

Statistical physics of periodic biological processes

Présentée le 27 janvier 2023

Faculté des sciences de la vie
Unité du Prof. Naef
Programme doctoral en physique

pour l'obtention du grade de Docteur ès Sciences

par

Lorenzo TALAMANCA

Acceptée sur proposition du jury

Prof. R. Houdré, président du jury
Prof. F. Naef, Prof. P. De Los Rios, directeurs de thèse
Prof. D. Rand, rapporteur
Dr T.-H. Collet, rapporteur
Prof. D. Suter, rapporteur

La vita è un problema alle grandi deviazioni.

— Giorgio Parisi

The truth is rarely pure and never simple.

— Oscar Wilde

Красота спасёт мир.

— Ф. М. Достоевский

To my family and my dearest friends ...

Acknowledgements

A doctorate is an important step in life, and many people have come together to make this possible for me. As I studied in Switzerland I will follow the pragmatic approach of the country and start by thanking Felix, without whom my past four years would have been drastically different.

Thank you Felix for all the help, suggestions, and guidance these past four years. I deeply enjoyed the discussions we had, not only strictly about science but also, sometimes, about the philosophical aspects and implications of the work we were doing. I think it wasn't an easy choice to offer me a position, as my very different background probably made you reflect on possibilities of integration in the lab and academic success; I hope you are now confident of the choice.

Another crucial piece of the puzzle is Paolo, who directed me to talk to Felix when I discussed with him during a visit at EPFL. Thank you for pointing me in this direction, and thank you for the insightful discussions at the beginning of my doctorate.

I want to finish the faculty thanks by expressing my gratitude to Gioele. We have had many discussions in the fertile periods of the collaborations, from which I learned a lot. Mainly I learned, at least a bit, a much more practical and implementable way of thinking, which I believe will greatly help me.

Another practical element to move forward in a doctorate is, at least in my opinion, a good lab, socially and scientifically.

Cedric, thank you for all the innumerable, maybe uncountable, discussions we had on a wide variety of topics. It was always interesting and refreshing to hear your opinion on science, football, building PCs, current events and many more. I also want to thank you for all the help, mental and practical, for the writing, submission, and revision of our paper; it was also very good to have someone with whom to share the anxiety of Bryan not answering.

Nick, thank you for all the conversations, scientific and not. It was always very nice go catch a beer at sat and discuss. I consistently appreciated your input, especially when it came as the pungent British humor.

Acknowledgements

And how can I forget the interminable discussions with Colas. With you and Clemence we animated many lab lunches just for pleasure of discussing, and honestly, it was very fun. I was very happy to share the TAs with you two, especially in the pre-covid times. It was very to end the long Tuesday afternoon with an aperò to relax after a long and stressful beginning of the week.

Elena and Maxine, although you arrived recently, your presence has changed the atmosphere in the lab, bringing much needed new energy; I am happy to have met you.

I am overall very happy with the people in the Naef lab and the personal and social interactions I had. Although my family and closest friendships are in Rome, they still had a deep impact on me during the doctorate. These acknowledgments will be in Italian as they are conceptually for them to read.

Il primo ringraziamento va alla mia famiglia, in particolare ai miei genitori e a mia nonna. Mi avete supportato tanto, da prima ancora di sapere dove o cosa avrei effettivamente fatto durante il dottorato. La solidità e tranquillità che mi avete trasmesso mi ha permesso di affrontare questo percorso nel migliore dei modi. Grazie per l'affetto che mi avete sempre dimostrato; non è necessariamente tipico avere un rapporto positivo con la famiglia, in particolare con i genitori; perciò sono davvero contento e riconoscente per il rapporto che abbiamo.

Sara, è sempre bellissimo tornare a Roma e vederti. Hai una quantità di energia e solarità che non ho mai incontrato in nessun altro. Il modo in cui tu riesci a portare forza e sorrisi nella vita delle persone mi stupisce ogni volta che ti parlo.

Ceci, ho un bellissimo ricordo di tante delle nostre chiacchierate, in particolare della serie di discussioni le sere d'estate. Le tue capacità di capire e spiegare le persone è incredibile a mio avviso. Mi hai insegnato tanto senza faticare e senza farlo pesare, è una cosa davvero rara. Tutto ciò che tu mi hai trasmesso mi ha molto aiutato nell'affrontare questa situazione ed una vita nuova, in un altro paese; grazie mille.

Fernanda, non so come sia possibile parlare con te dopo mesi come se fossero passate due ore dall'ultima volta che ci siamo visti. Il tempo passa sempre troppo in fretta quando ci vediamo, soprattutto ora che sono in svizzera. Grazie di tutto l'affetto che mi trasmetti, sapere che tu ci sei mi ha dato tanta forza in questi anni. Matteo, sei una delle amicizie più lunghe e costanti che ho. Ormai ci conosciamo da quasi quindici anni e il legame che si è creato è uno a cui tengo molto. È sempre interessante parlare e discutere, sei una delle poche persone di cui effettivamente mi interessa l'opinione e il ragionamento su ogni argomento. Da discussioni con te io trovo di aver imparato molto; abbiamo avuto discussioni in momenti così diversi

della vita e su temi così disparati. Grazie per un'amicizia che è stata un pilastro in tutti questi anni.

Lavinia, so che non leggerai queste parole, ma ci tengo comunque a scriverle. Sei stata a lungo al mio fianco da prima che cominciassi un dottorato, sei venuta in Svizzera per potermi stare più vicina; mi hai sopportato e supportato in tanti, troppi momenti diversi, anche quando praticamente chiunque se ne sarebbe andato. Abbiamo vissuto insieme in un piccolo appartamento il covid, e tu mi hai mantenuto sano; ho dei bellissimi ricordi di te e degli innumerevoli sorrisi di cui sei stata la fonte. Grazie di tutto questo e di tutto l'amore che mi hai mostrato in questi anni, senza di te non so se sarei qui oggi.

Lisa, con te accanto a me sto concludendo questa avventura dovendo sostenere il peso dell'ansia e della fatica di tutto ciò che porta con sé la conclusione di un lungo momento. Mi hai saputo far ritrovare la felicità dopo un periodo scuro dal quale l'uscita non era ovvia. Grazie per aver riportato luce e vita in me ed avermi accompagnato in questi ultimi e cruciali passi di questo dottorato.

Lausanne, January 16, 2023

Lorenzo

Abstract

Earth rotation around its axis imposes a 24-hours rhythmicity to all life on the planet. Rather than passively responding to these periodic changes, nature has given us an internal timekeeper, the circadian clock, to anticipate to our advantage the fluctuations in the environment. The circadian clock is a cell-autonomous pervasive molecular oscillator with a period of about a day. The core of the clock is a transcriptional-translational negative feedback loop involving a few dozen genes. This set of genes induces 24-hour rhythms in many downstream processes which are responsible for the daily oscillations in behaviour and physiology. A functioning timekeeper has been associated with well-being while disruptions of the clock have been linked with a variety of diseases, including cancer.

To study the properties and behaviour of the circadian clock in mammals the field has relied heavily on animal models and timed animal omics experiments, due to the difficulty of performing relevant human experiments. However, there is a vast set of human data coming from the clinic without a time stamp, so not directly exploitable to study circadian oscillations. Many methods have been proposed to assign time stamps to a set of omics snapshots; we take inspiration from their strengths and flaws to develop a new probabilistic method of circadian phase inference, CHIRAL.

The physiological impact of the circadian clock in humans is not fully characterised. We exploited the Genotype-Tissue Expression project (GTEx), assigning time stamps to the samples using CHIRAL. We used existing relationships among samples to both robustly infer one time stamp per donor, and transfer time information from robust clocks to weaker ones. This procedure allowed us to study human mRNA rhythms in 46 tissues and compare circadian behaviour across sexes and ages. Clock transcripts showed highly conserved phase and amplitude relationships across tissues, and were tightly synchronised across the body. Tissue rhythmic gene expression programs differed in breadth, covering global and tissue-specific functions, including metabolic pathways and systemic responses. The circadian

Abstract

clock structure and amplitude was conserved across sexes and age groups. However, overall gene expression rhythms were highly sex-dimorphic and more sustained in females. Moreover, rhythmic programs dampened with age across the body. Together, our stratified analysis unveiled a rich organization of sex- and age-specific circadian gene expression rhythms in humans.

To study the clock we took advantage of its low dimensional structure. In fact, projecting high dimensional biological data onto low dimensional manifolds is a widespread technique in biology. In particular, it has proved very useful to exploit single cell transcriptome data. In the single cell framework the RNA velocity technique arose. RNA velocity tackles the idea that cells move in the low dimensional manifolds; it infers the future state of a cell combining current spliced and unspliced RNA counts. However, it is still a very new technique, so not all of the possible adjustments have been made. We develop an analytical framework to add constraints to RNA velocity dynamical systems and differential geometry, allowing for trajectory reconstruction and calculation of time distances on the low dimensional manifold. We apply our method to the cell cycle and infer its period across different brain regions in mouse.

Keywords

Circadian clock, chronobiology, chronopharmacology, human rhythmic sex dimorphism, human age dependent circadian rhythms, omics, circadian phase inference, expectation-maximization, RNA velocity, cell cycle.

Sommario

La rotazione della Terra intorno al suo asse impone una ritmicità di 24 ore a tutta la vita sul pianeta. Invece di rispondere passivamente a questi cambiamenti periodici, la natura ci ha fornito un cronometro interno, l'orologio circadiano, per anticipare a nostro vantaggio le fluttuazioni dell'ambiente. L'orologio circadiano è un oscillatore molecolare autonomo presente in quasi ogni cellula con un periodo di circa un giorno. Il cuore dell'orologio è un ciclo con feedback negativo di trascrizione e traduzione che coinvolge poche dozzine di geni. Questo insieme di geni induce ritmi di 24 ore in molti processi a valle che sono responsabili delle oscillazioni giornaliere nel comportamento e nella fisiologia. Un orologio funzionante è stato associato al benessere, mentre le disfunzioni dell'orologio sono state collegate a una serie di malattie, tra cui il cancro. Il ripristino dei ritmi circadiani è stato recentemente tentato per trattare le malattie metaboliche e l'obesità, con risultati promettenti.

Per studiare le proprietà e il comportamento dell'orologio circadiano nei mammiferi ci si è affidati in larga misura a modelli e esperimenti temporali su animali, a causa della difficoltà di eseguire rilevanti esperimenti umani. Tuttavia, esiste una vasta serie di dati umani provenienti dalla clinica senza timbro temporale, quindi non direttamente sfruttabili per studiare le oscillazioni circadiane. Sono stati proposti molti metodi per assegnare timbri temporali ad un insieme di misure dell'abbondanza del trascrittoma; ci ispiriamo ai loro punti di forza e ai loro difetti per sviluppare un nuovo metodo probabilistico di inferenza della fase circadiana, CHIRAL.

L'impatto fisiologico dell'orologio circadiano negli umani non è completamente caratterizzato. Abbiamo sfruttato il progetto Genotype-Tissue Expression (GTEx), assegnando i timbri temporali ai campioni utilizzando CHIRAL. Abbiamo utilizzato le relazioni esistenti tra i campioni per dedurre in modo robusto un solo timbro temporale per donatore e trasferire le informazioni temporali da orologi robusti a orologi più deboli. Questa procedura ci ha permesso di studiare i ritmi dell'mRNA

umano in 46 tessuti e di confrontare il comportamento circadiano tra i sessi e le età. Il trascrittoma dell'orologio ha mostrato relazioni di fase e ampiezza altamente conservate tra i tessuti e sono strettamente sincronizzati in tutto il corpo. I programmi di espressione genica ritmica dei tessuti differivano per ampiezza, coprendo funzioni globali e specifiche dei tessuti, compresi i percorsi metabolici e le risposte sistemiche. La struttura e l'ampiezza dell'orologio circadiano erano conservate nei diversi sessi e gruppi di età. Tuttavia, i ritmi complessivi dell'espressione genica sono risultati altamente più sostenuti nelle femmine. Inoltre, i programmi ritmici si sono attenuati con l'età in tutto il corpo. Nel complesso, la nostra analisi stratificata ha svelato una ricca organizzazione di ritmi di espressione genica circadiana specifici per sesso ed età negli umani.

L'orologio circadiano è un esempio di dati ad alta dimensionalità proiettati su una varietà a bassa dimensionalità. L'uso di una rappresentazione a bassa dimensione dei dati è utile in biologia, in particolare per i dati del trascrittoma di una singola cellula. Nell'ambito delle singole cellule è nata la tecnica della velocità dell'RNA. La velocità dell'RNA affronta l'idea che le cellule si muovano in varietà a bassa dimensionalità; deduce lo stato futuro di una cellula combinando i conteggi attuali di mRNA processato e non. Tuttavia, si tratta di una tecnica ancora molto recente, per cui non sono stati apportati tutti i possibili aggiustamenti. Sviluppiamo una descrizione analitica per aggiungere vincoli dinamici e geometrici alla velocità dell'RNA, consentendo la ricostruzione della traiettoria e il calcolo delle distanze temporali sulle varietà a bassa dimensionalità. Appliciamo il nostro metodo al ciclo cellulare e ne deduciamo il periodo in diverse regioni cerebrali.

Parole chiave

Orologio circadiano, cronobiologia, cronofarmacologia, dimorfismo sessuale ritmico umano, ritmi circadiani umani dipendenti dall'età, omica, inferenza di fase circadiana, massimizzazione delle aspettative, velocità dell'RNA, ciclo cellulare.

Contents

Acknowledgements	i
Abstract (English/Italiano)	v
List of Figures	xiii
List of Tables	xv
I Introduction	1
1 Space, time, and velocity	3
1.1 The time of life	3
1.1.1 A molecular timekeeper	5
1.1.2 Animal experiments around the clock, atlases, and benchmarks	7
1.1.3 Methods of circadian phase inference	8
1.1.4 The clock across tissues	8
1.1.5 Circadian physiology	9
1.1.6 Timed health	12
1.1.7 Circadian sex-dimorphism	14
1.1.8 Age-dependent rhythmic physiology	14
1.2 The space of biological data	16
1.2.1 Low dimensional projections	16
1.3 Dynamics in biology	18
1.3.1 The velocity of cells	19
1.3.2 Cell cycle speed	22
1.4 Direction and aims	25
1.4.1 Circadian phase inference	25
1.4.2 Circadian clock in humans	25
1.4.3 Dynamically consistent RNA-velocity	26

II	Circadian phase inference	27
2	How to tell time: advances in decoding circadian phase from omics snapshots	29
2.1	Abstract	29
2.2	Introduction	30
2.3	Computational methods to infer circadian phase	30
2.4	Applications	33
3	The CHIRAL algorithm	37
3.1	Setting	37
3.2	Hypothesis	38
3.3	Multivariate harmonic regression model	38
3.4	Approximations on the covariance matrices	39
3.5	Rewriting the problem as a spin model	41
3.6	Mean field approximation	41
3.7	The emergence of a phase transition	43
3.8	The two-state model	44
3.9	The EM algorithm	44
3.10	An interesting equivalence	48
3.11	CHIRAL performance	48
3.12	Practicalities	52
III	Human chronobiology	53
4	Sex-dimorphic and age-dependent organization of 24-hour gene expression rhythms in human	55
4.1	Abstract	55
4.2	Introduction	56
4.3	Comprehensive 24h gene expression rhythms in humans	56
4.4	Human sexual dimorphism in circadian rhythms	68
4.5	Age dependent circadian reprogramming of human gene expression	79
4.6	Discussion	91
4.7	Methodological details	92
IV	Dynamically consistent RNA velocity	99
5	RNA velocity 2.0	101
5.1	Introduction	101

5.1.1	The general problem	101
5.1.2	Estimation of cell cycle period	102
5.2	The general model	102
5.2.1	Properties of the velocity field	103
5.3	Measuring cell cycle period	104
5.4	Implementation	109
5.4.1	Probabilistic programming with pyro	109
5.4.2	Preliminary results	110
5.4.3	Next steps	111
V	Conclusion & outlook	115
6	Conclusions and perspectives	117
6.1	Contributions of this thesis	117
6.1.1	Circadian phase inference	117
6.1.2	Human chronobiology	118
6.1.3	Dynamically consistent RNA velocity	120
7	Future directions	121
7.1	Cancer and circadian clock	121
7.2	Timed treatments	123
A	Gaussian integration	127
A.1	Univariate Gaussian integrals	127
A.2	Multivariate gaussian integrals	127
A.3	Particular case	128
B	Technical details for the improvement of RNA velocity	131
B.1	Biological constraints β and γ	131
B.2	Velocity and period	131
B.3	Rotation in the Fourier space	132
B.4	Constraints on γ	133
B.5	Pyro-accepted probability distributions	134
B.6	Scales across datasets	135
	Bibliography	177

List of Figures

1.1	Earth during summer solstice.	4
1.2	Circadian adaptation as a unifying model that integrates behaviour and physiology.	6
1.3	Model of how external rhythmic cues entrain circadian rhythms across tissue.	10
1.4	Circadian physiology throughout the day.	11
1.5	time-restricted feeding can improve health	13
1.6	Measuring dynamic changes in gene expression across complex tissues	20
1.7	Illustrative representation of single-cell temporal-omics approaches	21
1.8	The cell division cycle.	23
2.1	A summary of the steps followed by many phase reconstruction algorithms	34
2.2	A summary of the most important characteristics of the existing methods to decode circadian phase from omics samples	35
3.1	CHIRAL reconstruction using top 10000 expressed genes	49
3.2	CHIRAL reconstruction using clock reference genes	50
3.3	Comparison of CHIRAL and CYCLOPS on human muscle biopsies .	51
4.1	Failure of the TOD to capture circadian phase and necessity of the DIP	57
4.2	Scheme of the steps to infer donor internal phase	58
4.3	Importance of covariates in the GTEx data	59
4.4	Donor specific peaked TIP distribution	60
4.5	First results of DIP inference	60
4.6	Clock reference genes using DIPs	62
4.7	Heat shock response genes using DIPs	63
4.8	Gene rhythmicity across the body	64
4.9	Ultradian rhythms and transcription factor activity	65
4.10	Rhythmic GO terms across the body	67

List of Figures

4.11 Overview of clock, DIP, and rhythms in males and females	69
4.12 Distribution of rhythmicity between sexes	70
4.13 Close up on sex-dimorphism in the heart atrial appendage	71
4.14 Close up on sex-dimorphism in the liver	71
4.15 Heatmap of sex-dimorphic rhythms for heart atrial appendage . . .	72
4.16 Heatmap of sex-dimorphic rhythms for liver	73
4.17 Female-biased liver mRNA rhythms	74
4.18 Sex-dimorphism in cholesterol biosynthesis	75
4.19 Sex-dimorphism in detoxification	76
4.20 Sex-dimorphism in glucocorticoid signaling	77
4.21 Female-biased mRNA rhythms in the adrenal gland	78
4.22 Overview of clock, DIP, and rhythms in young and old	81
4.23 Distribution of rhythmicity between ages	82
4.24 Close up on age-dependent rhythms in adipose subcutaneous . . .	83
4.25 Close up on age-dependent rhythms in the coronary arteries	83
4.26 Heatmap of age-dependent rhythms for adipose subcutaneous . .	84
4.27 Heatmap of age-dependent rhythms for coronary arteries	85
4.28 Age-dependent cholesterol biosynthesis	86
4.29 Age-induced loss of rhythms in the coronary arteries	87
4.30 Age-related change of physiology in the ovary	88
4.31 Loss of metabolic function in elderly's ovary	89
4.32 Gain of heat shock response in elderly's ovary	90
4.33 Emergence of ultradian rhythms with age	91
5.1 Inference of a flat $\omega(\varphi)$	112
5.2 Inference of a phase dependent $\omega(\varphi)$	113
7.1 Clock structure in healthy and cancerous tissues	122



List of Tables

3.1	CHIRAL validation on human muscle biopsies	52
-----	--	----

Introduction **Part I**

1 Space, time, and velocity

1.1 The time of life

Time.

Everyone knows what time is, although to formulate the idea of time, clearly and without ambiguity, is complex. It is such an intrinsic concept that rarely schools or universities try to define it on a philosophical scale. We are taught how to measure it, and the units of time permeate our life. But what is the fundamental unit of time? The one that speaks to our core, as well as to the vast majority of life as we know it. Due to the societal evolution of the western cultures our lives revolve around hours. This starts especially in middle school, where every hour there is a different subject and we have to adapt our minds to this new time unit. Hours dictate our life until at least the end of university, all programs and schedules are indeed based on this societal basic time unit. Even after that, even if our life is now portioned in fractions longer than the hour, like the 9AM to 6PM working day, still the hour has a central role for anything different from working time, such as meetings, sports, events, and gatherings. Physics wholeheartedly disagrees with the hours and adopted the second as the standard time unit [1]. This unit is unchanged among a wide variety of branches of physics studying phenomena on our planet. Yet life does not understand these measurements that humans came up with. Life takes it a bit more slowly. The clear unit for the vast majority of life on our planet is the day. The rotation of the earth around its axis drives the environment to vast changes throughout the day, as encapsulated in figure 1.1. The day houses vast oscillations in light and temperature which are a rudimentary clock for life to follow. It is felt almost everywhere on our planet, and is short enough so that life, all life, from creatures that live months to those that live centuries, need to adapt to it. But

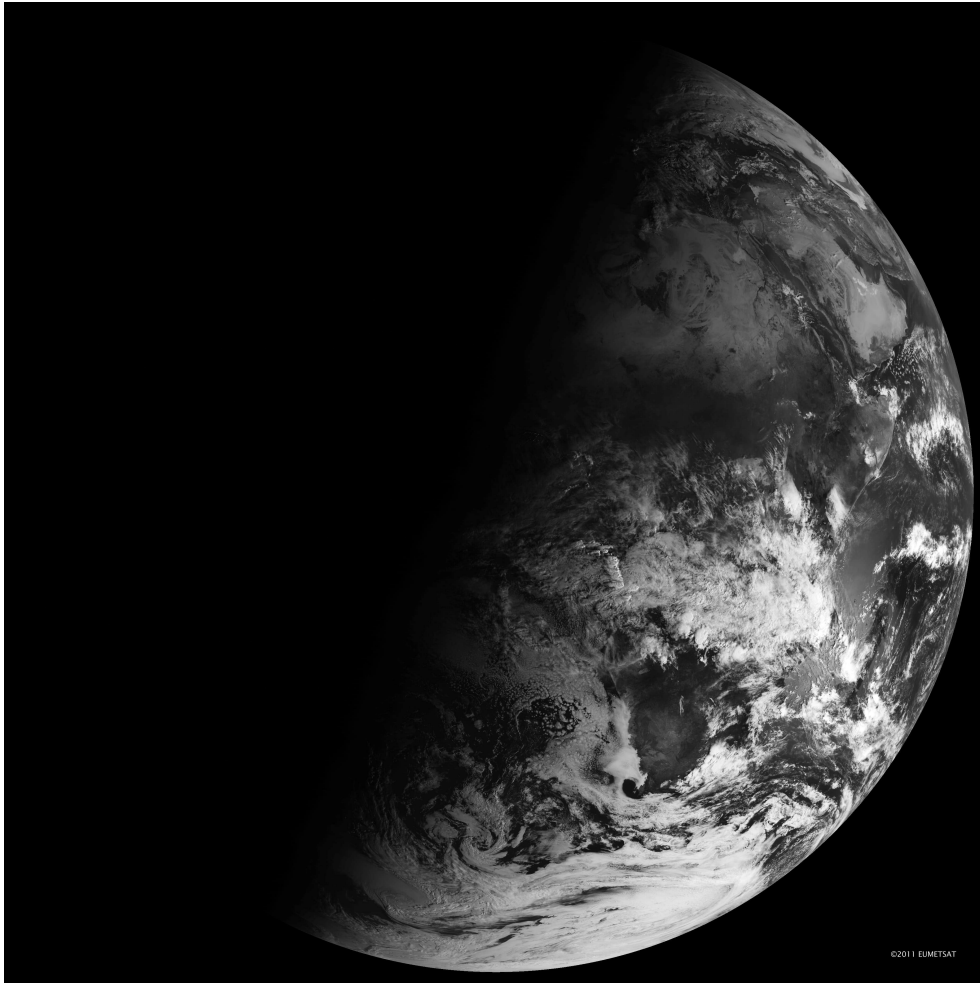


Figure 1.1: **Earth during summer solstice.** Visualization of light differences in the day-night cycle. Photo credit: NASA

a crucial question remains: do we only adapt our behaviours as a response to the oscillations, or did nature create something better, something smarter? Indeed nature found a way to encode the 24-hour rhythms in our DNA, so that we could not only respond to the environmental changes, but anticipate them and use them to our advantage. The circadian, from Latin “circa diem” (about a day), clock [2] is the name we gave to our internal timekeeping system.

