tuning of circadian phase specific gene expression in mouse liver by cis-regulatory modules using chromatin landscape

## Publications

Jonathan Sobel. Endometriosis biomarkers : discovery experiment and contextualization. Master's thesis, University of Geneva, December 2011.

Jonathan Sobel, Patrice Waridel, Ilaria Gori, Manfredo Quadroni, and Geraldine Olivia Canny. Proteome-wide effect of 17-beta-estradiol and lipoxin a4 in an endometriotic epithelial cell line. *Frontiers in Endocrinology*, 6(192), 2016.

Jonathan Aryeh Sobel, Irina Krier, Teemu Andersin, Sunil Raghav, Donatella Canella, Federica Gilardi, Alexandra Styliani Kalantzi, Guillaume Rey, Benjamin Weger, Frederic Gachon, Matteo Dal Peraro, Nouria Hernandez, Ueli Schibler, Bart Deplancke, Felix Naef, and the CycliX consortium. Transcriptional regulatory logic of the diurnal cycle in the mouse liver. *In preparation*, 2016.

## Awards

- PhD Award Merit Award at the SRBR conference in 2016  
Master Award UNIGE Master Award in 2012

## Computer skills

- Office Power Point, Excel, Word, Illustrator, Latex  
Languages R, Python, Bash, Perl, groovy, C++  
Web HTML, Wordpress, Dreamweaver  
Data Analysis Proteomics, NGS (ChIP-seq, RNA-seq, DNase-seq), time-series and differential analysis

## Languages

- French **C2** *Mother tongue*  
English **C1** *Professional skills*



## References

- Prof. Felix Naef (SIB/EPFL)
- Prof. Philipp Bucher (SIB/EPFL)
- Dr. Geraldine Canny (EPFL/CHUV)
- Prof. Ioannis Xenarios (SIB)
- Prof. Ueli Schibler (UNIGE)

## Interests

Martial Arts Judo, Ju-jitsu  
Reading Scientific literature, science popularization  
Traveling Discovery of other cultures  
Bio-hacking Citizen science