1.1.1 A molecular timekeeper

The idea that somehow the 24-hour periodicity is embedded in our behaviours rather than being a mere response to the external variations had to wait until the 18th century to come to light. The french scientist De Mairan removed the daily light oscillations from the environment of a mimosa and showed how the plant still continued to open and close its leaves every 24 hours [3]. This first pioneer could not see the impact of his observation; humanity had to wait two hundred years to witness a great accomplishment of genetics: the identification of the molecular mechanism driving these inner 24-hour rhythms in the fruit fly in 1971 [4]. Fifty years have passed and our knowledge has expanded from plants and flies and we have been able to decode the network of genes responsible for the circadian clock in mammals [5]. Now we know that the circadian clock is a molecular and pervasive cell autonomous oscillator; it ticks in virtually all cells of our bodies [6]. The clock gains its periodicity of almost 24 hours [7] from an as-yet incompletely understood mechanism, which most likely involves a transcriptional–translational negative feedback loop but probably also other regulatory layers [8, 9]. The full molecular mechanism of the clock may remain yet hidden, but the effort to detail it in the last 20 years has led to important advances. A series of key studies [6, 10–12] has shown that the core oscillator of the circadian clock involves a fairly small set (likely in the order of 20) of genes, called core clock genes. The interactions among these core clock genes and some of their most important outputs are depicted in figure 1.2. These genes in turn regulate the temporal expression of other genes, which drive programs of 24 hour tissue-specific rhythms in gene expression and physiology [15, 16]. The fact that only a handful of genes are able to control for such distinct oscillations in physiology and behaviour should be taken as a food for thought, even if we will not discuss the philosophical implications here. Under the control of the circadian clock life has adapted to the oscillations in the environment and has learned to thrive turning to its advantage the intrinsic variability in its surroundings. Sadly, humans in the modern world are under societal constraints (such as timing of social gathering and work schedule) which often interfere with the

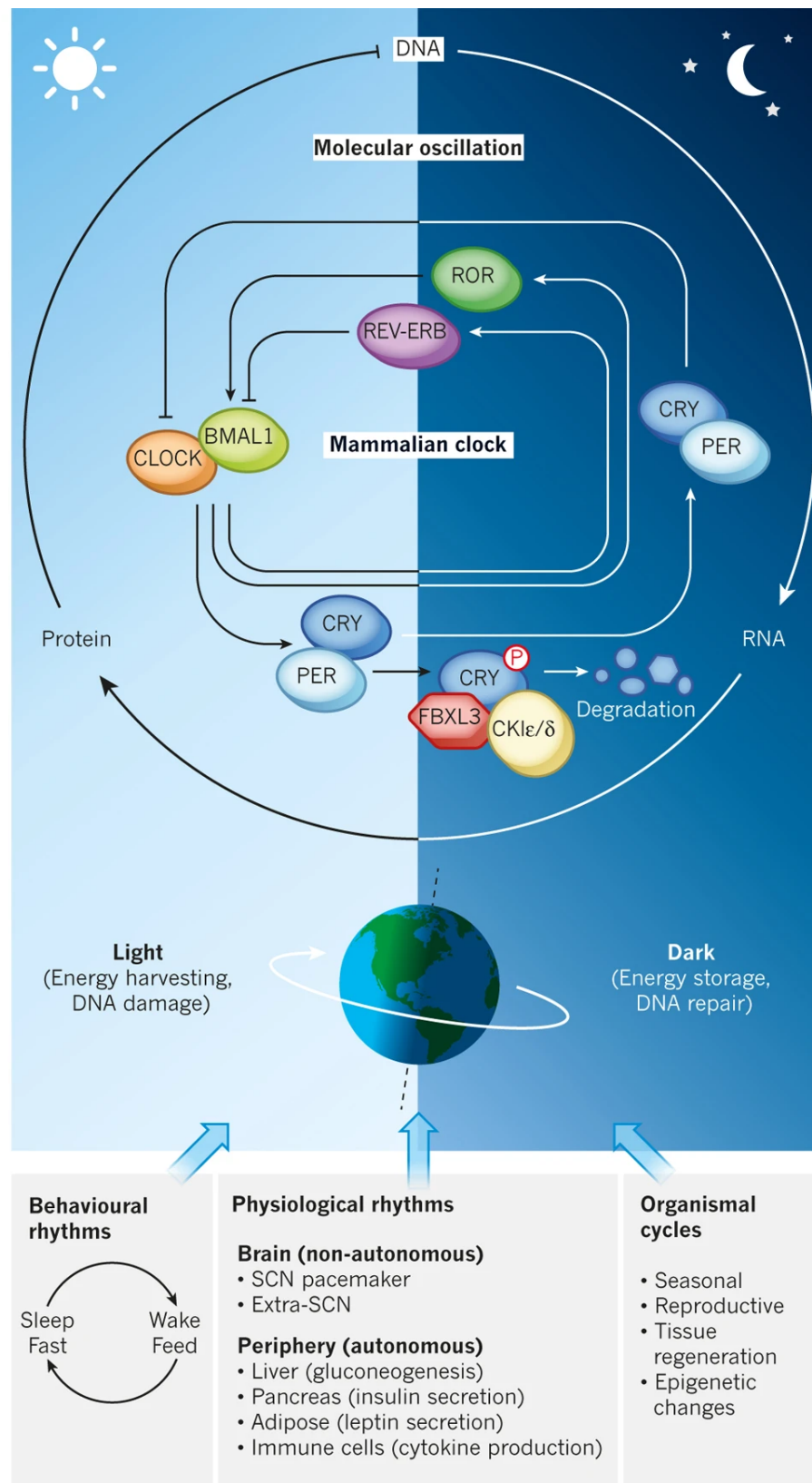


Figure 1.2: **Circadian adaptation as a unifying model that integrates behaviour and physiology.** Caption continued on the next page

Figure 1.2: The circadian clock allows light-sensitive organisms to synchronize their daily molecular oscillations, behavioural rhythms, physiological rhythms and organismal cycles with the rotation of Earth on its axis. Core molecular pathways dictate behavioural and physiological cycles. This core molecular clock in mammals, expressed both in brain and peripheral metabolic tissues, comprises a series of transcription–translation feedback loops that include opposing transcriptional activators (CLOCK–BMAL1) and repressors (PER–CRY) [13]. The non-phosphorylated PER–CRY complex represses CLOCK–BMAL1; phosphorylation, in turn, results in the degradation of PER–CRY and the turnover of these repressors. In addition, CLOCK–BMAL1 induces transcription of REV-ERB and of ROR, which regulate BMAL1 expression. During the night, PER–CRY is degraded through the ubiquitylation of CRY by FBXL3. The circadian clock coordinates anabolic and catabolic processes in peripheral tissues with the daily behavioural cycles of sleep–wake and fasting–feeding. SCN, suprachiasmatic nucleus. Figure and caption from *Circadian topology of metabolism* [14]

natural oscillations of the clock [16, 17]. The vast majority of knowledge pertaining to the clock has been obtained through animal experiments; the sacrifice of animals should not be easily forgotten when discussing biomedical advancements.

1.1.2 Animal experiments around the clock, atlases, and benchmarks

To study the mammalian clock, many experiments spanning the full day have been performed, mostly on mice. The general setup is to sacrifice mice every 2 to 6 hours over 24 hours. Most of these experiments use highly controlled conditions. Notably, the mice are typically of one genetic background and taken from the same gender; moreover, environmental light or temperature conditions are tightly controlled, as well as the feeding regimen and schedules. Although many experiments focus on specific organs or conditions, some transcriptome analyses provide comprehensive views on temporal gene regulation across an entire organism. Five years ago, the first atlas of gene expression around the clock was published [18]. In this experiment, two mice were sacrificed every 2 hours, then 12 tissues were analysed for their mRNA expression levels using RNA sequencing. This seminal dataset has since then served as a benchmark for other around-the-clock experiments on mice [19]. Given the strictly controlled experimental conditions typical of mice experiments, this atlas minimises all the variation in gene expression due to factors different from the circadian clock. More recently, an atlas on baboons [20] has been published in an effort to close the gap between mice and humans. In this case, one baboon was sacrificed every 2 hours and 64 tissues were collected and analysed. As baboons are more similar to humans than mice, these data may in principle provide

a better benchmark for future circadian studies in humans, especially considering the genetic diversity of the sacrificed baboons. In particular, mice are nocturnal animals and there are thus intrinsic differences in clock programs [21], though these are not yet fully characterized. To reduce the animal burden of experiments around the clock, some computational methods have been introduced to infer the circadian phase of omics samples.

1.1.3 Methods of circadian phase inference

Given the abundance of data, computational assignment of circadian phase to omics samples have been suggested. It is important to precise the definition of circadian phase: it is the molecular state of the clock with respect to its cycle. A variety of approaches have been proposed, but the community has not yet recognised a golden standard [15, 22–27]. The vast majority uses prior knowledge on clock gene phase and amplitude to infer the phase of a new sample; this is a solid approach although not necessarily general with respect to tissues or species considered. One method has taken an "unsupervised" approach and has improved on principal component analysis [28] to obtain a widely applicable method. We will discuss more in details these methods and applications, their weaknesses and strengths in chapter 2. The objective for all these methods is ultimately to reorder human samples to have access to the human circadian clock. However, it is important to keep in mind that the mammalian circadian clock has a non-trivial structure across the body.

1.1.4 The clock across tissues

Up to now we only focused on the temporal aspect of the clock. The spatial location of the circadian clock is an important part of the control of our 24-hour rhythmic behaviour. With a vast number of organs (> 20), mammals require precise coordination across the body to fully exploit the timing information given by the circadian clock. In fact, the mammalian structure for time-keeping is hierarchical [29]: a master clock [30] is hosted in the brain's suprachiasmatic nuclei (SCN) and peripheral clocks tick in virtually all of the other organs. The circadian clock's free running period is not exactly 24 hours. If not entrained¹, the human clock will drift from the environment, losing about 15 minutes a day [31]. However, under normal physiological conditions where the organism is subject to external 24 hour

¹Entrained in the circadian clock context means synchronised with the external 24h cycle.

periodic cues (called Zeitgebers in the circadian field), the clock synchronises to external time [32, 33]. The clock in the SCN responds to the external light intensity cycle [34] and synchronises the body's peripheral clocks [35]. Indeed it could be argued that light oscillations are by far the best possible cues (Zeitgebers) for the clock to entrain to. Nature has made sure that the main driver of our coupling to the surrounding environment is the strongest one. The internal positioning of an individual's circadian timekeeper with respect to the external time varies from individual to individual. This variability is an important consequence of the difference between the free running period of the circadian clock and the period of earth's rotation on its axis. The precise position depends on the combination of free running period and amplitude of that individual's clock and, importantly, the Zeitgeber strength such as the intensity of light. The difference (dephasing) between the circadian phase (internal time) and external time is called chronotype and varies in humans with a standard deviation of 2 hours [15, 36–39]. The other interesting phase relationships to explore are the ones among tissues. Especially as peripheral tissues are sensible to cues different than light [33] (figure 1.3) that might contrast with the signals from the clock in the SCN [41]. One relevant example is the clock in the liver: this peripheral clock is so responsive to feeding patterns that it can switch its clock with respect to the brain if the feeding time is reversed [42]. Overall, the network of clocks across the body shapes and regulates our 24-hour rhythmic physiology.

1.1.5 Circadian physiology

The interplay of the clock and our physiology is extensive and particularly relevant for metabolic tissues, as shown in figure 1.4. However, circadian oscillations are more relevant in some of these tissues than others. The liver plays a central role in nutrient metabolism and its transcriptome is highly rhythmic. In entrained mice the about 20% of the protein coding genes show circadian oscillations in the liver [18, 43]. Lipid metabolism in the liver has been shown to be rhythmic, and its oscillations are controlled at various steps [44–46]. The clock indirectly controls lipogenesis and energy homeostasis in the liver; it regulates lipid absorption reducing the risk of obesity and lipid accumulation [47–50]. In addition, circadian oscillations are also present in the intestine, where the clock regulates packing of triglyceride and macro-nutrient absorption [51, 52]. This timing ensures the best use of energetic resources and avoids the utilisation of lipids while glucose is readily available. The interplay of the liver circadian clock and glucose metabolism system has largely been studied [53–57]. While each study focuses on a specific interaction

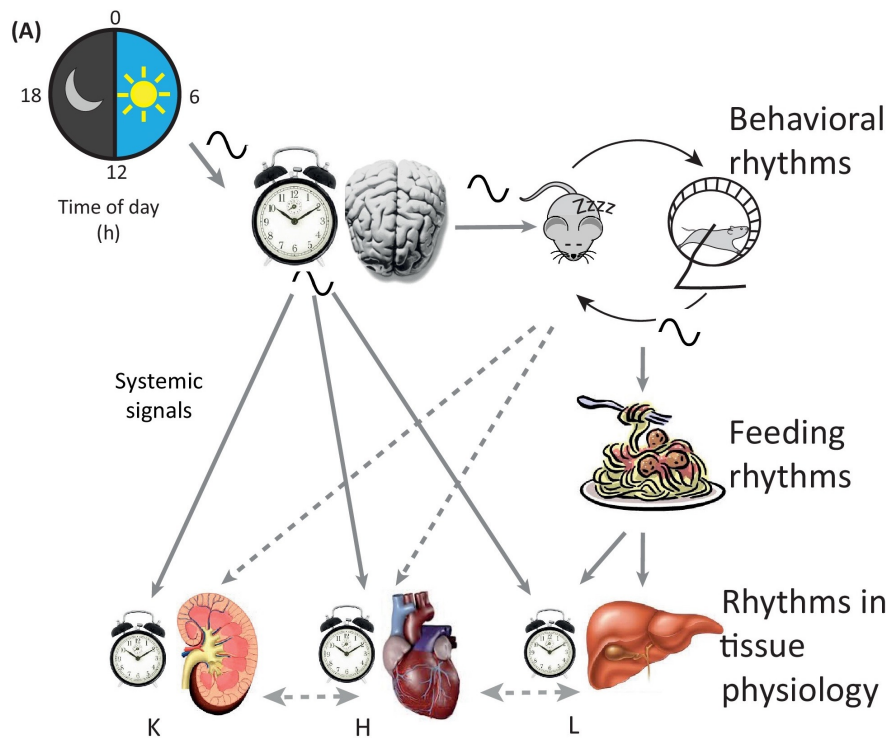


Figure 1.3: **Model of how external rhythmic cues entrain circadian rhythms across tissue.** The central clock, located in the suprachiasmatic nucleus of the hypothalamus, takes light from the environment and synchronises clocks in peripheral tissues by regulating locomotor activity and feeding rhythms as well as through systemic signals such as hormones and metabolites. Peripheral tissues, such as the kidneys, heart, and liver, orchestrate rhythms in tissue physiology, such as sodium homeostasis, blood pressure, and carbohydrate metabolism, respectively. Figure and caption adapted from *Rhythms of the Genome: Circadian Dynamics from Chromatin Topology, Tissue-Specific Gene Expression, to Behavior* [40]

between the clock and glucose metabolism pathways, a clear overall picture arises: to ensure physiological optimality, glucose metabolism needs to be timed during the active phase. On the other hand, if food availability is "unnatural" or unexpected from a circadian perspective the liver clock can adapt to these new rhythms. Circadian glucose level control and homeostasis are particularly relevant for human health, in particular to prevent the insurgence of diabetes [58]. Blood glucose levels are controlled by changes in the secretory capability for insulin of the pancreas, a process downstream of the circadian clock [59–61]. Another important system under the direct control of the circadian clock is the cardiovascular one. The clock is responsible for regulating amplitude and timing of the daily oscillations in blood pressure ensuring a peak in the middle of the active phase [62, 63]. Also the blood

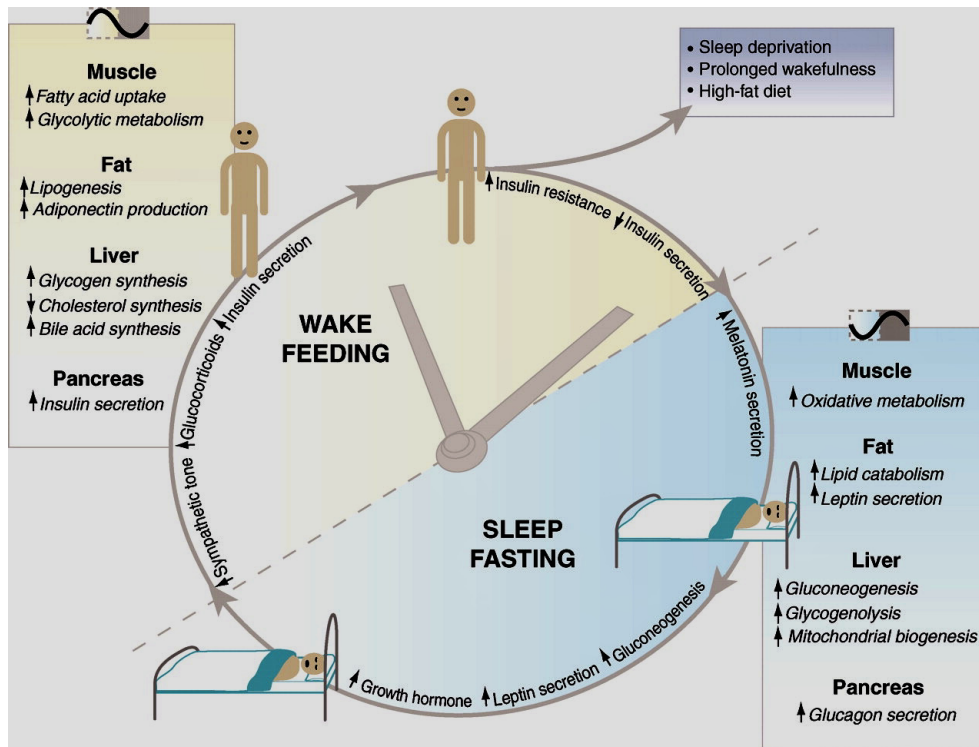


Figure 1.4: **Circadian physiology throughout the day.** The clock partitions behavioural and metabolic processes according to time of day. The clock coordinates appropriate metabolic responses within peripheral tissues with the light/dark cycle. For example, the liver clock promotes gluconeogenesis and glycogenolysis during the sleep/fasting period, whereas it promotes glycogen and cholesterol synthesis during the wake/feeding period. Proper functioning of peripheral clocks keeps metabolic processes in synchrony with the environment, which is critical for maintaining health of the organism. Different tissues exhibit distinct clock controlled properties; thus, ablation of the clock in certain tissues will cause opposing effects on metabolic function as uncovered through dynamic challenges at different times in the cycle under different nutrient conditions. Aging, diet, and environmental disruption such as shift work may also affect the integration of circadian and metabolic systems. Figure and caption from *Circadian Integration of Metabolism and Energetics* [14]

pressure response to external stress is controlled by our circadian pacemaker [64]. In addition, heart rate and heart rate variability follow daily oscillatory patterns [65–67]. All these factors could be linked with the uneven timing of a variety of cardiovascular events [68]. Clocks all around the body contribute to the oscillations in physiology and need to be well-timed to avoid negative interference [69]. A putative mechanism to ensure such coherence is signalling from the master clock via systemic cues such as hormones and glucocorticoids [70]. It might be of crucial importance for medicine to understand the interplay of clocks in different tissues [71]. It is believed that disruption of communication between the central clock and the vast network of peripheral ones is able to destabilise periodic oscillations across the body leading to a variety of metabolic diseases and neuropsychiatric disorders [72].

1.1.6 Timed health

During the last 20 years, two questions were repeatedly addressed: what are the most prominent effects on human health of a disrupted circadian clock, and what are the main causes in today's lifestyle that deteriorate our internal timing system? Only fairly recently the role of the circadian clock in human health has earned a more prevalent position in biomedical research, as ties have been discovered with a wider variety of diseases across the body [73–78]. Faults in the circadian clock have also been connected with the aberrant metabolism of cancerous cells at the end of last century [79]. It has been hypothesised that this interplay is due to reciprocal interactions with the cell-cycle, bestowing upon the circadian clock a tumour suppressor function [80]. Since the first link between cancer and a broken clock, more and more links between a disrupted circadian clock and cancerous growths have been identified [81–89]. These studies have been possible also thanks to the explosion in available data [90]. In addition, the synchrony both among all the internal clocks in different organs and with the external cycle is responsible for proper timing of downstream metabolic processes, thereby contributing to the health of the organism; the interplay of the circadian and the metabolic networks can be disrupted, especially in humans, by a variety of factors such as ageing, meal timing, jet-lag, and shift work [14, 91, 92]. The idea to restore circadian clocks, especially using time restricted feeding [93–95] to fight diseases where anomalous metabolism is pervasive is gaining momentum [96–99] and is summarised in figure 1.5. In addition, pharmacological modulations of the circadian clock have been introduced as a concept to fight cancer [77, 100]. To sum up, our current understanding suggests that the lack of a functioning circadian clock contributes











	Circadian rhythm disruption	Time-restricted feeding	Potential mechanism
	Obesity	↓Fat, ↑lean mass	↓Plasma- and ↓liver-triglycerides
	Glucose intolerance/ insulin resistance	Improved glucose homeostasis	↓Gluconeogenesis ↑PPP and ↑TCA cycle
	Gut dysbiosis	Diverse and dynamic	Altered digestion, absorption, and excretion of nutrients and bile acids
	Cardiovascular diseases	Arrhythmia and improved ↓cardiac function*	ATP-dependent chaperone and improved mitochondria function
	Chronic inflammation	↓Tissue inflammation	↓Macrophage infiltration of WAT ↓IL6 TNF α
	Liver diseases	↓Fibrosis and ↓hepatic fat deposit	Fatty acid synthesis, ↑ β oxidation mitochondria volume
	Increased cancer risk	↓Risk for breast cancer [#] and ↑breast cancer prognosis	Improved metabolic homeostasis, reduced inflammation
	Hypercholesterolemia	↓Cholesterol	Cholesterol metabolism to bile acids
	Sleep disorders	↑Sleep quality [#] and ↑quantity*	Consolidation of activity and rest
	Compromised muscle function	↑Endurance and ↑flight index*	Ketone bodies, creatine metabolism

Figure 1.5: **Chronic circadian rhythm disruption by erratic lifestyle whereas time-restricted feeding can restore daily rhythms and improve health.** The potential mechanisms are largely based on rodent studies. Few observations have been made in insects (*) and in humans (#). IL, interleukin; TNF, tumor necrosis factor. Figure and caption adapted from *Circadian physiology of metabolism* [92].

to irregular and spread-out food intake and metabolic disorders, and may lead to higher cancer rates and increased susceptibility to metabolic diseases. On the other hand, restoring circadian patterns seems to improve health and alleviate metabolic disorders. Many factors contribute to the weakening of the circadian rhythm; some are due to our lifestyle, such as sleep disruption, shift work, and absence of 12 hour fasting periods, and others are further from our control, such as ageing, chronic diseases characterised by altered metabolism, and cancer. It should be mentioned that, although promising in offering potential novel therapeutic avenues, complete biochemical proofs of these notions are still lacking in many cases. Lastly, it is important to remember that it is relevant, although often overlooked, to take into account the possible effects of sex and age on the circadian clock.

1.1.7 Circadian sex-dimorphism

The notion that female mice have stronger and more robust behavioural circadian oscillations is dominant in the field [101]; nonetheless still the vast majority of circadian animal experiments only involves male mice. With the wider availability of sequencing techniques a quantitative measurement of the differences in rhythmicity was brought forward [102]. Although the majority of transcripts shared rhythmic behaviour among the two sexes, 16% of transcripts was only rhythmic in females while only 9% exhibited rhythmicity only in male mice. Stronger circadian oscillations in females has also been recapitulated in humans behavioural rhythms [103, 104]. Indeed males had either higher night activity or lower day activity contributing to the erosion of circadian oscillations. Females also showed higher circadian rhythmicity on cognitive tests while undergoing forced-desynchrony experiments [105]. These experiments uncouple the natural oscillations from the sleep wake cycle and control for circadian phase, prior sleep, and elapsed time awake when interpreting cognitive test results. Lastly, the sleep hormone, melatonin, shows earlier and higher circadian oscillations in females, inducing earlier chronotypes and better compartmentalization of tasks [106]. Chronotype, the difference between the state of the internal clock and the external time, has been shown again to be dependent on both age and sex [107].

1.1.8 Age-dependent rhythmic physiology

The population, especially in first world countries, is ageing and thus increasing the risk and frequency of chronic diseases [108]. The influence of the clock on behavioural and sleep rhythms is well known and has also been correlated to the

ageing process [109, 110]. In fact, circadian rhythms in animal models shift in phase and decrease in amplitude as a result of ageing [110, 111]. Animal studies have also shown the decline of Zeitgebers reception as a result of ageing [112–115]. This in turn diminishes the ability to adapt and anticipate environmental changes in aged mice. High amplitude oscillatory rhythms are regarded as beneficial for animal well-being and longevity [116–118]. In addition, if the coupling between the internal and external cycles is not possible due to large differences in period, the expected lifespan of mice is reduced by about 20% [119]. On the other hand, interventions to restore circadian oscillations, like implanting fetal SCN into aged hamsters, increase longevity [30, 120, 121]. Lastly, circadian disruption due to chronic jet lag² has been associated with cancer and reduced lifespan in mice [122, 123]. In humans precise experiments on the interplay between the clock and ageing are complicated to replicate. However, the set of data and experiments, many of which are reported in section 1.1.6, suggests that a functioning clock increases health and longevity and that therapeutics aiming to restore circadian rhythms can revert some effects of ageing or disease.

Conclusions

We have encompassed the relevant aspect of the circadian clock, from its molecular and hierarchical structure to its importance in health and disease, and how it has been studied up to now. Now we will move on to describe how to approach highly multidimensional omics data (becoming omnipresent in biology), such as those introduced in section 1.1.2. The core idea to study omics data is to project them onto a suitable, known or inferred, low dimensional manifold that captures the properties of the biological system.

²Imagine living a week in Paris then one in Tokyo for a year.

1.2 The space of biological data

As humans progressed through history the knowledge we gained about the world has become increasingly quantitative [124, 125]. Biology has been no exception, especially in the last 100 years. The sheer amount and quality of data we have been able to collect has grown astonishingly [126–128]. Many different quantities have been measured more and more reliably as time and technology advanced, spanning almost all the questions and scales of biological systems. Our focus is on systems and molecular biology [129–131], the origin, almost at chemical level, of the living world around us. To correctly characterise the state of systems or cells we would, ideally, measure protein activity levels. This is not yet achievable at the scale and cost that would allow protein activity measurements to be ubiquitous and universal in biological studies, although we have been working tirelessly in this direction for half a century [132–135]. After the advent of affordable technologies to measure relative mRNA abundance [128, 130], the microarray technique [136–142] first and RNA-seq [143–147] with its single cell counterpart [148–153] later, mRNA abundance has often been used as a proxy for protein concentration. In this thesis we developed our methods and conclusion on the backbone of genome-wide mRNA abundance measurements and their relationship with protein activity. Each point in the mRNA space of measurement is a highly multi-dimensional vector, with about 10^4 entries. All the currently available data lives in the high dimensional "gene space", and if left there is of difficult interpretation. Therefore we need techniques or assumptions on how to reduce the dimensionality of the data without distorting the meaningful properties of the system of interest.

1.2.1 Low dimensional projections

A general assumption is that there exists a low dimensional manifold [154] \mathcal{M} that encapsulates the relevant properties of the high dimensional data. All the variance induced in the data outside of \mathcal{M} is either measurement or biological noise, or is due to other processes. In order to be able to project the high dimensional data onto \mathcal{M} we can follow two distinct approaches. We can try to infer the best N dimensional manifold \mathcal{J}_N and hope that there is a mapping³ from \mathcal{J}_N to \mathcal{M} . On the other hand we can assume to know beforehand the topological structure of the low dimensional manifold. In this case the difficulty lies in finding the correct projection of the data in order to capture the biological process we want to study

³By mapping we mean a continuous mapping with a continuous inverse, the very definition of homeomorphism. We did not use the technical term in the text to facilitate reading.

on the manifold.

Inferring low dimensional manifolds

The task of inferring a low dimensional manifold to reasonably represent high dimensional data living in the gene space is not a trivial one. Efforts of almost a century have brought today a variety of methods, which can be divided into two categories depending on the conceptual approach to dimensionality reduction. Some, especially older approaches, want to keep the overall distance structure of the data [28, 155–158], the most famous being PCA [28]. Modern methods favour a preservation of local geometry disregarding long distance relationships [159–164]. Among these there are the most commonly used in single cell biology, also due to their ability of recapitulate important similarities and differences between cells, T-SNE [163] and UMAP [164]. Although these [163, 164] methods often capture crucial properties of datasets in their inferred low dimensional space, they fail in some cases. This is especially important if we are studying a process which is not the main driver of variance in our dataset. In such a case it would be useful to pre-process the data to allow the process of interest to be the main source of variance in the data. If we cannot remove the undesired variance it would be precious to know, or at least hypothesise, how the low dimensional is shaped to ease the task of projecting the data.

Known low dimensional manifolds

There are cases in which we know the structure of the low dimensional manifold. For example cells switching from state A to state B can be represented by a segment going from A to B. In the rare events where the topology of the low dimensional manifold is known or can be assumed, the projection can be obtained in a more controlled way. In addition, a more mathematical approach for the analysis of the data can be put forward if we know the low dimensional manifold. However, it is still not an easy task to correctly project high dimensional data onto a known low dimensional manifold capturing the relevant variation induced by the process. We are particularly keen on exploiting circular manifolds, homeomorphic to the unit circle S^1 . Such manifolds also have a clear and convenient basis (Fourier) to express functions. Crucial biological processes of great importance to us, such as the circadian clock (described in section 1.1) and the cell cycle (described in section 1.3.2), can be represented on the unit circle, S^1 . These two processes can be studied simultaneously in a low dimensional manifold that is the cross product

of two circles, a torus, and have been shown to be linked in single cells [165]. The projection onto low dimensional manifolds also allows us to study new descriptions of the system, like the statistical physics one.

Statistical physics on low dimensional manifolds

We assume we managed to have high dimensional data projected onto a low dimensional, and we want to study the statistical physics of the biological system. By statistical physics we mean the description of a system, in terms of statistical mechanics of equilibrium, that has characteristics determined by the biological data. To obtain such descriptions we can, for example, map the parameters of a known system in statistical physics to a function of biological data (as we will do in section 3.5). The mapping between low dimensional representation of biological data and known problems in statistical physics can provide insights on the behaviour of emerging properties of the biological system, that might not be obtained otherwise. Statistical physics is especially useful at determining and calculating changes in the collective behaviour due only to local interactions, such as transitions from disorder to order, and has already contributed to the advancement of biology [166, 167]. However, there is more to biology than a static description. Biology describes life in its essence, but life is ever-changing and deserves a dynamical description.

1.3 Dynamics in biology

Introducing the dynamical descriptions of biological data we address a crucial question: how do systems move in the low dimensional space, thus how do they evolve in the high dimensional (gene) space? The main difficulty is to correctly write the equations of motion. We do not necessarily know the correct parametrization to write a humanly interpretable set of equations that encompasses the key features of the biological system we are studying. Physics can come to our aid suggesting reasonable quantities to measure experimentally and match with the theoretical predictions. However, no one guarantees that the quantities relevant in physics are also the correct one to describe biology, and most probably they aren't⁴. Physics has guided the idea that cells as they differentiate descend a potential landscape, which has been promoted more than fifty years ago [159, 168–170]. It was proposed that cells should navigate a static landscape, this can't account for biological systems such as switches. Recently, this limitation has been addressed by a low dimensional

⁴This idea has been developed also through conversations with Prof. S. Rahi

dynamical system encapsulating the behaviour of biological switches and transitions between cell types [171]. Such mathematically detailed systems are of great interest but do not cover many scenarios in biology. In fact, the variables governing a vast portion of biological systems are still not clear. With the arrival of affordable single cell technologies the question about the dynamics of single cells arose in a vast number of distinct branches of biology.

1.3.1 The velocity of cells

With the advent of scRNA-seq [149] ten years ago a new resolution of measurements was introduced in biology. The ability to measure genome-wide transcriptome in single cells has allowed a more precise study of many biological processes, such as the introduction of atlases for a variety of tissues and species [172–176]. However, the temporal aspect is still complicated to investigate as systems do not survive an RNA-seq measurement. So we are left with only a snapshot of cells at a defined time point, which limits the possibility to infer and characterise their dynamics [177]. By dynamics of cells we mean the dynamics in the high dimensional gene space or in a suitable low dimensional manifold where high dimensional data has been projected (section 1.2.1). We want to understand how and why cells navigate the transcriptional landscape. Thus, the velocity of cells is the velocity with which cells move through the gene space or, depending on the situation, a low dimensional manifold where biological properties of interest are conserved. The idea to use unspliced mRNA as a proxy for mRNA abundance in the future (although not a precise future), an approach already described for bulk measurements [178, 179], was applied to single cells five years ago [180] with brilliant results and is recapitulated in figure 1.6. This technique has been widely used in the so called "single-cell temporal-omics approaches" [20, 176, 182–186] (summarised in figure 1.7), especially in developmental studies. Temporal-omics methods study development and pathology by integrating temporal information existing in an omic snapshot with the transcriptome data. RNA-velocity has also been upgraded [188] and expanded to tentatively include proteins in the equation [189, 190]. An important remark on all the formulations of RNA-velocity is that the arrows of the "velocity" actually represent transitions probabilities. Longer arrows signify more probable transitions, not faster ones. Also, the manifold chosen to represent the interesting variations in the data is always an inferred one, with little to no external constraints, as described in section 1.2.1. This choice is due to the complex architecture and yet unknown topology of developmental manifolds. The general theory and applications of the RNA-velocity technique are still open to improvement, especially with regards to dy-

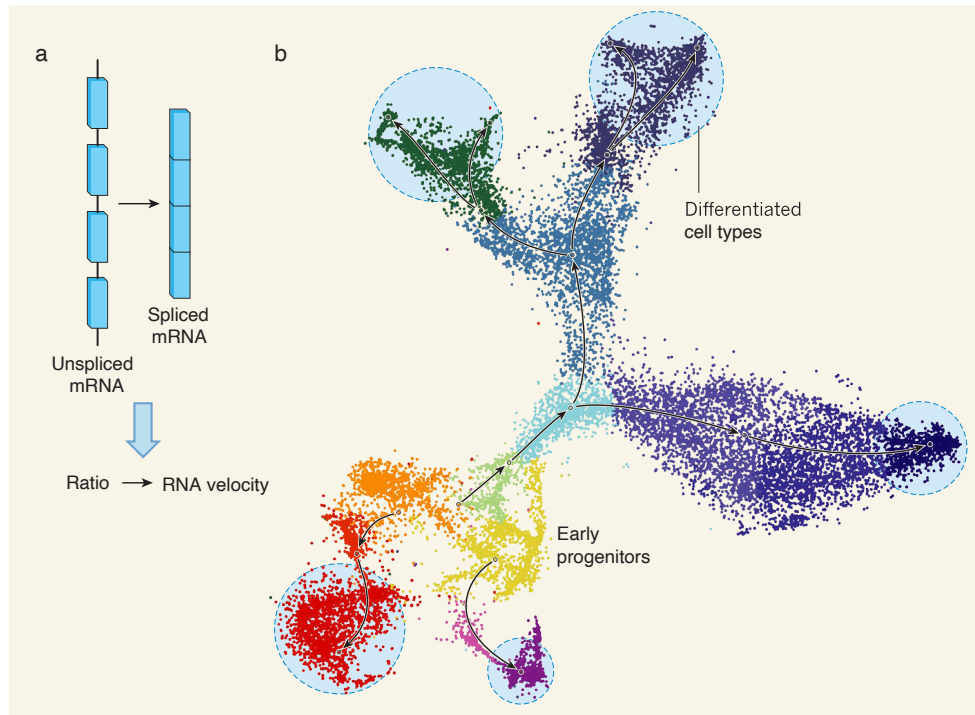


Figure 1.6: Measuring dynamic changes in gene expression across complex tissues. **a**, As messenger RNA matures, sections of the immature transcript are removed — a process called splicing. When the expression of a gene increases, a transient increase in the proportion of immature, unspliced transcripts compared with that of mature, spliced transcripts is observed in the cell. By contrast, a higher proportion of spliced transcripts is seen for a short time when expression of the gene decreases (not shown). La Manno et al [180] measured the ratio of unspliced to spliced transcripts for each gene in a single cell to calculate a quantity called the RNA velocity, which reveals how gene expression is changing. **b**, By measuring RNA velocity in thousands of cells in a tissue (here, in neurons in the developing mouse brain), the authors could generate maps that show not only how closely related cells are to one another (with closeness indicated by similar colours), but also which cells they will become similar to in the future (indicated by arrows), according to the gene-expression changes they are undergoing. RNA velocity successfully tracks early progenitors (orange and yellow) that eventually give rise to a range of differentiated cell types (blue dashed circles). (**b** adapted from Fig. 3c of [180]), Figure and caption from *Technique to measure the expression dynamics of each gene in a single cell* [181]

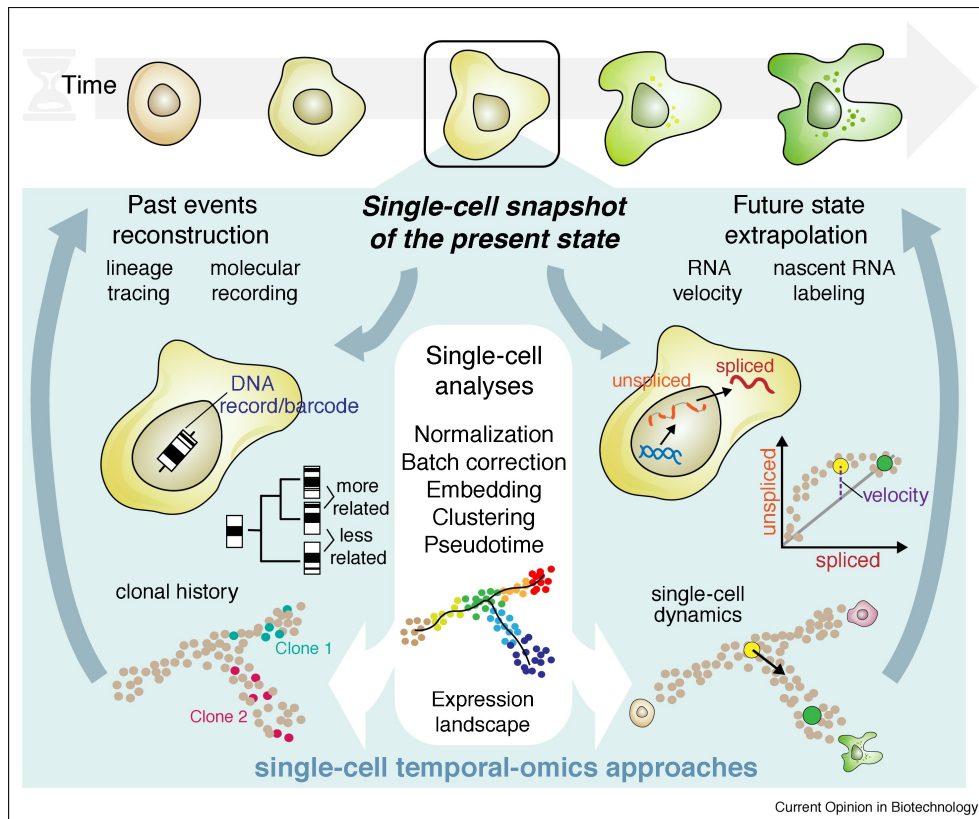


Figure 1.7: Illustrative representation of single-cell temporal-omics approaches. Given the snapshot obtained from single-cell RNA sequencing, data analyses enable characterization and classification of the gene expression landscape in a heterogeneous population of cells. Recent methods further enable the extrapolation of future gene expression states (right) and reconstruction of past cellular events (left). Together, these approaches permit greater inference of the temporal changes within a single cell while still relying on measurement from a single time point. Figure and caption from *The emergence and promise of single-cell temporal-omics approaches* [187]

namical constraints (i.e. give the arrows in the velocity field the dynamical meaning of actual velocity). The cell cycle is a perfect testing ground for RNA-velocity and its derivatives as it is a system with an evident dynamic and a known low dimensional manifold.

1.3.2 Cell cycle speed

Arguably the most ubiquitous process in nature, the cell cycle and has been studied for almost two hundred years since its discovery in the middle of the nineteenth century [191, 192]. The cell cycle is divided into four phases or better two active phases and two gap phases [130, 193–196] summarised in figure 1.8. The active phases are mitosis (M phase) and DNA synthesis (S phase); the cell after mitosis undergoes G1 (gap 1) phase where it can exit or re-enter the cell cycle, then DNA synthesis occurs followed by the second gap phase, G2 (gap 2), and finally mitosis again. The reason why the cell cycle is viewed as a series of steps rather than a continuous oscillation lies in the crucially different mechanism that separates the cell cycle from the circadian clock. In fact, the cell cycle progresses through its phases due to a timed degradation of relevant proteins that are otherwise expressed in the cell [128, 130] in opposition to a negative feedback loop. Given the periodic structure of the cell cycle, even if it is biologically divided into discrete phases, the temptation of mapping it to the unit circle S^1 is too great to resist. Indeed this mapping could bring new insight on the cell cycle, especially if the discrete phases are not a sufficient characterization. The FUCCI system was a turning point to directly evaluate the cell cycle state (and phase) of a cell by comparing the intensity of two fluorescent proteins in microscopy experiments [197]. These measurements cannot be easily coupled with scRNA-seq to match transcriptome information with cell cycle phase. Although if combined, information about cell cycle phase and transcriptome bring insights on how individual cells are able to shape their gene expression dynamics [198]. Many computational approaches, generally lacking generality and robustness have been attempted [199–203] but no golden standard has been recognized by the community. Although the cell cycle has a clear periodic structure [204] its period is highly variable across cells; even among renewing cells 10 – 100 hours can pass between two subsequent divisions [205]. This variability is highly attributable to variation in the length of G1, as S and G2 last between 2 to 4 hours each and M phase lasts only 1 – 2 hours.

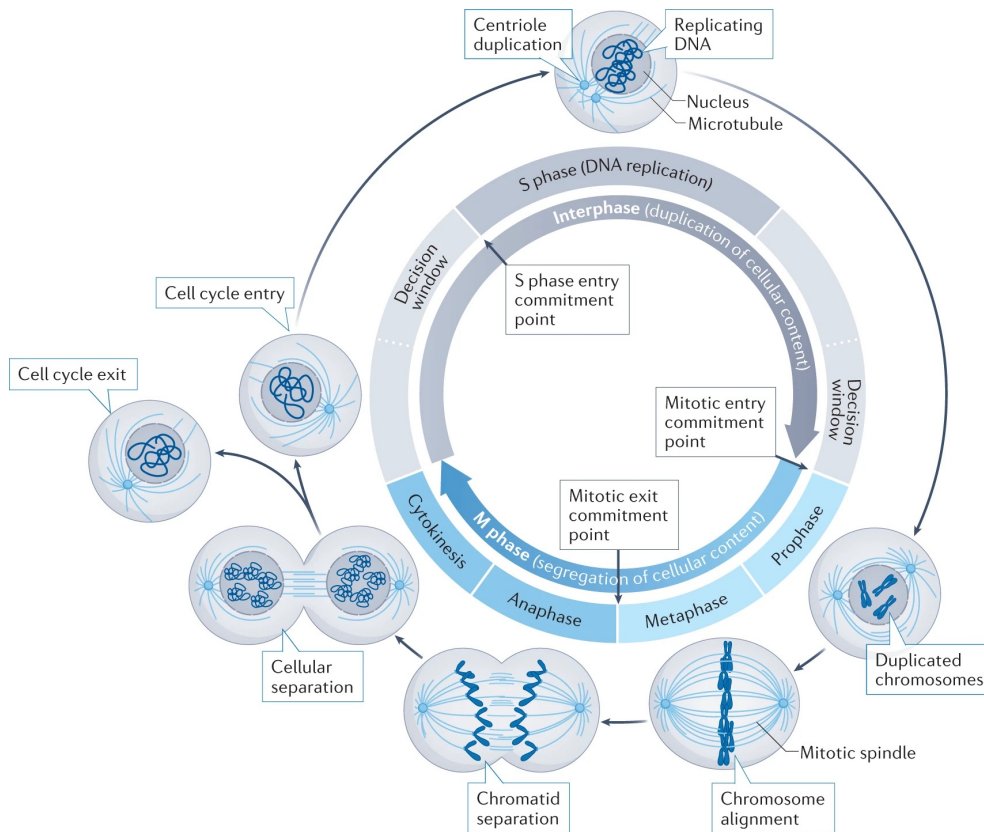


Figure 1.8: The cell division cycle. The eukaryotic cell cycle is the process during which a cell duplicates its entire cellular content during interphase, and through division in M phase creates two genetically identical cells. The two main events, DNA replication and the segregation of the replicated DNA, are separated during the cell cycle. DNA replication happens during a distinct phase in interphase, called ‘S phase’, and DNA separation happens in mitosis (M phase). Segregation of the cellular content happens during cytokinesis at the end of M phase to complete a cell cycle, after which a cell can either exit the cell cycle or enter a new round of cell division. During interphase, cell cycle progression is controlled before and after S phase. Commitment to enter a new cell cycle (that is, S phase entry) is made during a decision window preceding S phase. Similarly, during a decision window following S phase, a cell can commit to mitotic entry. Commitment to mitotic exit happens during M phase at the metaphase–anaphase transition. *Cell cycle control in cancer, Matthews et al. [196]*

Conclusions

We have now explored the setting of the various problems we want to address and investigate. They range from methodological questions to biomedical ones, spanning scales from human population to single cells. There is one thread to bring them all and in this thesis bind them: their nature as high dimensional problems that we study on opportune low dimensional manifolds. Now we formulate more clearly the questions driving the work presented here.

1.4 Direction and aims

Here we discuss the main aims that drove the research described in this thesis; not all of the questions presented here will find an answer here.

1.4.1 Circadian phase inference

We want to reconstruct the circadian phase of biological omics samples. By circadian phase we mean the molecular state of the circadian clock with respect to its cycle. This is equivalent to finding the best projection of high dimensional data living in the gene space on the unit circle that preserves the circadian clock and its related processes. One of the main problems in the field is to find a method to correctly infer the circadian phase given mRNA abundance measurements. We want to achieve this in a conceptually different way from what has already been done [15, 22–27]. Published methods are mainly supervised⁵; our aim is to create a general, unsupervised⁵ method of phase reconstruction. We aim to find a robust and general way to reorder any given dataset in which the oscillations due to the circadian clock exist. The key methodology is to develop a probability model, which allows us to derive joint probability distribution on the quantities of interest (phases). In a probabilistic framework this means to find the maximum, or ideally the expected value, of $\mathcal{P}(\text{circadian phases}|\text{mRNA abundance})$. To be biologically useful our method must be robust and work well on real data. The bridge between these two requirements is a computational framework, where we will need to implement our calculations once analytic computation will become unfeasible. To test the method we find, we will apply it to datasets in which the circadian phases are known. Once a working method is found we will apply it to human data.

1.4.2 Circadian clock in humans

The main objective is to apply our method of circadian phase inference to human data and learn new chronobiology. We want to understand the behaviour of the circadian clock and gene expression rhythmicity in humans. Although many experiments have been performed in mice, the human clock is still vastly understudied. We are also keen on exploring the circadian difference between males and females or during ageing. We are finally interested in the medical application of our findings,

⁵These words are borrowed from machine learning: with unlabelled we mean datasets where no temporal information about the samples is known, with supervised we mean that we have a set of labelled data to use, and with an unsupervised approach we do not have any

in particular regarding age or sex related differences. These differences could give crucial medical insights regarding treatment timing differences for the different groups, explaining the different incidence of certain illnesses, or even suggest novel treatments approaches. Using all our previously constructed methods, we aim to investigate the state of the clock in diseased tissues. We want to study the behaviour of the circadian clock in unhealthy samples, to more precisely link the clock with various illnesses. In order to do this in a significant way, we need to introduce a metric to evaluate the integrity of the clock. We would need to develop two different metrics of clock integrity: one for a single sample, and one for a whole dataset.

1.4.3 Dynamically consistent RNA-velocity

As RNA velocity [180] is a very new technique, not all of the possible adjustments have been made, although many have been tried. the inner consistency of the velocity field should be exploited, both arising from differential geometry and derived from a dynamical systems perspective. We want to be able to integrate the velocity field through a trajectory and determine its duration. We start by using a convenient system: the cell cycle.

Cell cycle properties

The cell cycle is an almost ubiquitous process in biology, so a more precise knowledge of how exactly cells progress through it could be a great tool to better understand the steps of life. It would be relevant to infer a velocity field that can be integrated along one cell cycle to calculate its period. This approach should be validated using inhibitors[206, 207] or temperature [208, 209] to slow down the cell cycle and compare the period of cell cycle measured with the FUCCI system [197] to the one inferred computationally. Once this can be done, a much more comparative and in depth study of the cell cycle can begin. We want to infer both the period of the cell cycle and of the single phases in developmental atlases to better understand the interplay of the cell cycle and differentiation. It would also be interesting to compare the shape of the speed of cells undergoing the cell cycle across cell types and species.

Circadian phase inference Part II

2 How to tell time: advances in decoding circadian phase from omics snapshots

This section has been adapted from the review article *How to tell time: advances in decoding circadian phase from omics snapshots* [210] written by Lorenzo Talamanca and Felix Naef. The work was envisioned, supervised and checked by Felix Naef, it was compiled, written, and illustrated by Lorenzo Talamanca. Sections of this review have been incorporated in the introduction to improve the overall flow of this thesis. What follows from section 2.3 is the core of the review modified as little as possible from the original text. The abstract and introduction have been adjusted to improve readability.

2.1 Abstract

The ability of organisms to keep track of external time is essential for health. This ability is achieved through the circadian clock interacting with the environment. The focus of this review is recent methods to detect the internal circadian time of an omics sample. We analyze an important methodological question: how to infer the circadian phase of unlabeled omics snapshot measurements. Answering this question could both significantly increase our understanding of the circadian clock and allow the use of this knowledge in biomedical applications. We review existing methods following a historical trajectory, concentrating on the more recent ones. We explain the basic concepts underlying the methods, as well as some crucial technical aspects of each. We conclude by reporting how some of these methods have, more or less effectively, enabled furthering our understanding of the clock and given insights regarding potential biomedical applications.

2.2 Introduction

The circadian clock is the celebrated evolutionary response to earth's intrinsic 24-hour periodicity [4]. It is a cell-autonomous molecular oscillator driven by transcriptional-translational negative feedback loop with a period of about 24 hours [7], due to only a few dozen genes [6]. It drives behaviour and physiology [16] adapting them to the periodicity of the environment. Disruption of the circadian clock have been associated with a variety of diseases [91]; time restricted feeding has been used in recent years to rescue circadian oscillations and counteract metabolic diseases [92]. However, due to the impossibility of performing experiments around the clock on humans, we have often relied on mice data to infer properties of the circadian clocks in mammals. On the other hand there is a vast availability of human data, which often come without a time stamp. Thus a series of computational methods, that we will here review, have been brought forward with the intent of inferring the circadian phase, i.e. the molecular state of the circadian clock with respect to its cycle, of omics samples, therefore allowing to study the clock in humans.

2.3 Computational methods to infer circadian phase

Since the importance of circadian timing in human health was established, an interesting question arose: how can we detect the internal time of a tissue sample from its gene expression level? Answering this question might open up significant novel medically relevant opportunities, especially for diagnosis, prognosis, and potentially therapeutic strategies for a variety of illnesses, most importantly cancer [211]. In addition, reliable methods to infer circadian phase would allow us to further study the structure and effects of the clock by leveraging the vast quantity of existing unlabeled RNA sequencing data on humans available in public databases. We could gain important insights on clock biology in healthy and diseased human tissues, notably useful for future biomedical applications. Although many outstanding works have been recently published, how reliably and robustly this can be achieved remains an open question. An array of methods is currently available. The majority of algorithms available are supervised in the sense that the parameters are trained on time-labelled data (the training set). A word of caution concerns the intrinsic difference between the circadian phase and the external time, as explained above, which is generally not considered: circadian phase and the external time are often taken as equal in labelled datasets. One natural method to assign a circadian phase to unlabeled samples consists of applying the four-quadrant

arc-tangent function to the two first principal components [212]; we will refer to this unsupervised method as PCA. In 2004, a first significant step was made [213] by developing a method, molecular timetable, to infer the circadian phase based on the temporal and expression patterns of known clock-related marker genes in mice livers. The core idea is shared with several other supervised methods that will follow it. Namely, it consists of finding a suitable low-dimensional representation (also called the low-dimensional manifold) and then, in this coordinate system, finding the one-dimensional trajectory (termed here circadian trajectory) that best matches the known labels of the samples in the training set. A graphical representation of this core idea is shown in Figure 1. To find the phase of a new sample, it must be suitably projected onto the circadian trajectory in the low-dimensional space. In [213], the low-dimensional manifold consists of a set of standardised “time-telling” marker genes for which a sinusoidal shape is assumed and the peak phase fitted from the training set. The inferred circadian phase is obtained by minimising the least square error across all the time-telling genes. Although simple, this method deserves praise for the pioneering work, also considering the scarcity of sequencing data available compared to modern day. For the next important contribution, we have to wait more than 10 years. The availability of data and the increased interest in the circadian clock attracted many researchers to this problem, which yielded four new methods, each claiming to be better than the last, in less than 4 years. The first was ZeitZeiger [23] in 2016. This method adds a supervised twist to PCA. The low-dimensional space is a set of sparse principal components (SPCs), each component being a sparse linear combination of genes. The linear combination of each component and the non-linear mapping from the SPCs to the circadian phase is optimized on the labeled training set, while the number of SPCs and the maximum nonzero entries in each of them are optimized by cross validation. To infer the phase of new samples, the algorithm linearly projects the gene expression levels to the SPC space and then finds the phase at which the circadian trajectory is closest to the projected gene expression levels. This method exhibits good performance in inferring the circadian phase; in addition, because of its easy and clear structure, it is able to identify which genes carry the most information about the circadian oscillation in a labelled sample. A few months after, a machine learning method was developed: BIO_CLOCK [15]. This method is made of a three-layer dense neural network taking as input the normalised expression of 16 clock genes and yielding as an output the sine and cosine of the circadian phase. This method was trained and tested on mice microarray data from different organs with a 70 – 30 training-test split. Being a dense neural network, the training can be performed via backpropagation. To infer the phase of a new standardised sample, one simply

inputs the required gene expression levels and takes the four-quadrant arc-tangent of the network's output. Although this method uses a black box neural network, its performance is on a par with ZeitZeiger. Finally, the need for an unsupervised method (no labelled dataset is available) was recognized, and in 2017 CYCLOPS was invented. This method uses a linear autoencoder with a circular node, constructed with two coupled neurons, so the low-dimensional space is the two-dimensional space of the code neurons and the circadian trajectory is the unit circle in the two-dimensional space from which the circadian phase is extracted via the arc-tangent. CYCLOPS does not need a training set to learn its governing parameters; however, it cannot predict the phase of only one new sample. The general scheme of this algorithm is as follows: project to a low-dimensional space, from this project to a circadian trajectory, and from the circadian trajectory try to project back onto the original space, making the smallest possible error across all samples. For CYCLOPS, the low-dimensional representation is reached by linear projection onto two dimensions and the circadian trajectory is simply the unit circle where all points of the two-dimensional space are orthogonally projected; from the two-dimensional unit circle, points are linearly projected back to try to match the measured gene expression levels. The difference between CYCLOPS and PCA is the non-linearity from the two-dimensional space to the one-dimensional circadian trajectory included in the autoencoder. CYCLOPS offers one crucial advantage over supervised methods: it is more easily generalizable. This means that it can be directly applied to a much wider range of datasets, even ones with biologically different periodic behaviours, like the cell cycle, without the need for enough labelled data to retrain the algorithm. However, one must note that to obtain the best results the input of CYCLOPS should be restricted to a subset of genes, which have been previously implicated with time. Still, in 2017, a new supervised method, PLSR [214], was published. The idea of this method is to linearly project both the sequencing data and the two-dimensional representation of the phase (its sine and cosine) onto a five-dimensional feature space and maximise the correlation between the two. In this case, the projection to the feature space is optimised on the training set and the dimensionality of the feature space is selected via cross validation. This method was built, trained, and tested on in-house blood sample measurements, which, unlike all the other training sets, used the dim light melatonin onset (DLMO), the gold standard assay to estimate circadian phase in humans, as the measure for circadian phase, not the external time. Using the gold standard for circadian phase detection in the training set gives this method a good advantage; however, it was never applied to datasets not collected in the authors' lab. Lastly, another supervised method to infer circadian phase from blood RNA sequencing data was developed: TimeSignature

[26]. This approach is reminiscent of [213], as the low-dimensional space consists of a subset of genes, then, using the training set sinusoidal patterns are fitted for each gene with a ridge and lasso penalization to control the amplitude of each gene and reduce the number of genes kept. To infer a new phase, one must plug the gene expression level and determine the best fitted phase from the shapes of the reference genes. This method is trained and tested on human blood, with external time as the label. In addition, there is one disadvantage: each sample needs to be paired with another one, from the same person, taken 10–14 hours apart; without this pairing, the method cannot work, as the first step is a within-subject renormalization. This renders the method inapplicable to most existing data and deeply constrains its use in a medical context. The latest, still unpublished, method, TimeTeller, is a supervised method that exploits the periodic expression and covariance of 10–15 key genes to infer external time from a single tissue sample [27]. A summary of the main characteristics of each of these methods is presented in Figure 2. Although it is generally required for methods of the same category (supervised, unsupervised) to show that they perform better than their predecessor, this fact needs to be taken with a pinch of salt. A common issue with comparisons between methods is that all methods are built on different datasets and optimize a slightly different quantity which further bias any comparison.

2.4 Applications

Finally, we dedicate a few words to the impact that some of these methods had and how they have been applied. As we have seen, ZeitZeiger is a supervised method originally trained on various mouse tissues; in 2017, it was trained on existing whole human blood microarray data and a small set of marker genes was found to reliably reconstruct the circadian phase of each sample [24]. A year later, in another study using blood samples, ZeitZeiger was trained on monocytes purified from the blood to accurately obtain the circadian phases from single blood samples [215]. The resulting improved accuracy can be explained by the fact that peripheral blood mononuclear cells (PBMCs), which are used by most studies, consist of a complex mixture of many cells including T and B cells, which together make up more than 80% of the cells; however, the clock in T or B cells is weak. On the other hand, monocytes host a high amplitude clock, which likely explains the superior accuracy at predicting the DLMO. Other researchers focused on human skin: in a first step, they applied CYCLOPS to reorder a large set of unlabeled samples, then they applied ZeitZeiger to find a small subset of biomarkers for

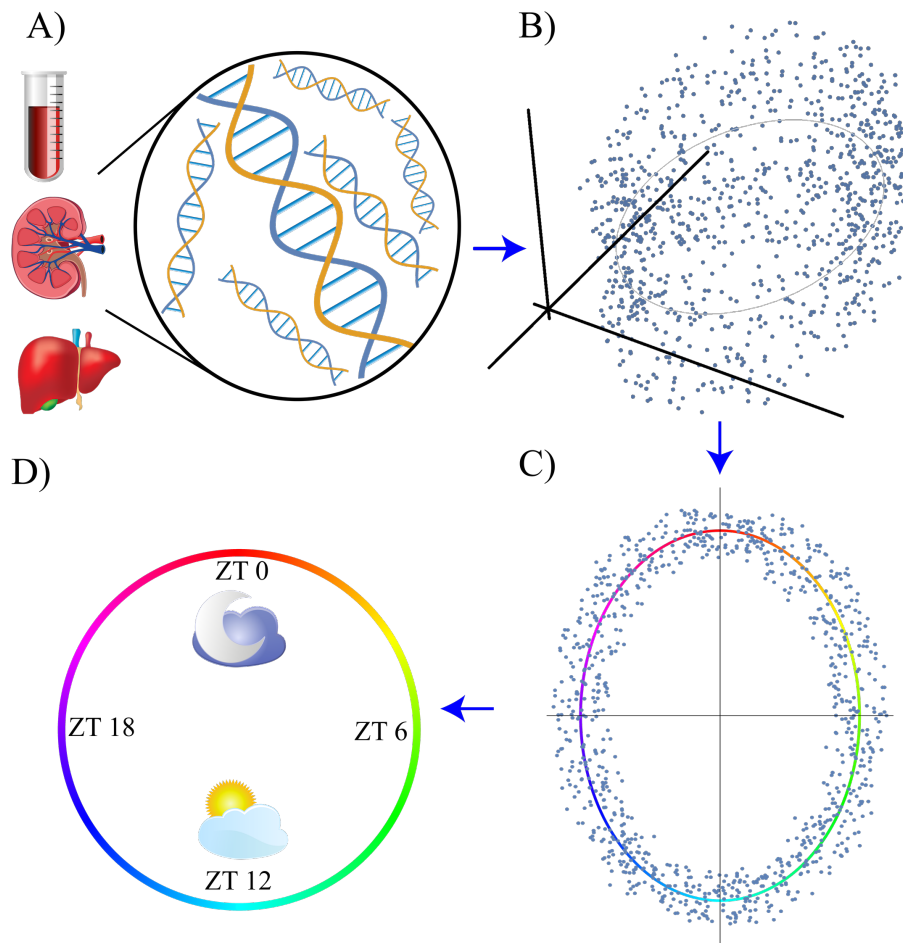


Figure 2.1: **A summary of the steps followed by many phase reconstruction algorithms.** The first step is the bulk mRNA extraction and sequencing from omics samples (A); thus, high-dimensional data are generated (B). The data are then projected in a low-dimensional representation and the circadian trajectory is identified (C). Lastly, the data are projected onto the circadian trajectory and the internal time (circadian time, CT) of each sample is identified (D).

Algorithm Kind	Year Author	Low-dimensional space	Circadian trajectory (CTR)	Training strategy	Phase prediction
Molecular timetable Supervised	2004 Ueda <i>et al.</i>	Set of 168 marker genes	Circle embedded in 168 dimensions	Fit the marker genes parameters (phase, amplitude)	Project gene expression vector on CTR
ZeitZeiger Supervised	2016 Hughey <i>et al.</i>	2 Sparse Principal Components (SPCs)	Ellipse embedded in SPC space	Optimize SPCs and CTR	Projection of data to SPC space, then to CTR
BIO_CLOCK Supervised	2017 Agostinelli <i>et al.</i>	2-dimensional nonlinear transformation of data	Unit circle in the 2-dimensional space	Backpropagation	Neural network prediction
CYCLOPS Unsupervised	2017 Anafi <i>et al.</i>	2-dimensional linear combination of data	Unit circle in the 2-dimensional space	Not applicable	Optimize gene expression reconstruction with 1-dimensional CTR bottleneck
PLSR Supervised	2017 Laing <i>et al.</i>	5-dimensional linear combination of data	5-dimensional interpolation of linear combinations (sin, cos) using the time labels.	Maximise correlation in 5-dimensional space between projections of data and CTR	Projection to 5-dimensional space, then to CTR
TimeSignature Supervised	2018 Braun <i>et al.</i>	Set of ~40 marker genes	Unit circle embedded in 40 dimensions	Fit the marker genes parameters (phase, amplitude)	Project patient-wise mean-centred gene expression vector on CTR

Figure 2.2: **A summary of the most important characteristics of the existing methods to decode circadian phase from omics samples.** d, dimensional

circadian phase in human skin [216]. Application of these algorithms to data from blood or skin samples is particularly relevant, as they are easily accessible and can potentially be easily used in biomedical research. Of note, the different studies listed differ by the type of readouts used (candidate genes vs. whole genome); in the context of clinical applications, it appears that candidate gene approaches such as NanoString assays [43] might have a clear advantage in terms of costs and complexity. An interesting application of these methods was published in 2018. CYCLOPS was applied to temporally order human RNA sequencing data from 13 different post-mortem tissues and used to perform a comparative rhythmic analysis across them [217]. In particular, the authors focused on drug target genes to propose improvements in the effectiveness of therapies by delivering drugs at the optimal time of day. In conclusion, the discussed advances in biomedical circadian biostatistics application promise to sprout important advances in around-the-clock treatments, especially those concerning the maximisation of drug efficacy and the minimization of their side effects [218].

3 The CHIRAL algorithm

We follow an historical trajectory and firstly focus on our algorithm for circadian phase inference, followed by its application on human data in the next chapter. This chapter has been adapted as little as possible from the preprint appendix of [219] written by Lorenzo Talamanca, Cedric Gobet and Felix Naef. The appendix was envisioned, supervised and checked by Felix Naef, it was compiled, written, calculated and implemented computationally by Lorenzo Talamanca. Although we tried to modify this as little as we could, given the role of appendix in the submitted version this chapter is the one with the most prominent changes.

3.1 Setting

Here we discuss the theory, mathematical steps, and approximations of our Circular Hierarchical Reconstruction ALgorithm, CHIRAL. This algorithm takes as input an ensemble of sets of measurements of quantities in a periodic system, like mRNA expression levels of core clock transcripts from tissue samples across the 24h. Each set of measurements represents a contemporary measurement of all the quantities in our system; In this example the set of measurements is the set of mRNA expression levels (RNA-Seq) for all clock genes for one sample. The aim of the algorithm is to assign to each set of samples in the ensemble a phase along the period of the system; i.e. a phase between 0 and 24hs for the circadian clock. We will keep in mind the circadian clock example throughout the calculations as that is our main practical objective.

3.2 Hypothesis

We start by clearly stating the hypothesis behind our method. The main hypothesis, underlying all methods of (circadian) phase reconstruction, is a simil-ergodic assumption. The ergodic hypothesis in general states that the average of a quantity across an ensemble is the same as the average of the same quantity over time of one single entity. In physics, generally, the ensemble is a set of particles and the entity is a single particle. In our case the ensemble is the set of donors (or sacrificed mice) from which we collected samples and the entity is one of them. This means that we assume that the behaviour of the circadian clock in one donor across time is the same as its behaviour across donors sampled at different points (conditions) spanning the (quotient) space of interest.

3.3 Multivariate harmonic regression model

As a starting point, we consider a genes by conditions (samples) $N_g \times N_c$ data matrix E_c^g . As we are considering a periodic system, we use the Fourier basis and consider up to N_f nonzero Fourier coefficients.

We model the data matrix as follows:

$$E_c^g = \alpha_f^g \zeta^f(\varphi_c) + \epsilon_c^g = \alpha_f^g \zeta_c^f + \epsilon_c^g = \sum_{f=0}^{N_f} \alpha_f^g \zeta_c^f + \epsilon_c^g \quad (3.1)$$

using the differential geometry notation where high indices only sum with low ones and vice versa. In this equation the α s represent the Fourier coefficients of the genes associated with the Fourier functions $\zeta^f(\varphi_c)$. The latent phases are denoted φ_c . In this notation, the first 5 functions take the form $(\zeta^0(\varphi), \zeta^1(\varphi), \zeta^2(\varphi), \zeta^3(\varphi), \zeta^4(\varphi)) = (1, \cos(\varphi), \sin(\varphi), \cos(2\varphi), \sin(2\varphi))$. The aim of the model is to obtain the phases $\{\varphi_c\}$ from the data $\{E_{gc}\}$. For the moment the error has two indices, as its distribution can depend both on the gene and on the condition. In addition we want a model that is able to weigh rhythmic and non-rhythmic genes, only the latter ones carrying information about the latent phases. In the notation of Eq.(3.1), the difference lies in the prior over α , as for non-rhythmic genes only the flat (0th harmonic) Fourier harmonic is considered non-zero.

Before proceeding, let us specify the relevant probability distributions. We will use Gaussian error models and Gaussian priors on the Fourier coefficients of the genes.

We start with the distribution of random errors

$$\mathcal{P}(\varepsilon) \sim e^{-\frac{1}{2}\varepsilon_c^g S_{gg'}^{cc'} \varepsilon_{c'}^{g'}}, \quad (3.2)$$

where the covariance matrix is a 4-dimensional tensor. Let us introduce here $\{s\}$, a discrete state for each gene indicating whether the gene is rhythmic, $s_g = 1$, or non-rhythmic, $s_g = 0$. The probability on the $\{\alpha\}$ conditioned on the states $\{s\}$ takes the form:

$$\mathcal{P}(\{\alpha\}|\{s\}) \sim e^{-\frac{1}{2}\alpha_f^g A(\{s\})_{gg'}^{ff'} \alpha_{f'}^{g'}}, \quad (3.3)$$

where the covariance matrix is written as A . While these covariance matrices were kept arbitrary to show the most general formulation, we will make various approximations to allow us to proceed with analytical calculations. The prior on the $\{\varphi_c\}$ is taken as uniform, i.e. $\mathcal{P}(\varphi_c) \sim 1$, though this could be adapted easily, e.g. when it is known that samples are non-uniformly sampled (hospital working hours).

3.4 Approximations on the covariance matrices

Here we discuss approximations that will make calculations more tractable. We start by simplifying the error covariance matrix:

$$(S^{-1})_{cc'}^{gg'} = \delta_{cc'} \delta^{gg'} \sigma^2, \quad (3.4)$$

and we truncate the Fourier expansion at the first harmonic; the full model can then be rewritten as:

$$\begin{aligned} E_{gc} &= v_g + \varepsilon \quad \text{if } s = 0, \\ E_{gc} &= \mu_g + a_g \cos(\varphi_c) + b_g \sin(\varphi_c) + \varepsilon_{gc} = \alpha_g^\top \zeta_c + \varepsilon \quad \text{if } s = 1; \end{aligned} \quad (3.5)$$

with $\alpha_g^\top = (\mu_g, a_g, b_g)$, and $\zeta_c^\top = (1, \cos(\varphi_c), \sin(\varphi_c))$. We also simplify the prior on the gene parameters depending on s :

$$\mathcal{P}(\alpha|s=1) = \sqrt{\frac{\det(T^{-1})}{(2\pi)^3}} e^{-\frac{1}{2}\alpha^\top T^{-1} \alpha} \quad (3.6)$$

$$\mathcal{P}(v_g|s=0) = \delta(v_g - \langle E_{gc} \rangle_c) \quad (3.7)$$

with

$$T = \begin{pmatrix} u^2 & 0 & 0 \\ 0 & \tau^2 & 0 \\ 0 & 0 & \tau^2 \end{pmatrix}. \quad (3.8)$$

The error distribution is taken as:

$$\mathcal{P}(\varepsilon) \sim \prod_{gc} e^{-\frac{\varepsilon_{gc}^2}{2\sigma^2}}. \quad (3.9)$$

For the moment we will consider all genes to have $s = 1$ (one state model), and we will reintroduce the two state model with the Expectation-Maximization algorithm in paragraph (3.9). From $\mathcal{P}(\varepsilon)$ we can write:

$$\mathcal{P}(E|\{\varphi\}, \{\alpha\}) \sim \prod_{gc} e^{-\frac{(E_{gc} - \alpha_g \zeta_c)^2}{2\sigma^2}} = \prod_g \mathcal{P}(E_g|\{\varphi\}, \alpha_g). \quad (3.10)$$

After some algebra we obtain:

$$\mathcal{P}(E_g|\{\varphi\}, \alpha_g) \mathcal{P}(\alpha_g) \sim e^{-\frac{1}{2\sigma^2} (\alpha_g^T (M + \sigma^2 T^{-1}) \alpha_g - 2\alpha_g \cdot W_g)}, \quad (3.11)$$

with

$$M = \sum_c \zeta_c^T \zeta_c =: X^T X, \quad W_g = \sum_c E_{gc} \zeta_c. \quad (3.12)$$

As the goal is to identify the sample phases φ_c (ζ_c), we can marginalize over the α 's. After a multidimensional Gaussian integration we get:

$$\int \mathcal{P}(E_g|\{\varphi\}, \alpha_g) \mathcal{P}(\alpha_g) d\alpha_g \sim e^{\frac{1}{2\sigma^2} W_g^T (M + \sigma^2 T^{-1})^{-1} W_g}. \quad (3.13)$$

To obtain the set of ζ 's, we rewrite this as:

$$\mathcal{P}(\varphi|E) \sim e^{\sum_{ij} \zeta_i K_{ij} \zeta_j} \sim e^{-\beta H[\varphi]} \quad (3.14)$$

with

$$K_{ij} = M_\sigma^{-1} \sum_g E_{gi} E_{gj} \quad M_\sigma = M + \sigma^2 T^{-1}. \quad (3.15)$$

Here, we introduced the physics notation where β is the inverse temperature and H a Hamiltonian. We remind that each K_{ij} is a matrix, thus K is a 4-dimensional tensor of the form K_{ij}^{ab} .

3.5 Rewriting the problem as a spin model

For the moment we wrote the problem in function of $\zeta_c^\top = (1, \cos(\varphi_c), \sin(\varphi_c))$. However we can define a new variable that still carries all the phase information: $\eta_c^\top = (\cos(\varphi_c), \sin(\varphi_c))$. Doing this transformation will bring a more informative, although slightly more complicated, Hamiltonian. We define the J_{ij} matrix as the bottom right 2×2 matrix of K_{ij} , i.e. $J_{ij}^{ab} = K_{i,j}^{a+1,b+1}$ with $a, b \in \{1, 2\}$, and the h vector as the last two elements of the first column of K , i.e. $h_i^a = K_i^{a+1,3}$ with $a \in \{1, 2\}$. We notice how $K_{1,1}$ just contributes to a constant¹ that will be included in the normalisation of the probability which we are ignoring for the moment. So we have:

$$\mathcal{P}(\varphi|E) \sim e^{-\beta H[\varphi]} \quad (3.16)$$

with

$$H[\varphi] = - \left(\sum_{i,j=1}^{N_c} \eta_i J_{ij} \eta_j + \sum_{i=1}^{N_c} h_i \eta_i \right), \quad \beta = 1 \quad (3.17)$$

with J and h defined above. Here we observe two canonical contributions to an XY spin model. The first term in the Hamiltonian is an all-to-all spin interaction term with couplings J_{ij} . Due to the nature of the J_{ij} matrix, with both positive and negative entries of different magnitudes this Hamiltonian appears at a first glance as the one of a spin glass system [220–224]. Then we have a site-dependent external field term h_i that attempts to align all spins to an external local field.

3.6 Mean field approximation

Here we will derive a mean field approximation that allows us to calculate the unknown sample phases from a simple recursion relation. First, we take a closer look at the M matrix, which can be written as:

$$M = \begin{pmatrix} 1 \cdot 1 & 1 \cdot C & 1 \cdot S \\ C \cdot 1 & C \cdot C & C \cdot S \\ S \cdot 1 & S \cdot C & S \cdot S \end{pmatrix} \quad A \cdot B = \sum_c A_c B_c, \quad (3.18)$$

with $C = (\cos(\varphi_1), \dots, \cos(\varphi_{N_c}))$, $S = (\sin(\varphi_1), \dots, \sin(\varphi_{N_c}))$. The problem greatly simplifies when the samples are fairly uniformly distributed around the $24h$ cycle,

¹A constant shift in energy in a system is generally regarded as useless to study the behaviour of the system and especially its ground state.

which leads to the approximation that

$$C \cdot 1 = S \cdot 1 = C \cdot S = 0, \quad C \cdot C = S \cdot S = N_c/2 \quad (3.19)$$

Then, M , and also M_σ become diagonal, and thus the above matrix inversion (Eq.3.15) will be much easier.

Since $h_i = 0$ we obtain:

$$\mathcal{P}(\varphi|E) \sim e^{-\beta H[\varphi]} \quad (3.20)$$

with

$$H[\varphi] = -\sum_{ij} \eta_i J_{ij} \eta_j, \quad (3.21)$$

as from (3.19) we have $h = 0$ and J_{ij} defined as before. In addition we can now write precisely all the terms involved in the probability of interest, as some of them have changed slightly compared to (3.17): where

$$J_{ij} = \frac{1}{N_c N_g} \sum_g E_{gi} E_{gj}, \quad \beta = \frac{N_g}{\sigma^2} \frac{(N_c \tau^2)}{(\tau^2 N_c + 2\sigma^2)}. \quad (3.22)$$

The two normalizations on J_{ij} are only due to the physics of spin systems: they are there to keep the energy of the system finite regardless of its size. This is at first sight an Heisenberg spin-glass system. Fortunately, it will not be in the end a spin-glass, but only a spin system; this is due to the consistency present in the J matrix due to the E structure [225, 226], as we are basically measuring the angle between two spins (phases).

We now can find a first solution, a first approximation to the energy minimum of our system that should give us some ordering of our samples. The approximation we will exploit is one of the most versatile and is widely used in physics: the mean field [227–232]. We now derive the mean field approximation to obtain the expected phases contained in $\langle \eta_i \rangle$ (as $\eta_i = \eta_i(\varphi_i)$).

We remind for clarity:

$$\mathcal{P}(\eta|E) \sim e^{\beta \sum_{ij} \eta_i J_{ij} \eta_j} \quad (3.23)$$

We start from the set of coupled equations:

$$\langle \eta_i \rangle = \frac{1}{Z} \int d\varphi_i e^{\beta \sum_j J_{ij} \eta_i \eta_j} \eta_i(\varphi) \quad (3.24)$$

which then reduces using the mean-field for η_j in the exponent to:

$$\langle \eta_i \rangle = \frac{\int d\varphi_i e^{\beta \eta_i \sum_j J_{ij} \langle \eta_j \rangle} \eta_i}{\int d\varphi_i e^{\beta \eta_i \sum_j J_{ij} \langle \eta_j \rangle}}. \quad (3.25)$$

Since

$$\begin{aligned} \eta_i \sum_j J_{ij} \langle \eta_j \rangle &= W_i \cos(\varphi_i - \bar{\varphi}_i) \\ \text{with } \vec{W}_i &\equiv W_i \hat{W}_i \equiv W_i \begin{pmatrix} \cos(\bar{\varphi}_i) \\ \sin(\bar{\varphi}_i) \end{pmatrix} := \sum_j J_{ij} \langle \eta_j \rangle, \end{aligned} \quad (3.26)$$

where we have decomposed the vector \vec{W}_i into the product of a unit vector \hat{W}_i and a modulus W_i . Eq.(3.25) can be solved easily using Bessel functions, in particular:

$$\langle \eta_i \rangle = \frac{I_1(\beta W_i)}{I_0(\beta W_i)} \hat{W}_i. \quad (3.27)$$

We can solve equation (3.27) recursively starting from a random initial condition. Doing so we rapidly approach the minimum of the energy of our system and obtain a first circadian phase assignment for the samples in our dataset.

3.7 The emergence of a phase transition

This section is not present in the original appendix, however it contains interesting insights cut for length constraints. The recursive solution we found from equation (3.27) sheds a clear light on the existence of a phase transition in β . In the disordered phase our samples do not have a proper circadian phase so it is impossible for us to infer it; there simply isn't enough information in the system to order the samples. While in the low temperature (high β) phase the state that minimizes the energy is a good proxy for the actual state of the system. Thus, we can reliably infer the circadian phases of our samples if β becomes big enough. The existence of the phase transition can be easily seen with a high temperature expansion. Namely we take the first order in β :

$$\langle \eta_i \rangle = \frac{1}{2} \beta W_i \hat{W}_i = \frac{\beta}{2} \vec{W}_i = \frac{\beta}{2} \sum_j J_{ij} \langle \eta_j \rangle. \quad (3.28)$$

Which is nothing more than

$$\vec{\eta} = \frac{\beta}{2} J \vec{\eta} \quad (3.29)$$

with

$$\vec{\eta} := (\eta_1 \dots \eta_i \dots \eta_{N_c})^\top. \quad (3.30)$$

This shows us how the first non trivial solution appears at a finite temperature β_c :

$$\beta_c = \frac{2}{\lambda_1} \quad (3.31)$$

with λ_i the eigenvalues of the J matrix, ordered in a decreasing order. The existence of a phase transition signifies the need for a high enough β in order to have an ordered solution. Given the meaning of β this means that, as we expect, a signal to noise ratio too low doesn't allow for a re-ordering of our samples. However, the mean-field solution plays a crucial role as we use it to seed a much more accurate algorithm, which we now describe.

3.8 The two-state model

While the phases inferred in by this XY mean-field model are typically ordered properly when tested in datasets with known time stamps (e.g. we used [22]), the actual values are not sufficiently accurate. In the model below, we will relax the assumption that phases are homogeneously along the $24h$ (Eq.3.19).

3.9 The EM algorithm

To develop a more precise method to infer latent phases we will proceed with an Expectation Maximisation[233, 234] approach [235]. In the following calculations we have mean centered our data matrix E by rows. However, this does not mean that μ_g is zero in this model (Eq. 3.5) since the density of phases may not be constant along the cycle while it imposes $v_g = 0 \forall g$. We will no longer assume that the phases are uniformly distributed around the cycle, nor the genes are all in state one ($s = 1$) and consider a mixture model for states $s \in \{0, 1\}$. The rationale for the mixture model is to allow that some genes may not carry phase information in all the sets of samples, for example if one gene is less expressed or noisy in a given tissue. In the iterative EM scheme, we will denote the new objects at any given step of the algorithm with a prime (''). We will also drop the gene index and the vector notation, as now the objects in play should be clear. The so called "Q-function" in the EM algorithm takes the form:

$$Q(\theta|\varphi) = \langle \log(\mathcal{P}(E, \alpha, s|\theta)) \rangle_{\mathcal{P}(\alpha, s|E, \varphi)} . \quad (3.32)$$

In particular, the phases at the next step are:

$$\varphi' = \underset{\theta}{\operatorname{argmax}}(Q(\theta|\varphi)) . \quad (3.33)$$

From now on we will abuse our notation and generally substitute θ with φ' so that (3.32) becomes:

$$Q(\varphi'|\varphi) = \langle \log(\mathcal{P}(E, \alpha, s|\varphi')) \rangle_{\mathcal{P}(\alpha, s|E, \varphi)} \quad (3.34)$$

and (3.33)

$$\varphi' = \underset{\varphi'}{\operatorname{argmax}}(Q(\varphi'|\varphi)) . \quad (3.35)$$

As our aim is to maximise Q which involves taking a logarithm we can forget about some normalisation and we can write from now on:

$$\mathcal{P}(E, \alpha, s|\varphi') = \mathcal{P}(E|\alpha, s, \varphi')\mathcal{P}(\alpha, s|\varphi') = \mathcal{P}(E|\alpha, s, \varphi')\mathcal{P}(\alpha|s)\mathcal{P}(s) . \quad (3.36)$$

Our prior on the gene parameters is, depending on the state:

$$\mathcal{P}(\alpha_g|s=1, \varphi) = \sqrt{\frac{\det(T^{-1})}{(2\pi)^3}} e^{-\frac{1}{2}\alpha_g^\top T^{-1}\alpha_g} , \quad (3.37)$$

or

$$\mathcal{P}(v_g|s=0, \varphi) = \delta(v_g) . \quad (3.38)$$

As

$$\log(\mathcal{P}(E, \alpha|\varphi')) = \log\left(\prod_g \mathcal{P}(E, \alpha|\varphi')_g\right) = \sum_g \log(\mathcal{P}(E, \alpha|\varphi')_g) , \quad (3.39)$$

taking into account (3.36), we have for each gene (considering only what we will need for the maximization):

$$Q(\varphi'|\varphi) = \int d\alpha (E - X'\alpha)^2 e^{\frac{1}{2\sigma^2}(\alpha - \hat{\alpha})^\top M_\sigma(\alpha - \hat{\alpha})} \mathcal{P}(s=s^1|E, \varphi) , \quad (3.40)$$

where X' is the same as implicitly defined in (3.12) and we defined $t_g := \mathcal{P}(s_g = s^1|E, \varphi)$.

Solving the integral and putting everything together yields for the part that will be relevant for the maximisation of Q :

$$Q(\varphi'|\varphi) = - \sum_g t_g (E_g^\top E_g - 2E_g^\top X' \hat{\alpha}_g + \hat{\alpha}_g^\top M' \hat{\alpha}_g) - \operatorname{Tr}(\sigma^2 M_\sigma^{-1} M') \quad (3.41)$$

So all we need to do is calculate the t_g weights, which are equal to the probability

that each gene is rhythmic ($s = 1$). To do so we need to apply Bayes theorem:

$$\mathcal{P}(s|E, \varphi) = \frac{\mathcal{P}(E|s, \varphi)\mathcal{P}(s)}{\sum_s \mathcal{P}(E|s, \varphi)\mathcal{P}(s)}, \quad (3.42)$$

where:

$$\mathcal{P}(E|s, \varphi) = \int d\alpha \mathcal{P}(E|\alpha, s, \varphi)\mathcal{P}(\alpha|s, \varphi), \quad (3.43)$$

and where we have used that the prior $\mathcal{P}(s|\varphi) = \mathcal{P}(s)$. We also need to correctly normalise all the probabilities in play:

$$\int \mathcal{P}(E|\alpha, s = 1, \varphi) \prod dE = \int e^{-\frac{(E-\alpha\zeta)^2}{2\sigma_1^2}} dE = (2\pi\sigma_1^2)^{m/2} = Z_E^1 \quad (3.44)$$

with m the number of entries of the E matrix. For state zero ($s = 0$) we find

$$\int \mathcal{P}(E|\alpha, s = 0, \varphi) \prod dE = \int e^{-\frac{E^2}{2\sigma_0^2}} dE = (2\pi\sigma_0^2)^{m/2} = Z_E^0. \quad (3.45)$$

We now do all the calculations omitting the gene index, as it only burdens the notation. Marginalising the α yields

$$\begin{aligned} \mathcal{P}(E|s = 1, \varphi) &= \int d\alpha \mathcal{P}(E|\alpha, s = 1, \varphi)\mathcal{P}(\alpha|s = 1, \varphi) \\ &= \frac{1}{Z_E^1} \sqrt{\frac{\det(T^{-1})}{\det(M_\sigma/\sigma_1^2)}} e^{\frac{1}{2\sigma_1^2}(E^\top X M_\sigma^{-1} X^\top E - E^\top E)}. \end{aligned} \quad (3.46)$$

In addition, we can also easily calculate the other probability ($s = 0$):

$$\mathcal{P}(E|s = 0, \varphi) = \frac{1}{Z_E^0} e^{-\frac{E^\top E}{2\sigma_0^2}}. \quad (3.47)$$

Finally we obtain the gene weight

$$t = \frac{\gamma q e^{\frac{1}{2\sigma^2}(E^\top X M_\sigma^{-1} X^\top E)}}{1 - q + \gamma q e^{\frac{1}{2\sigma^2}(E^\top X M_\sigma^{-1} X^\top E)}}, \quad (3.48)$$

with

$$\gamma = \left(\frac{\sigma_0^2}{\sigma_1^2}\right)^{M/2} \sqrt{\frac{\det(T^{-1})}{\det(M_\sigma/\sigma^2)}} e^{\frac{E^\top E}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)}. \quad (3.49)$$

The final expression for Q and its derivatives reads:

$$Q(\varphi'|\varphi) = -\sum_g t_g \left(E_g^\top E_g - 2E_g^\top X' \hat{\alpha}_g + \hat{\alpha}_g M' \hat{\alpha}_g - \text{Tr}(\sigma_g^2 M_{\sigma_g}^{-1} M') \right) \quad (3.50)$$

$$\begin{aligned}
 0 &= \frac{\partial Q(\zeta'|\zeta)}{\partial C'_i} = \\
 \sum_g t_g (\hat{a}_g(-E_{gi} + \hat{\mu}_g + \hat{a}_g C_i + \hat{b}_g S_i) + (M_\sigma^{-1})_{2,1} + (M_\sigma^{-1})_{2,2} C_i + (M_\sigma^{-1})_{2,3} S_i) &= 0
 \end{aligned} \tag{3.51}$$

$$\begin{aligned}
 0 &= \frac{\partial Q(\zeta'|\zeta)}{\partial S'_i} = \\
 \sum_g t_g (\hat{b}_g(-E_{gi} + \hat{\mu}_g + \hat{a}_g C_i + \hat{b}_g S_i) + (M_\sigma^{-1})_{3,1} + (M_\sigma^{-1})_{3,2} C_i + (M_\sigma^{-1})_{3,3} S_i) &= 0.
 \end{aligned} \tag{3.52}$$

This set of equation can be compactly rewritten as:

$$K \bar{\zeta}_i = \bar{O}, \tag{3.53}$$

with the evident definitions.

However, since we want to keep the constraint that $|\zeta|^2 = 1$ we need to introduce Lagrange multipliers. As a reminder: In general, we want to maximise $f(x)$ where x satisfies $g(x) = 0$. To do this, we can maximise $h(x, \lambda) = f(x) + \lambda g(x)$. If there is more than one constraint, we need to introduce one multiplier for each constraint. In our case:

$$h(\zeta', \lambda) = Q(\zeta'|\zeta) + \sum_i \lambda_i (C_i^2 + S_i^2 - 1). \tag{3.54}$$

So, as before, if we consider sample i we have:

$$(K + \lambda I) \bar{\zeta}_i = \bar{O} \tag{3.55}$$

and

$$C_i^2 + S_i^2 = 1. \tag{3.56}$$

So we need to find the λ such that

$$\left| (K + \lambda I)^{-1} \bar{O} \right|^2 = 1. \tag{3.57}$$

To solve (3.57) we calculate explicitly the fourth order polynomial in λ and numerically find its roots. Among these roots there are the local maxima and minima of the Q function. Although not trivial, we can find the exact solution for this problem. If we write:

$$\bar{O} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \quad K = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \tag{3.58}$$

we find

$$\begin{aligned} & \lambda^4 + \lambda^3(2A + 2D) + \lambda^2(D^2 + A^2 + 4AD - \alpha^2 - \beta^2 - 2BC) + \\ & + \lambda(2AD(A + D) - 2BC(A + D) - 2D\alpha^2 - 2A\beta^2 + 2\alpha\beta(B + C)) + \\ & + B^2C^2 + A^2D^2 - \alpha^2(D^2 + C^2) - \beta^2(A^2 + B^2) + 2\alpha\beta(AB + CD) - 2ABCD = 0. \end{aligned} \quad (3.59)$$

Solving this numerically provides us with 2 or 4 different possible values of φ' . In order to find which φ' gives rise to the max of Q we simply evaluate the Q function in all the points. This can be done separately for each condition as the derivatives become independent from one another. With this we have solved the minimization process and have found our φ' thus we have concluded an iteration of the EM algorithm.

3.10 An interesting equivalence

This section is not present in the original appendix, however it contains a curious equivalence which some might like to read. The normalisation factor of (3.46) could be also obtained in a direct way. In fact, we could simply integrate:

$$\int e^{\frac{1}{2\sigma_1^2}(E^\top X M_\sigma^{-1} X^\top E - E^\top E)} dE = \sqrt{\frac{(2\pi\sigma_1^2)^M}{\det(1 - X M_\sigma^{-1} X^\top)}} \quad (3.60)$$

Thus our new γ as defined by direct integration above would be:

$$\gamma = \left(\frac{\sigma_0^2}{\sigma_1^2}\right)^{M/2} \sqrt{\det(1 - X M_\sigma^{-1} X^\top)} e^{\frac{E^\top E}{2} \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)} \quad (3.61)$$

Although not completely intuitive for multidimensional case the two normalization factors are, as they should be, identical:

$$\frac{\det(T^{-1})}{\det(M_\sigma/\sigma^2)} \equiv \det(1 - X M_\sigma^{-1} X^\top). \quad (3.62)$$

3.11 CHIRAL performance

We benchmarked our algorithm using time-labelled samples from 12 mice tissues collected around the clock applying CHIRAL in a tissue-by-tissue manner [18]. Our first approach was to try CHIRAL without the use of seed clock genes. We can see in figure 3.1 that the reconstruction is suboptimal for tissue with weak rhythmic

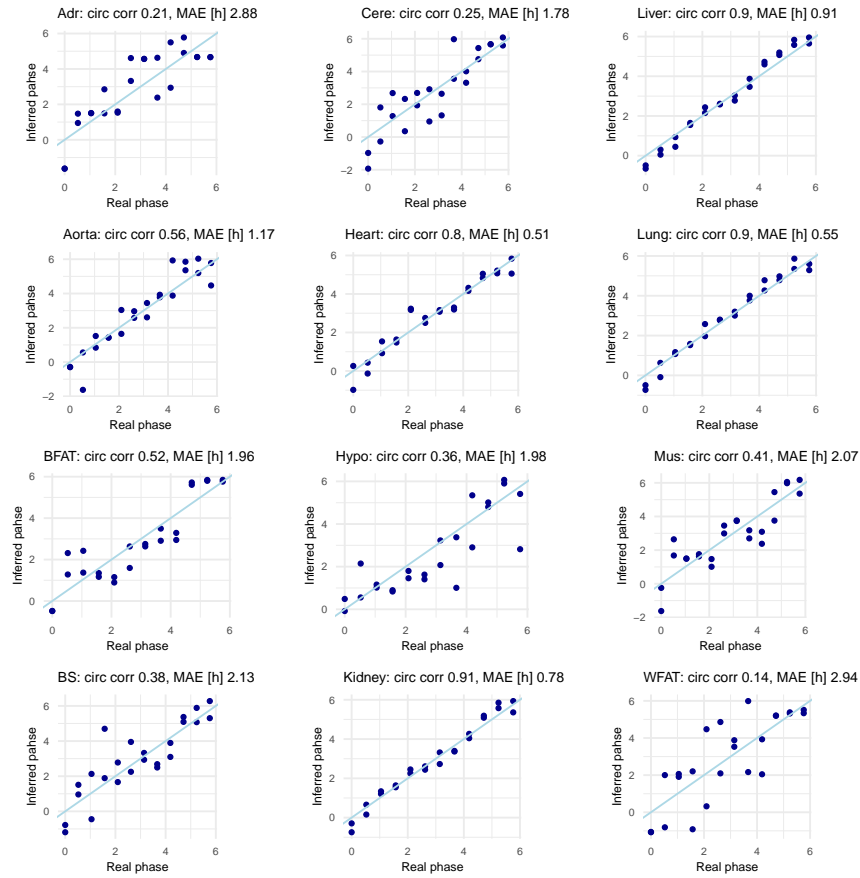


Figure 3.1: **CHIRAL reconstruction on top 10000 expressed genes.** Samples from tissues with strong clocks are correctly ordered, while weaker clocks do not carry dominant information. The names are abbreviated as: Adr (adrenal gland), Cere (cerebellum), BFAT (brown fat), Hypo (hypothalamus), Mus (skeletal muscle), BS (brainstem), WFAT (white fat) [18].

behaviour such as brain tissues. However, for some metabolic tissues (Liver, Lung) the Median Absolute Error (MAE) is comparable or better than what is reported in the literature [15, 22–26, 213].

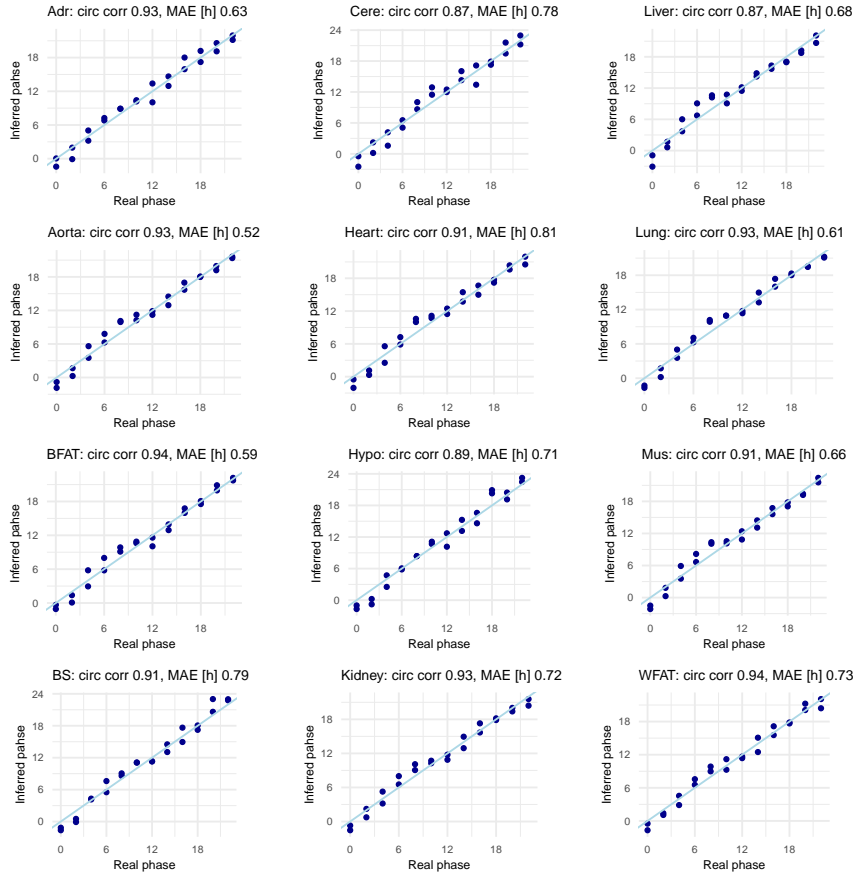


Figure 3.2: **CHIRAL reconstruction using clock reference genes.** Samples from all tissues are well-ordered. The names are abbreviated as: Adr (adrenal gland), Cere (cerebellum), BFAT (brown fat), Hypo (hypothalamus), Mus (skeletal muscle), BS (brainstem), WFAT (white fat) [18].

This re-assures us that our method has some merit. To decisively outperform the only unsupervised counterpart, CYCLOPS [25], we will reduce the set of genes to ones that we know from literature carry temporal information. With this update in seed genes we can see in figure 3.2 the increased performance of CHIRAL on the 12 mice tissues used previously. Our objective is to use CHIRAL on human data RNA-seq, so mice microarray data might not be the best testing bench. For this we exploited one of the few labelled human dataset comprising 54 samples for human muscle biopsies around the 24 hours [236].

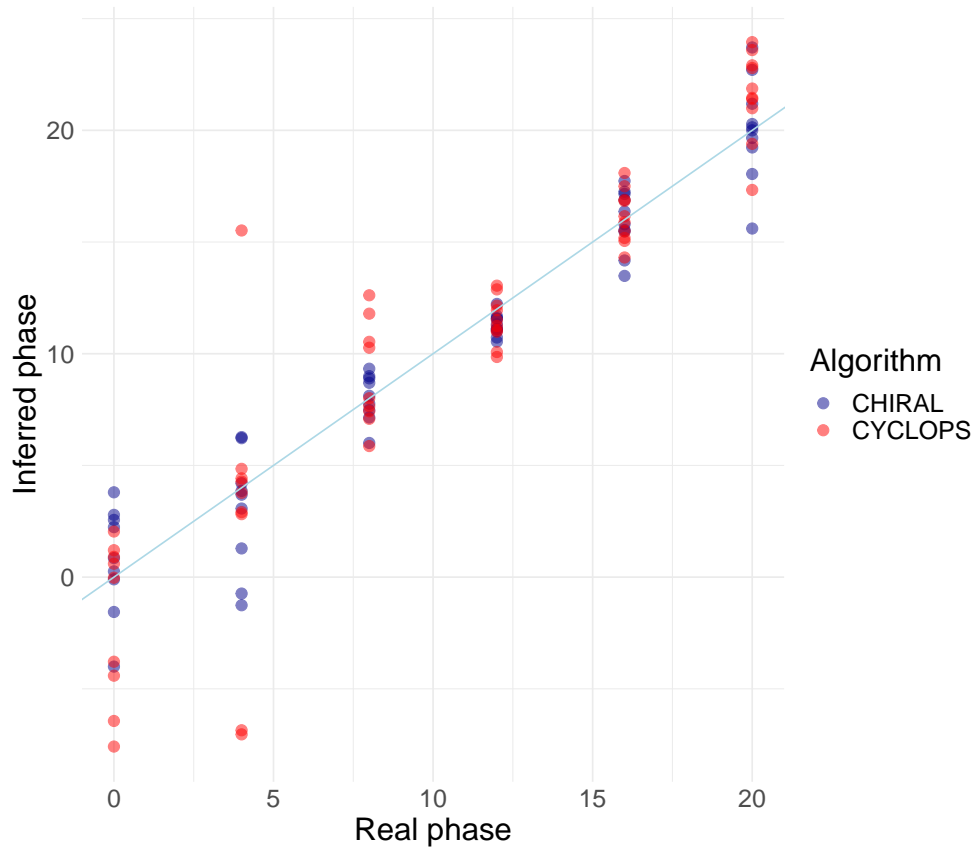


Figure 3.3: **Comparison of CHIRAL and CYCLOPS on human muscle biopsies.** The reconstructed phase of CHIRAL is generally more representative of the true phase at which the measurement was taken. Some differences can be due to chronotype and inflammatory response. X and y axes are hours.

We also applied CYCLOPS to this dataset and compared the two algorithm to ensure that our procedure is the best currently available; the results of both inferences are shown in figure 3.3 and some relevant metrics calculated for both algorithms are summarised in table 3.1. The results clearly confirm that CHIRAL is the better choice of phase inference algorithm.

Table 3.1: **CHIRAL validation on human muscle biopsies.** Important metrics for phase reconstruction algorithms shown for CHIRAL and CYCLOPS show the better performance of CHIRAL.

	CHIRAL	CYCLOPS
Circular Correlation	0.76	0.53
Mean absolute error [h]	1.49	2.16
Median absolute error [h]	0.92	1.14
Root of square mean error [h]	1.94	3.16

3.12 Practicalities

Up to now we have considered the set of all genes in the given set of mRNA measurements. However, given the completely unsupervised nature of this algorithm we cannot, a priori, be sure of which oscillatory process will be picked up (e.g. cell cycle or circadian clock). To be sure that we are capturing the circadian time as a source of variation we can apply our algorithm to a subset of genes. In particular, in this case, we selected the 12 clock reference genes (CRGs): DBP, PER3, TEF, NR1D2, PER1, PER2, NPAS2, ARNTL, NR1D1, CRY1, CRY2, CIART. Also, we found that the prior on the α s needs to be tightly controlled as it "competes" against the experimental data and we know that genes do not have oscillation of more than 10 – 20 log2 units. So for the prior we pick $u^2 = 0.2$, $\tau^2 = 4/(24 + n)$ where n is the number of samples.

Human chronobiology **Part III**

4 Sex-dimorphic and age-dependent organization of 24-hour gene expression rhythms in human

Now we will go into the applications of CHIRAL to human post-mortem bulk RNA-seq data. This chapter has been adapted as little as possible from the preprint of [219] written by Lorenzo Talamanca, Cedric Gobet and Felix Naef. This work has at present been accepted for publication in *Science*. The research was developed along the way by interactions of the three; the original research question envisioned by Felix Naef has been answered in chapter 3.

This work in its entirety had been supervised and checked by Felix Naef; the analysis, coding, writing and illustrations were made by Lorenzo Talamanca and Cedric Gobet in equal parts.

4.1 Abstract

The circadian clock modulates human physiology. However, the organization of tissue-specific gene expression rhythms and how these depend on age and sex is not defined in humans. We leveraged the Genotype-Tissue Expression project (GTEx) with an algorithm that assigns circadian phases to 914 donors, by integrating temporal information from multiple tissues in each individual, to identify mRNA rhythms in 46 tissues. Clock transcripts showed conserved timing relationships and tight synchrony across the body. mRNA rhythms varied in breadth, covering global and tissue-specific functions, including metabolic pathways and systemic responses. The circadian clock was conserved across sexes and age groups. However,

overall gene expression rhythms were highly sex-dimorphic and more sustained in females. Rhythmic programs generally dampened with age across the body.

4.2 Introduction

The circadian clock allows evolutionary adaptation of life to the 24-hour periodicity of earth rotation. The clock synchronizes internal body rhythms in behavior and physiology with 24-hour environmental, societal or feeding cues [29, 237, 238]. Perturbations of the clock, as caused by sleep disruption and shift work can lead to pathologies [218]. Sexual dimorphism exists in gene expression across the body [239] and many complex phenotypes, including diseases, exhibit sex-dependent characteristics [240]. However, interactions between sexual dimorphism and circadian rhythms in humans are unexplored [101]. Likewise, the effects of aging on human physiology are well studied [241], but the interplay between circadian oscillations and aging processes are still poorly understood [242]. We combined GTEx transcriptomes [243] with an algorithm that assigns circadian times to individuals and tissues [25, 236] to obtain a whole organism view of 24-hour gene expression rhythms in 46 human tissues. A stratification by sex and age revealed rich picture of group-specific rhythms, especially in metabolic and cardiovascular tissues that may provide insights into differential disease incidence rates.

4.3 Comprehensive 24h gene expression rhythms in humans

To study the breadth of rhythmic mRNA programs across the human body, we leveraged 16k human RNA-seq experiments from 914 donors in the Genotype-Tissue Expression (GTEx) collection, and computed one circadian reference phase for each donor. This phase corresponds to the expected circadian phase in skeletal muscle, hereafter named the donor internal phase (DIP). The algorithm exploits that the circadian phases of tissue samples (typically 10 to 20 per individual) from the same donor are correlated, and makes the assumption that relative circadian phases of tissues are conserved across donors. Time of death (TOD) (available from GTEx) may not reflect circadian phase due to the varying individuals' chronotypes [17], positions in a time zone and type of death. In relation to the TOD, clock genes such as *PER2* and *NR1D1* exhibited arrhythmic profiles in most tissues (figure 4.1a), and mRNA rhythmicity at the genome-scale was nearly absent (figure 4.1b).

However, the pairwise correlations between clock transcripts were indicative of a

4.3 Comprehensive 24h gene expression rhythms in humans

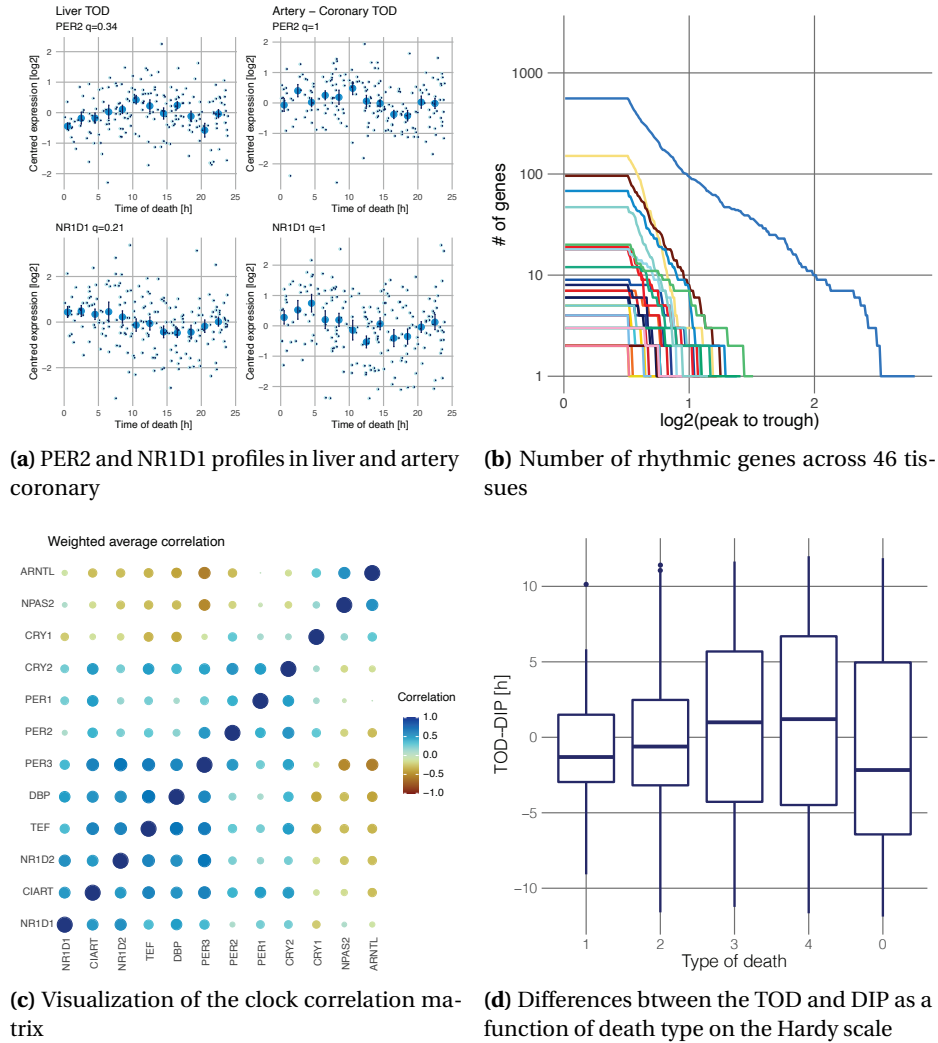


Figure 4.1: Failure of the TOD to capture circadian phase and necessity of the DIP. (a) Clock genes PER2 and NR1D1 exhibit statistically arrhythmic profiles in liver and artery coronary as a function of GTEx time of death (TOD). Log2-centered expression for each donor (light blue). Mean and standard error (SE) computed in 2h bins are also shown (dark blue). (b) Number of rhythmic genes ($q > 0.2$ & \log_2 peak-trough > 0.5 with peak-trough amplitude higher than a threshold (x-axis, \log_2)) across 46 tissues using indicated time of death. (c) Weighted correlation matrix of CRGs, see section 4.7, is indicative of a functioning circadian clock across all tissues of the GTEx collection. (d) Boxplot of the difference between time of death (TOD) and donor internal phase (DIP) as a function of type of death (hardy scale).

Chapter 4. Sex-dimorphic and age-dependent organization of 24-hour gene expression rhythms in human

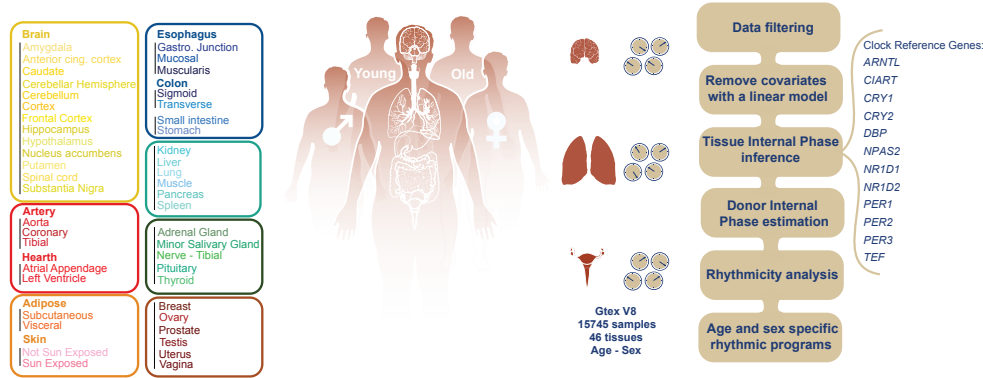


Figure 4.2: Scheme of the steps to infer donor internal phase Summary of the steps taken to assign one circadian phase (DIP) to 914 donors in the GTEx v8 RNA-seq dataset. This allows us to analyze 46 tissues, and to compare phases of gene rhythms across tissues. Left: Color scheme and list of all tissues present in our analysis.

functional clock (Fig. S1C). We therefore developed an algorithm to assign DIPs to all donors (figure 4.2). After correcting for sample covariates such as age, sex, ischemic time (time between death and sample collection), and type of death to reduce variability not due to circadian oscillations (figure 4.3), we applied two steps. For each tissue independently, we used CHIRAL, an algorithm we developed to estimate the tissue internal phases (TIPs) of all samples. We benchmarked CHIRAL using time-labeled human samples from muscle [236] figure 3.3 and table 3.1. As the TIPs in a donor's tissues often showed one primary mode, we assigned the DIPs to that mode (Fig. S1G-H). With the DIPs, we characterized gene expression rhythmicity in 46 tissues by harmonic regression. DIPs were distributed fairly uniformly along the 24-hour cycle (Table S2).

TODs and DIPs showed better concordance for fast compared to slow death (Hardy scale), figure 4.1d. With the DIPs, PER2 and NR1D1 showed clear circadian oscillations across all donors (figure 4.5a), indicating that the DIPs captured circadian phase more reliably than TODs. Clock transcript oscillations across all tissues showed that, although the amplitudes varied, the peak times were aligned, the tightest being TEF and ARNTL (BMAL1), and the most variable NR1D1 (REVERBA) (figure 4.5b).

We used the complex-valued singular value decomposition (cSVD) to summarize the multi-gene structure of clocks across multiple conditions, explained in detail in section 4.7. The first mode, which captured > 95% of the variance, showed that the

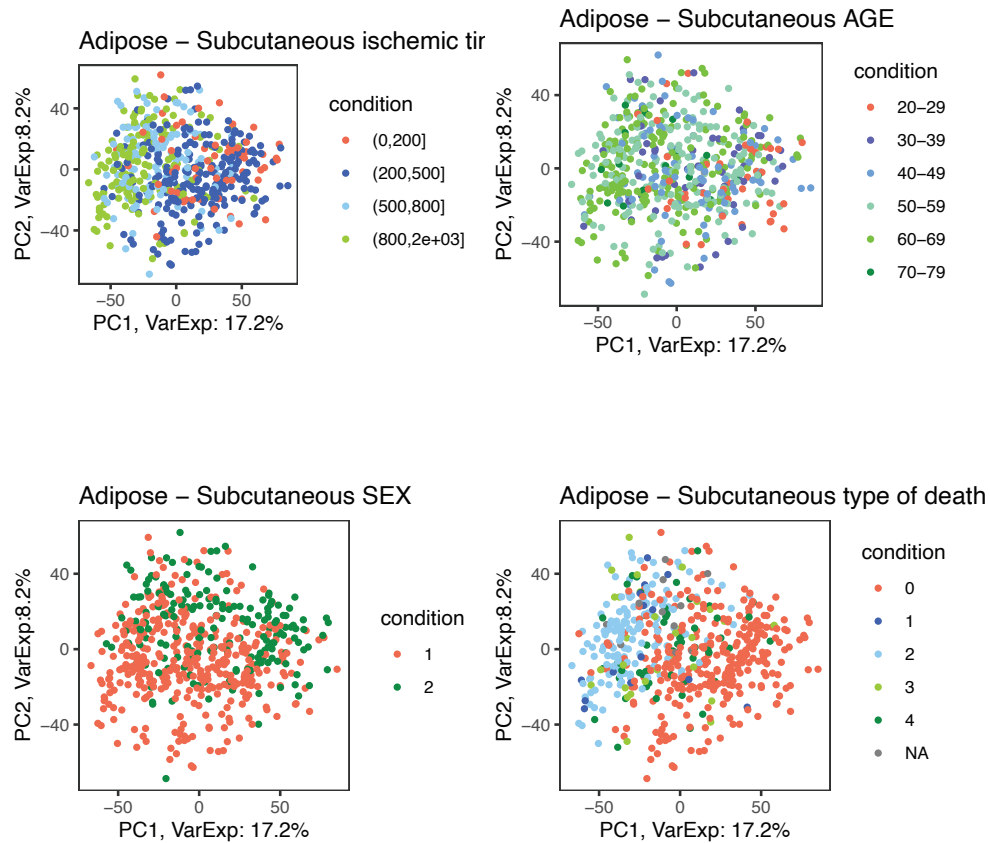


Figure 4.3: **Importance of covariates in the GTEx data** Principal Component Analysis (PCA) analysis of samples for adipose subcutaneous in PC1-PC2 space color coded according to different meta variables calls for removal of undesired sources of variability in the data. PC1 and PC2 explain respectively 17.2% and 8.2% of the total variance.

Chapter 4. Sex-dimorphic and age-dependent organization of 24-hour gene expression rhythms in human

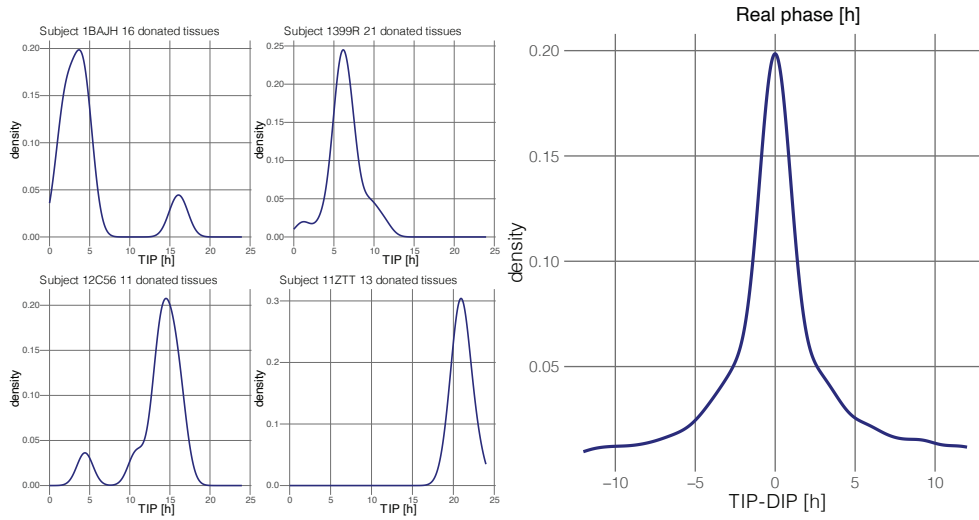


Figure 4.4: Donor specific peaked TIP distribution. Illustrative TIP distributions with one clear peak suggest the possibility of inferring the DIP. Distribution of the TIP (shifted by the DIP) for all samples and donors shows a clear and tight peak.

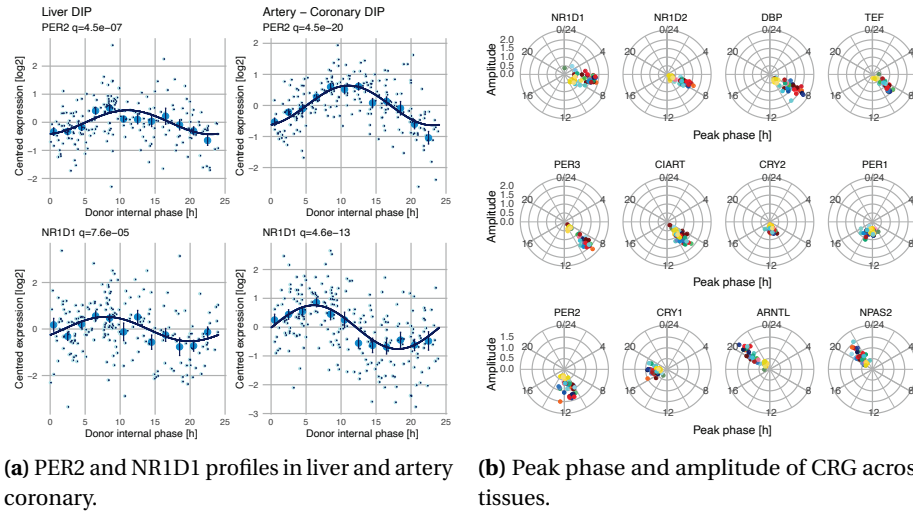


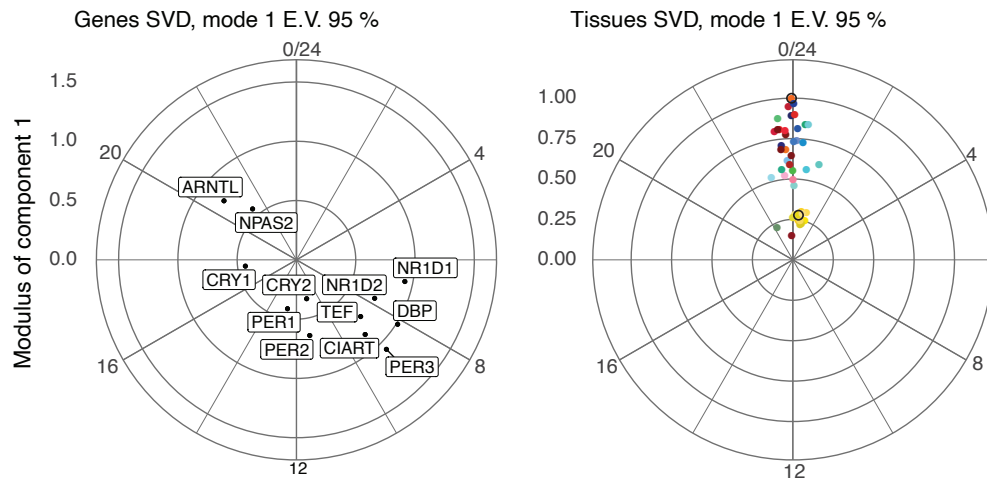
Figure 4.5: First results of DIP inference. (a) Clock reference genes PER2 and NR1D1 exhibit robust oscillatory behavior in metabolic tissues as a function of DIP. Mean and standard error (SE) computed in 2h bins, and harmonic regression fits, are also shown as in figure 4.1a. (b) With the DIP, clock gene phases can be compared across tissues. Phases (clockwise) and amplitudes (radial) of CRGs across tissues highlight overall tight phases. NR1D1 is the most variable gene. Tissues are color coded as in figure 4.2

human clock module comprises two main groups of antiphasic genes, plus fewer genes with a phase angle (figure 4.6a). Clocks across tissues were well synchronized showing relative offsets of only a few hours; the adrenal gland had the earliest phase, perhaps related to the distinct role of adrenal glucocorticoids in systemic clock organization [244]. Metabolic tissues (adipose tissue, esophagus, cardiovascular tissues) showed the highest amplitudes whereas brain tissues and testis had lowest 4.6a [245], as reflected by the clock genes *PER3* and *ARNTL* (figure 4.6b).

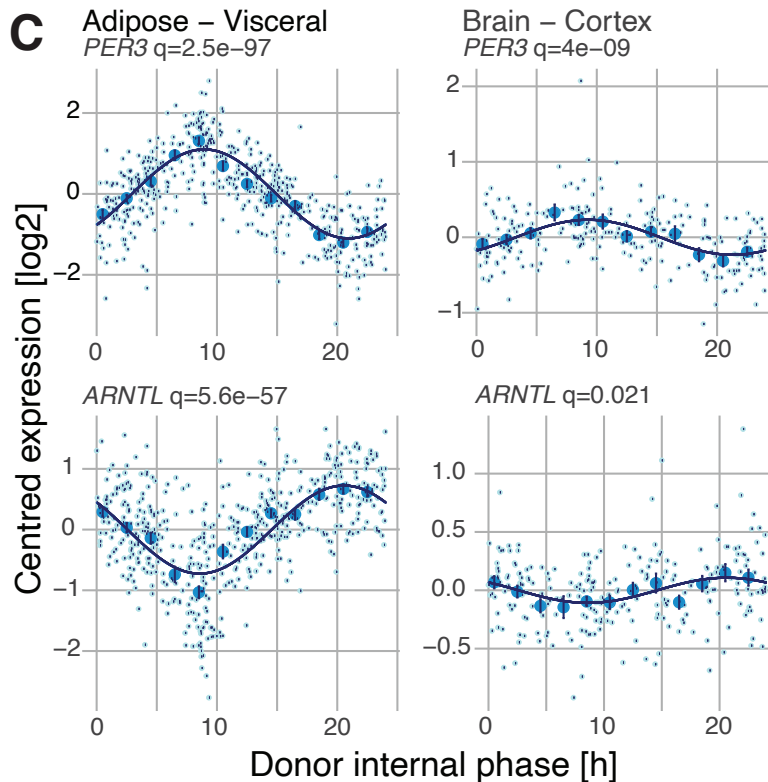
To test whether the DIPs also unveil mRNA rhythms associated with systemic signals, we considered heat shock response genes, which exhibit rhythmic activity in mice [246]. Heat shock genes (*HSF1* targets) showed clear diurnal expression patterns with highest oscillatory amplitude in brain tissues, peaking between 8 and 10pm, near the time of highest body temperature in humans [247] figure 4.7a. Compared to the clock, we observed a larger spread in peak phases across tissues; the peak times of both *HSPH1* and *HSPA1B* were almost antiphasic in spleen and amygdala (figure 4.7b). The high amplitude heat shock program in many brain regions may reflect a pressure for high proteome integrity in non-renewing tissues [248]. Genome-wide 24-hour rhythmicity (Fig. S1L, Table S3 and S4) across tissues showed morning (centered on 7am) and evening (7pm) waves of gene expression throughout the body (figures 4.8c), with metabolic tissues showing the most and brain tissues the least rhythmicity [18, 20] (figure 4.8c). These waves showed slight temporal shifts following the phases of the core clock (figure 4.6a): several glands showed early phases, followed by cardiovascular, metabolic and brain tissues (figure 4.8c). Depending on the tissue, we found between tens and several hundreds of rhythmic transcripts with peak-to-trough amplitudes higher than 2-fold (figure 4.8c). Besides clock genes, more than 100 transcripts were rhythmic in at least 20 tissues, including known rhythmic genes such as *NFIL3* and *PDK4* as well as glucocorticoid responsive genes such as *FKBP5* and pro-inflammatory cytokine receptors, Interleukin 1 Receptor Like 1/2 (*IL1RL1*, *IL1R2*) (Table S4). 12-hour ultradian mRNA rhythms were detected in several tissues, notably in ovary and liver (Fig. S1M).

To characterize regulatory mechanisms, we used cSVD to integrate transcription factor (TF) targets, further details in section 4.7. Putative regulators of both the morning and evening waves were involved in immunity, core clock, carbohydrate metabolism and cell proliferation (Fig. S1N, Table S5). Among TFs explaining the most variance were the core clock dimer *CLOCK:BMAL1* (peak target accumulation at 10am) and glucocorticoid receptor (GR) *NR3C1* (5pm) corresponding to GR

Chapter 4. Sex-dimorphic and age-dependent organization of 24-hour gene expression rhythms in human



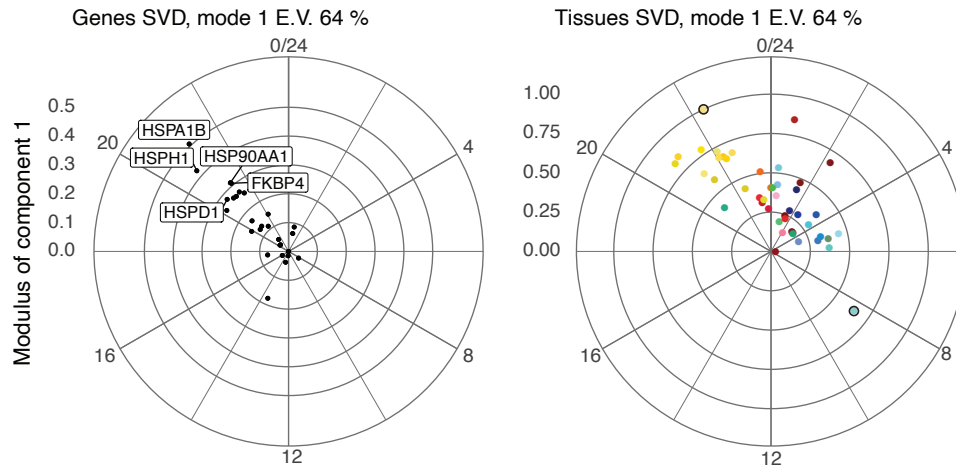
(a) Structure of clock reference genes.



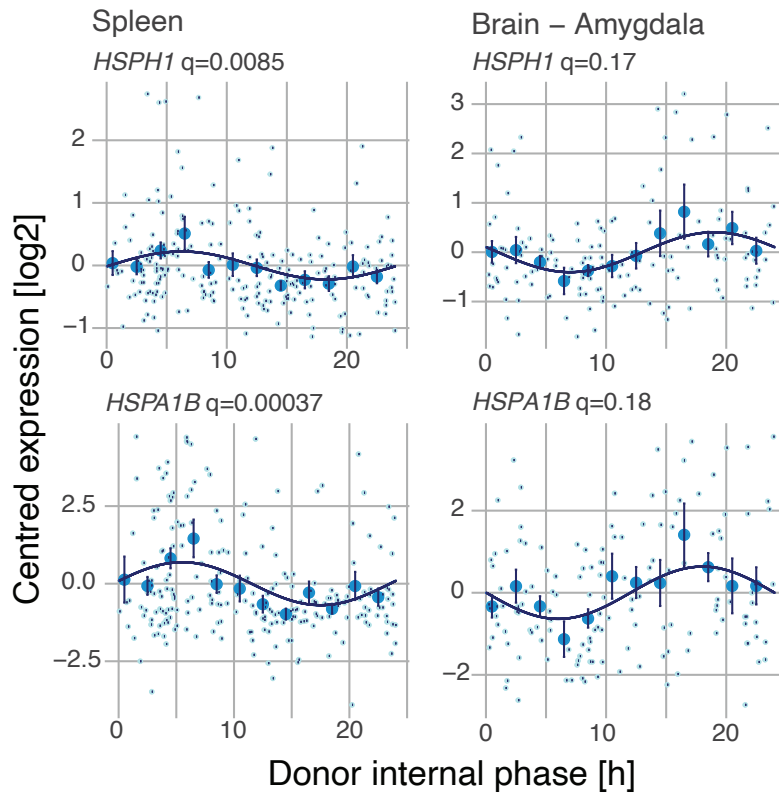
(b) PER3 and ARNTL profiles in adipose tissue and cortex.

Figure 4.6: **Clock reference gene structure using DIPs.** (a) First gene (left) and tissue (right) complex vectors from the cSVD performed on clock reference genes (CRGs) show coherent phase and amplitude relationships across tissues (first module captures 95% of the 24-hour variance, E.V.). The core clock is also tightly synchronized across the human body (spread of 2 hours, tissue vector). Brain tissues have the lowest clock gene amplitudes and the adrenal gland peaks earliest. Tissues are color-coded as depicted in figure 4.2. Polar plots: time runs clockwise, phases have been converted to 24-hour time. (b) mRNA expression levels (log2, centered) of two clock genes (PER3 and BMAL1) in two representative tissues. Mean and standard error (SE) computed in 2h bins, and harmonic regression fits are shown (dark blue).

4.3 Comprehensive 24h gene expression rhythms in humans



(a) Structure of heat shock response genes.



(b) *HSPH1* and *HSPA1B* profiles in spleen and amygdala.

Figure 4.7: **Heat shock response genes using DIPs.** (a) Heat shock gene module displayed as in figure 4.6a. The genes representing the top 20 HSF1 targets (ChIP-Atlas) show time of day specific expression in all tissues, with higher amplitudes and earlier peak times in brain tissues (yellow). (b) mRNA expression levels (log2, centered) for two heat shock response genes, represented as in figure 4.6b

Chapter 4. Sex-dimorphic and age-dependent organization of 24-hour gene expression rhythms in human

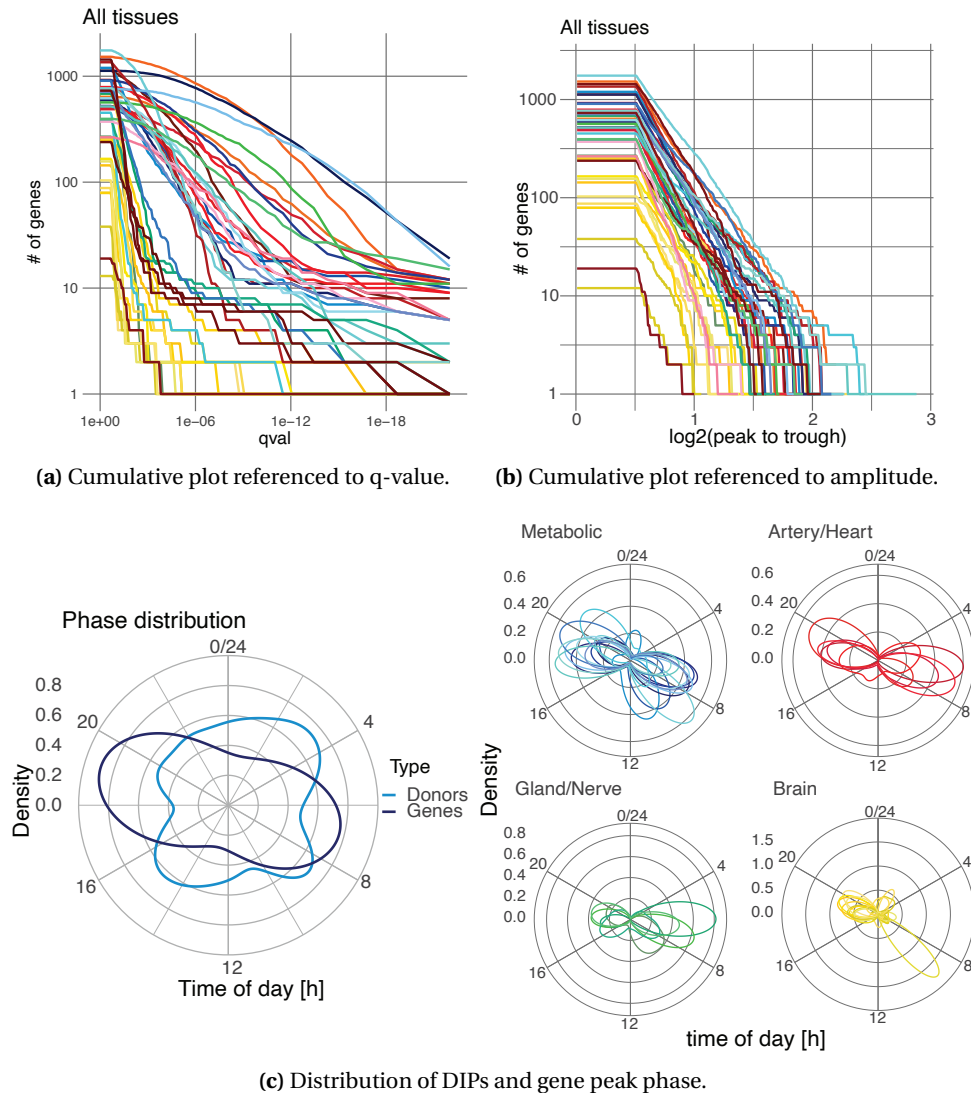


Figure 4.8: Gene rhythmicity across the body. (a) Number of rhythmic genes ($q(BH) > 0.2$ and \log_2 peak-trough > 0.5) with q-value smaller than a threshold (x-axis) across 46 tissues shows numerous and highly significant rhythms in the metabolic tissues. Tissues are color coded as in figure 4.2. (b) Number of rhythmic genes ($q(BH) < 0.2$ and \log_2 peak-trough > 0.5) with peak-trough amplitude higher than a threshold (x-axis, \log_2) across 46 tissues show different intensities of rhythmic gene expression programs across the human body. (c) Left: polar density plot of the predicted DIPs of the 914 donors and of the peak times of all identified rhythmic genes. Right: phase densities of rhythmic genes ($q(BH) < 0.2$ and \log_2 peak-trough > 0.5) for various groups of tissues show biphasic programs, with small rotations.

4.3 Comprehensive 24h gene expression rhythms in humans

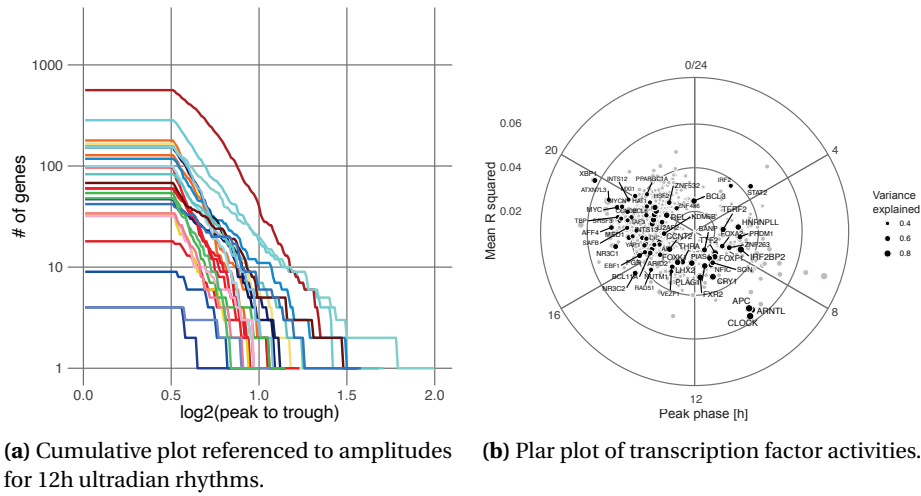


Figure 4.9: Ultradian rhythms and transcription factor activity. (a) Number of 12h-rhythmic genes ($q(BH) < 0.2$ and \log_2 peak-trough > 0.5) with peak-trough amplitude higher than a threshold (x-axis, \log_2) across 46 tissues show different intensities of rhythmic gene expression programs across the human body. (b) Transcription factor activities. Polar plot of mean phase and mean R^2 values for the top 100 targets of all transcription factors (ChIP-Atlas) across the body obtained via cSVD (sections 4.7 and 4.7, Table ST3). These factors are mostly one with pan-rhythmic activity, i.e. the target genes are rhythmic in multiple tissues.

repressed genes [249]. In the evening, MYC and MYCN (7pm, cell proliferation), X-Box-Binding Protein 1 (XBP1) (8pm, response to Unfolded Protein Response), and Peroxisome Proliferator-Activated Receptor Gamma Coactivator 1 (PPARGC1) (8pm, energy metabolism) were activated. During the night, IRF2 (2am, interferon regulatory factor) and STAT2 (3am, cytokine response) activities peaked. Similar TFs showed rhythmic activities in mouse liver [250]. Enriched gene functions showed coherence across many tissues (Table S5, figure 4.10a). Starting at midnight, immune response genes peaked early during the night consistent with the above IRF2 and STAT2 TFs, followed by a response to cholesterol in the early day coinciding with peak times of serum cholesterol levels [251]. Around 9am, we observed a peak for caffeine response, followed by energy homeostasis, gluconeogenesis, and lipid metabolism genes. mRNAs involved in amino acid and glucose metabolism, as well as protein synthesis and folding peaked in the early afternoon, extending into the evening. Cell-cycle pathways peaked in the evening to late night, coinciding with the predicted MYC and MYCN activities. Therefore, pan-rhythmic gene functions in humans largely consist in timed metabolic processes reflecting a switch between low and high energy states during the rest-activity cycle. Among functions that showed tissue specificity, lipid metabolism was particularly rhythmic in the liver,

Chapter 4. Sex-dimorphic and age-dependent organization of 24-hour gene expression rhythms in human

as was amino acid metabolism in the intestine and heat shock response across the brain tissues (figure 4.10b).

4.3 Comprehensive 24h gene expression rhythms in humans

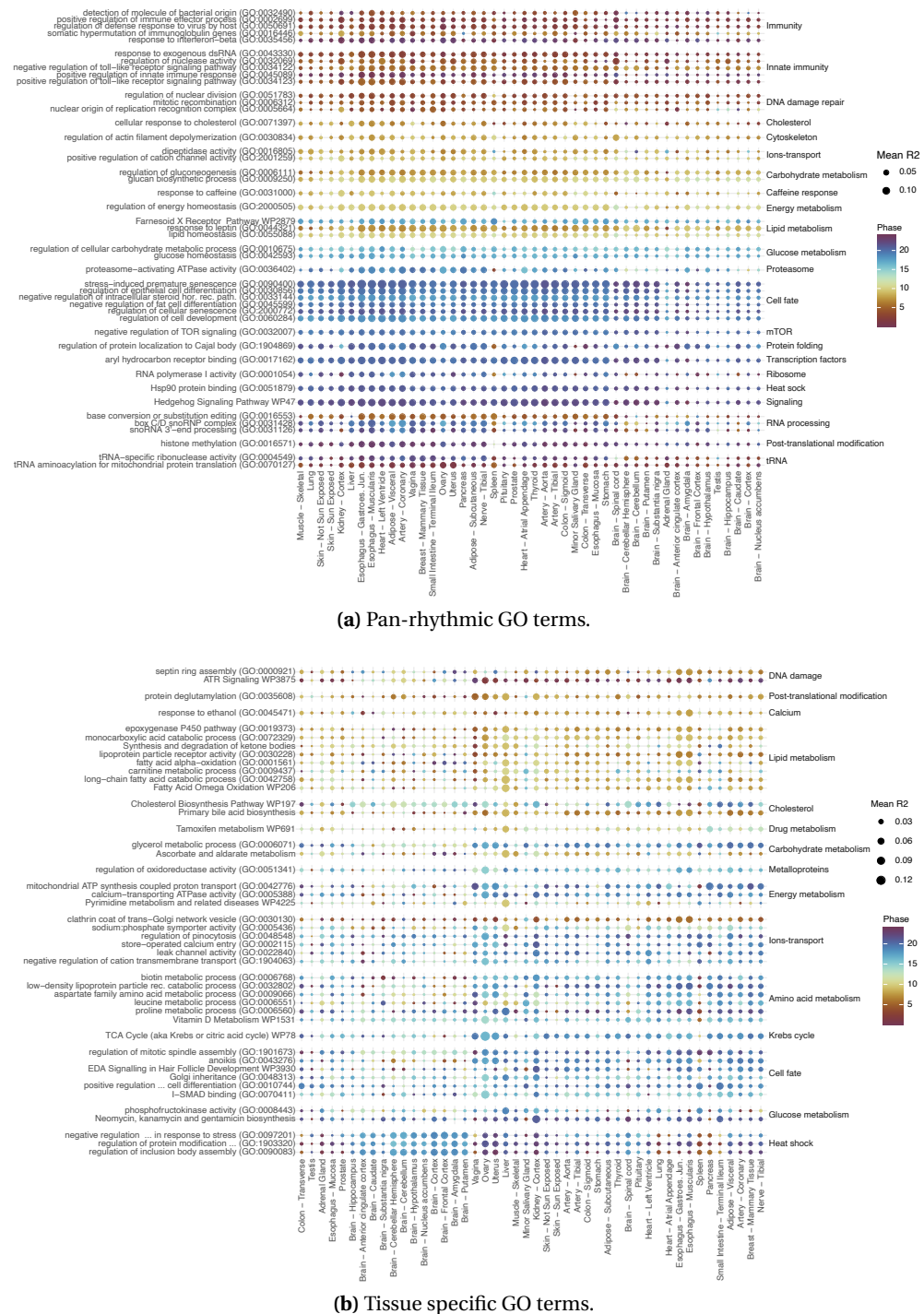


Figure 4.10: Rhythmic GO terms across the body. (a) Set of GO terms showing a coherent rhythmicity across the body in the cSVD approach (section 4.7). Size of the circle is the mean R squared (harmonic regression) of the genes, color is the mean phase. (b) Set of tissue specific rhythmic GO terms highlight enhanced rhythmicity of some function in specific tissues. Representation as in (a).

4.4 Human sexual dimorphism in circadian rhythms

We leveraged DIPs to analyze sex-dimorphic mRNA rhythms. The first cSVD component (93% of variance explained) indicated a conserved clock in males and females (figure 4.11a) and the distributions of DIPs were similar for males and females (figure 4.11b). To analyze sex-dimorphic clock programs, we used a model selection approach to classify each transcript into five statistical scenarios (models) depending on the rhythmic behavior in both sexes, section 4.7[102]. Core clock genes were largely classified as having identical rhythms in both sexes (model 4, Table S3), but the overall number of rhythmic genes was almost double in females at all amplitude thresholds (figure 4.11b).

Although tissues such as esophagus, skeletal muscle and adipose tissue did not differ much, the stratification by sex unveiled several highly dimorphic tissues (figure 4.12). Notably, females had significantly more rhythmic transcripts in the adrenal gland and liver (figure 4.12). Cardiovascular tissues are known sites of circadian regulation [252] exhibiting circadian rhythmicity in GTEx [217]. In the heart (atrial appendage), the total number of rhythmic genes and their peak phases were similar in males and females (figure 4.13). However, only about 50% of rhythmic genes were shared between male and female, with the remaining rhythms being either specific to one sex, or showing different rhythmic patterns (figure 4.15). Liver exhibits marked circadian physiology and sex-dependent gene expression in humans [239]. We found a strong enrichment of mRNA rhythms in females at all amplitudes, mostly as an extensive morning wave (figure 4.14 and 4.16). Three pathways, which have sex-dependent mRNA rhythms in mice [102], showed enriched rhythmicity in female livers: xenobiotic detoxification, fatty acid oxidation, and cholesterol synthesis (figure 4.17). In the latter nearly all enzymes, including the rate limiting and statin target 3-Hydroxy-3-Methylglutaryl-CoA Reductase (HMGCR), showed rhythmicity in females that was damped or absent in males (figure 4.18). In detoxification, many phase I and II enzymes were strongly enriched in female-specific cyclers (figure 4.19) [253]. Female-specific rhythms in the liver were predicted as driven by the Heat Shock Transcription Factor 1 (HSF1) and Peroxisome Proliferator Activated Receptor Gamma (PPARG) (figure 4.17). The adrenal gland also exhibited more rhythmic mRNAs in females than males, centered at midday (figure 4.20). Among those, GR targets were enriched, which could reflect autocrine signaling (figure 4.21). As glucocorticoid signaling is a systemic synchronizer and organizer of peripheral rhythms [244], this might corroborate with the overall increased rhythmicity in females at transcriptional and physiological levels [101].

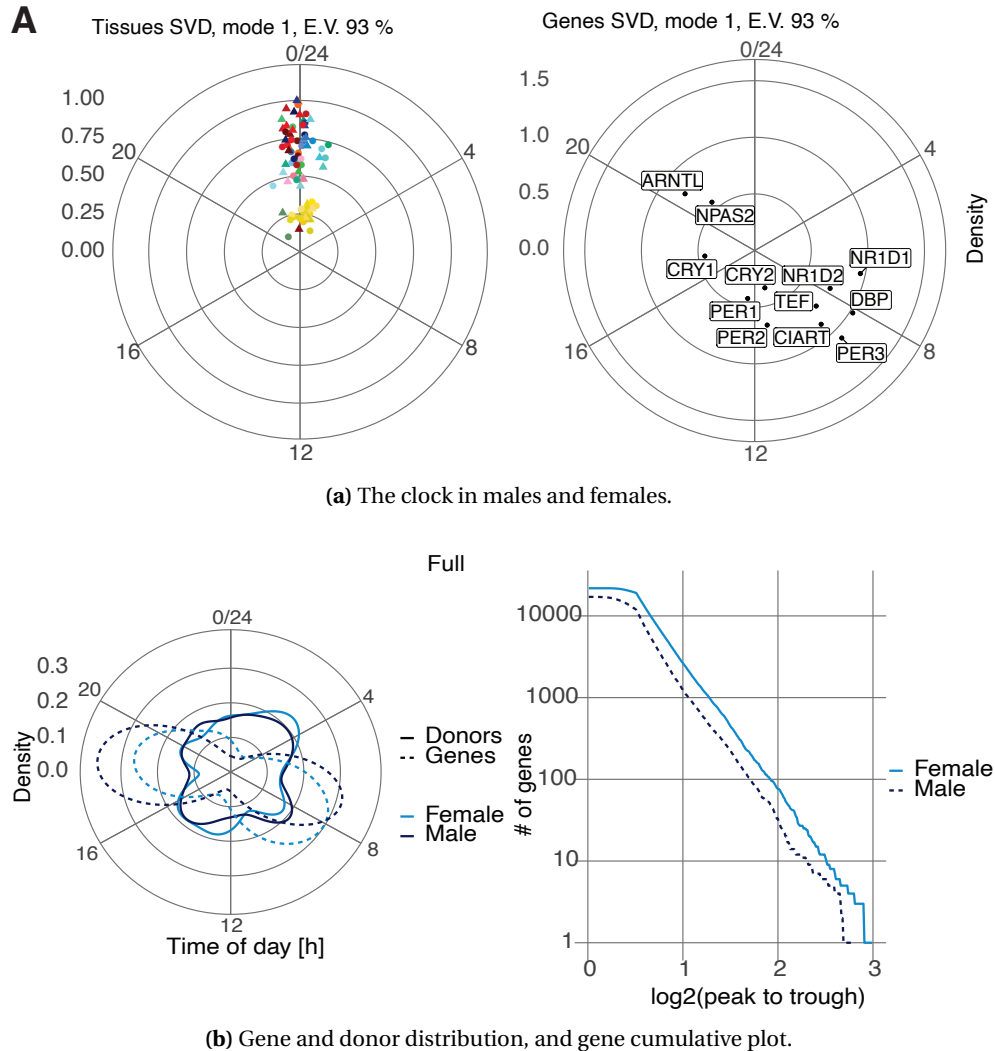


Figure 4.11: Overview of clock, DIP, and rhythms in males and female. (a) First gene (left) and tissue (right) vectors of cSVD performed on CRGs indicate invariant clocks between males (triangle) and females (circle). Males and female samples for one tissue were treated as two tissues. The first module captures 93% of the 24-hour variance, E.V.. (b) (right) Polar densities of DIPs (solid line) and gene peak phases (dashed line) in male (dark blue) and female (light blue) donors show morning and evening waves of gene expression. (left) Number of 24-hour-rhythmic genes with amplitude higher than a threshold plotted as a function of the threshold in all tissues combined show a globally higher rhythmicity in females (solid line) than males (dashed line).

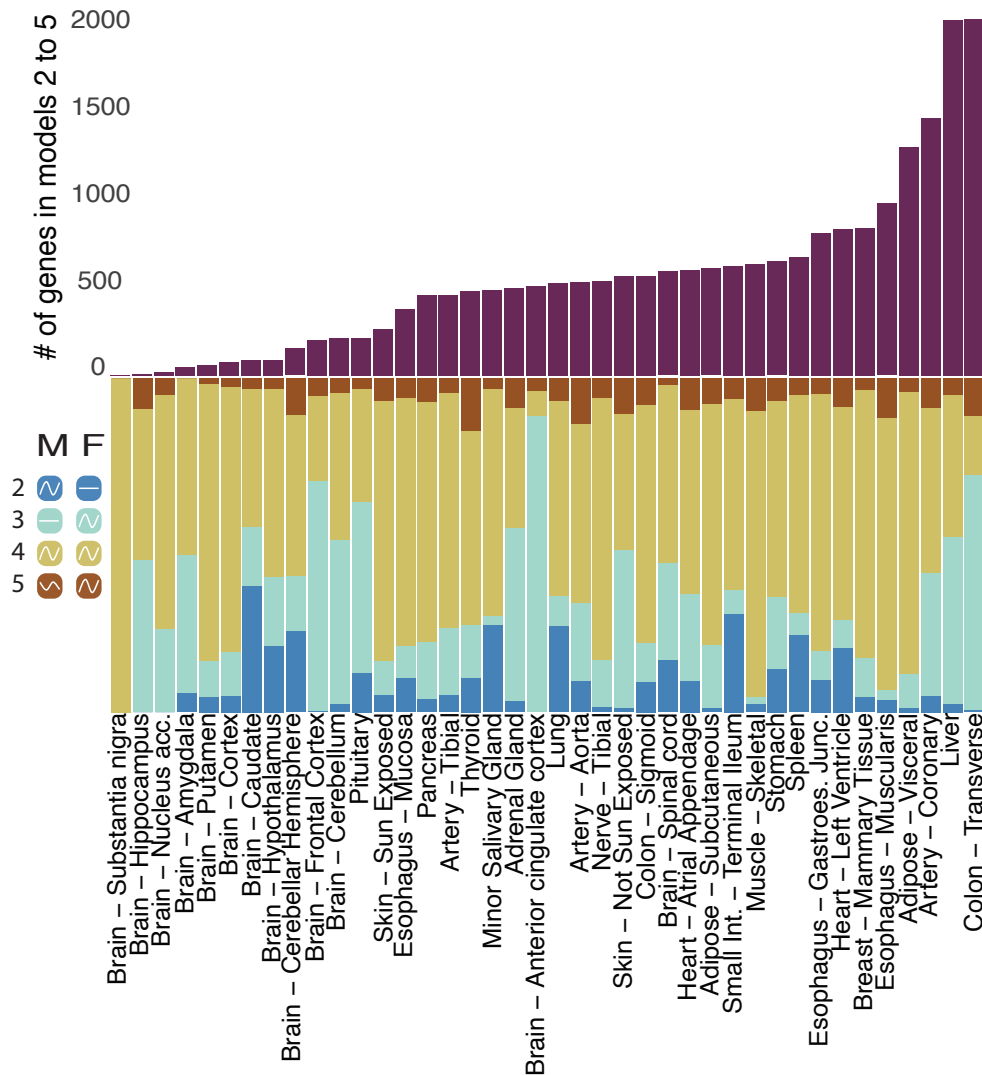


Figure 4.12: **Distribution of rhythmicity between sexes.** Summary of total number of rhythmic genes in each tissue (top, bordeau) divided according to four statistical models (bottom) show tissue-specificity of sex dimorphic mRNA rhyhtms. Models: 2/blue, 3/cyan, 4/mustard, 5/brick

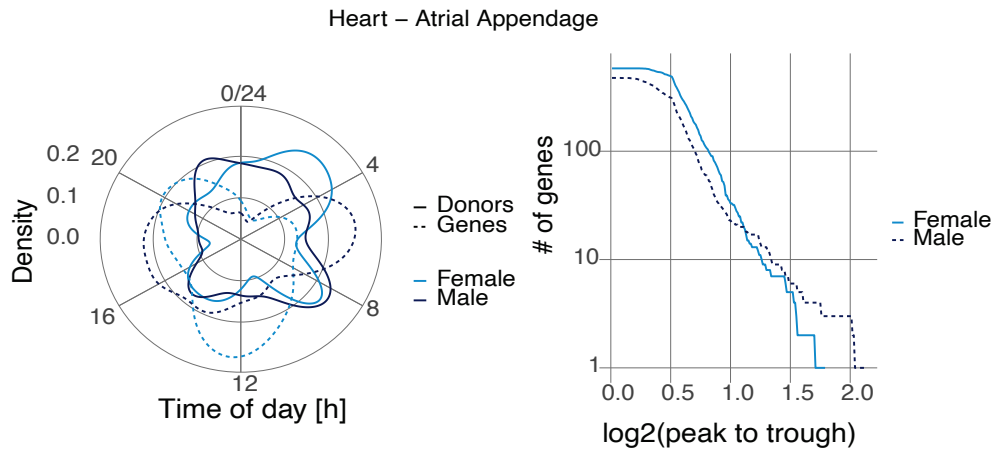


Figure 4.13: **Close up on sex-dimorphism in the heart atrial appendage.** (left) Heart atrial appendage: polar densities of DIPS (solid line) and peak phase of rhythmic genes (dashed line) in for males (dark blue) and females (light blue) show similar overall rhythmicity. (right) Heart atrial appendage: number of genes with amplitude higher than a threshold plotted as a function of the threshold for males (dark blue) and females (light blue) reveals an overall similar amount of rhythmic genes in the two sexes.

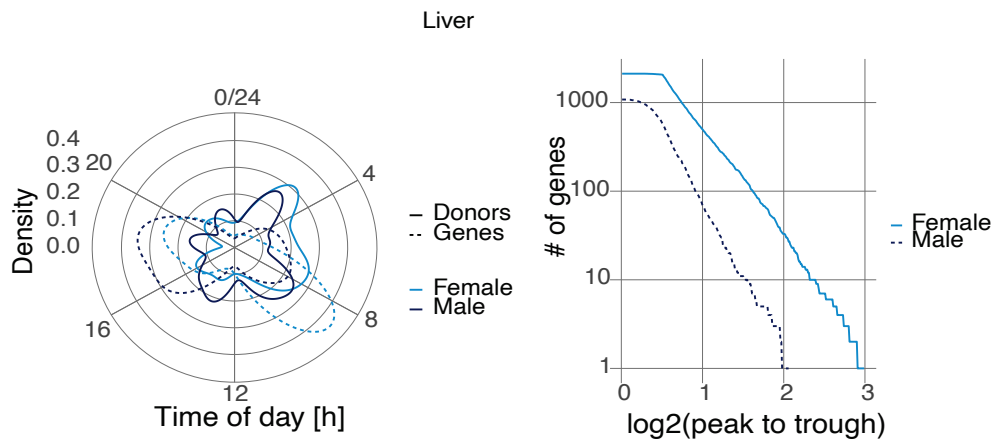


Figure 4.14: **Close up on sex-dimorphism in the liver.** (left) Liver: polar densities of DIPS (solid line) and peak phase of rhythmic genes (dashed line) in for males (dark blue) and females (light blue) show similar overall rhythmicity. (right) Liver: number of genes with amplitude higher than a threshold plotted as a function of the threshold for males (dark blue) and females (light blue) reveals an overall similar amount of rhythmic genes in the two sexes.

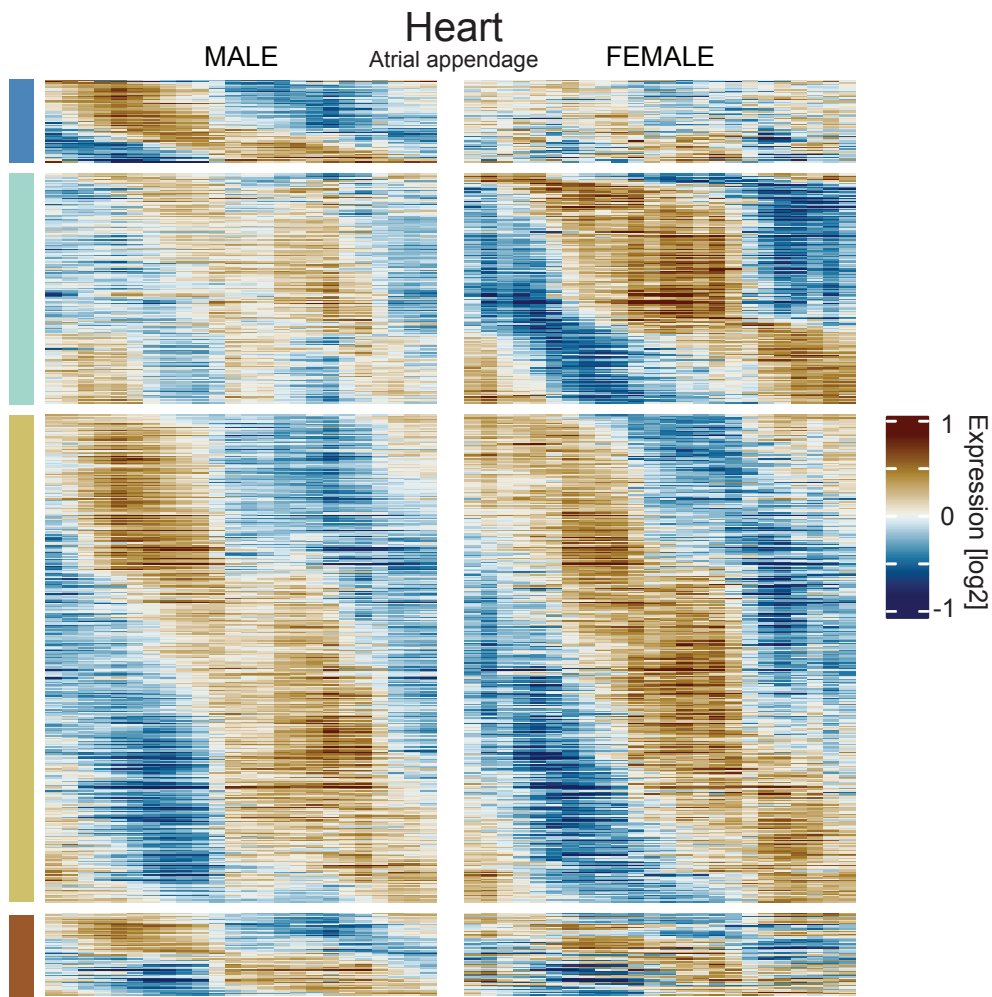


Figure 4.15: **Heatmap of sex-dimorphic rhythms for heart atrial appendage.** Liver: heatmap of genes in models 2 to 5 recapitulates increased rhythmicity of gene expression in females especially in xenobiotic, fatty acid metabolism and cholesterol related pathways (figure 4.17). Log2 mean centered expression, low/blue to high/brown, 1h bins plotted with a 4h window moving average.

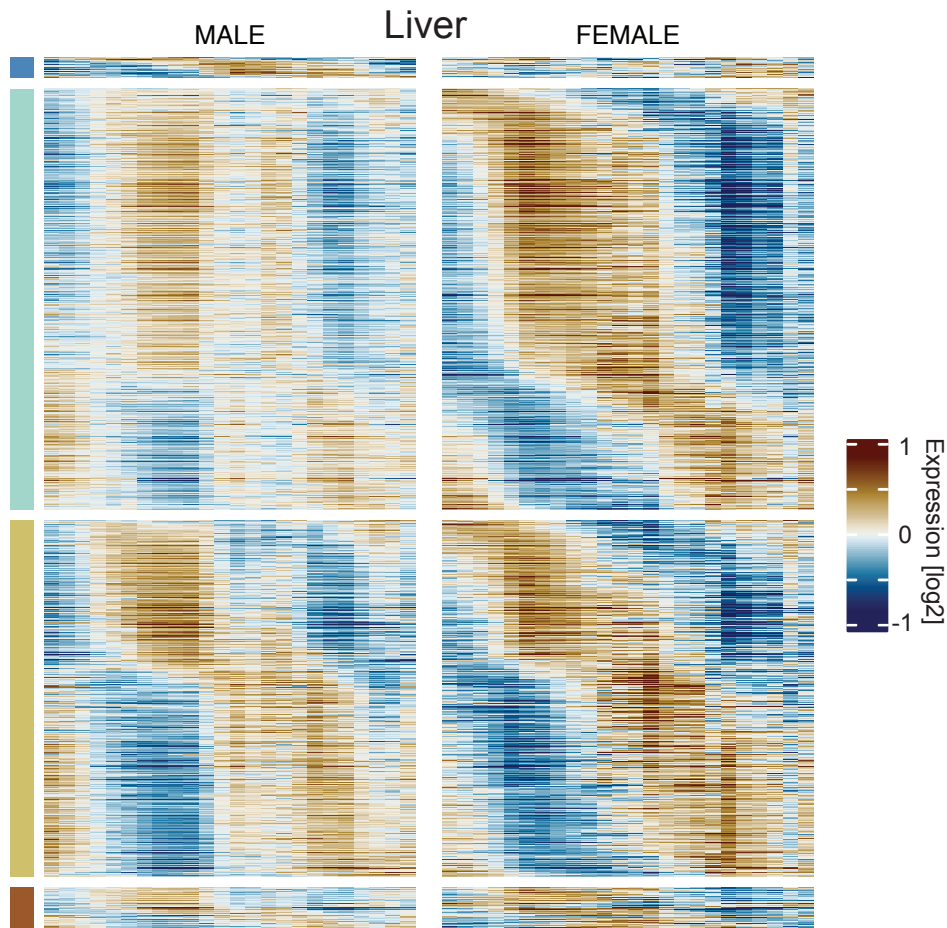
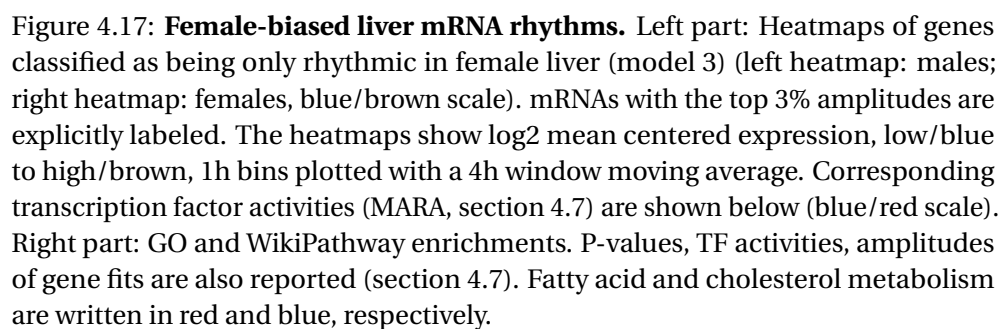


Figure 4.16: **Heatmap of sex-dimorphic rhythms for liver.** Heart atrial appendage: heatmap of mRNA levels for genes in models 2 to 5 illustrates model selection, and shows dimorphic (blue, cyan) and conserved (mustard) rhythmic gene expression. Log2 mean centered expression, low/blue to high/brown, 1h bins plotted with a 4h window moving average.



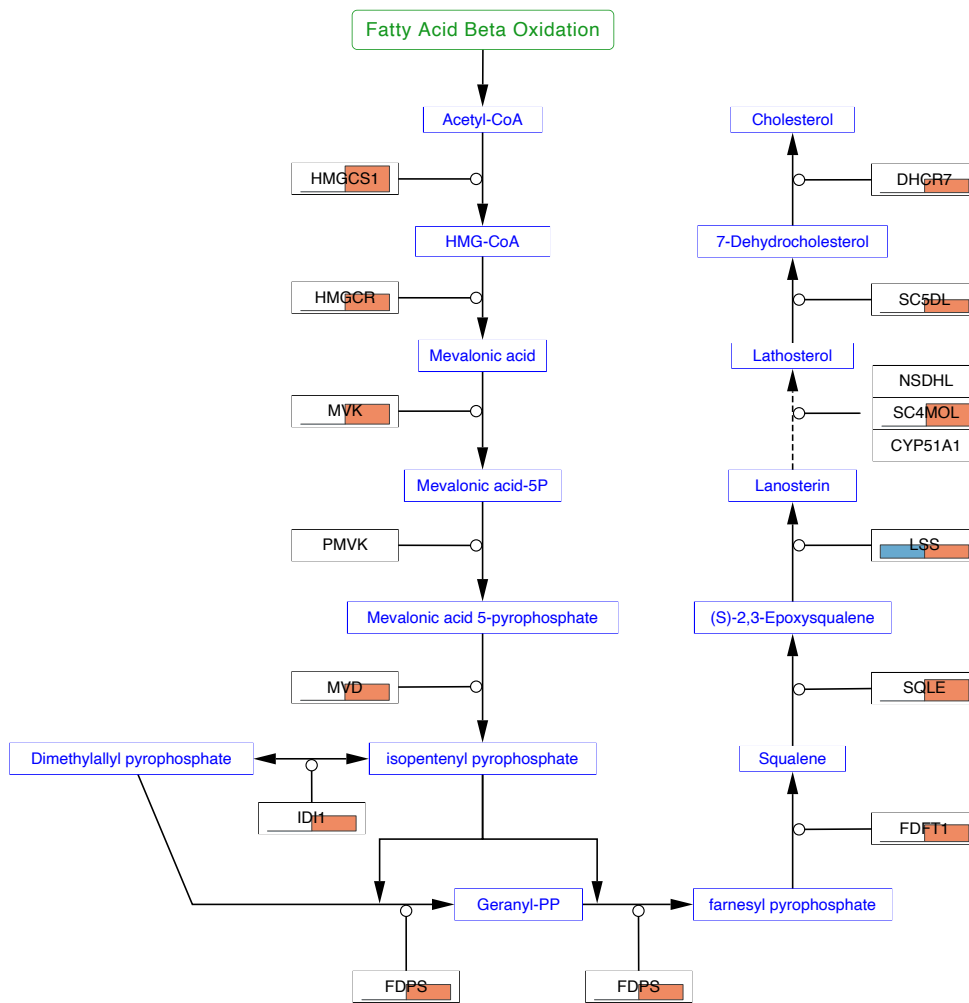
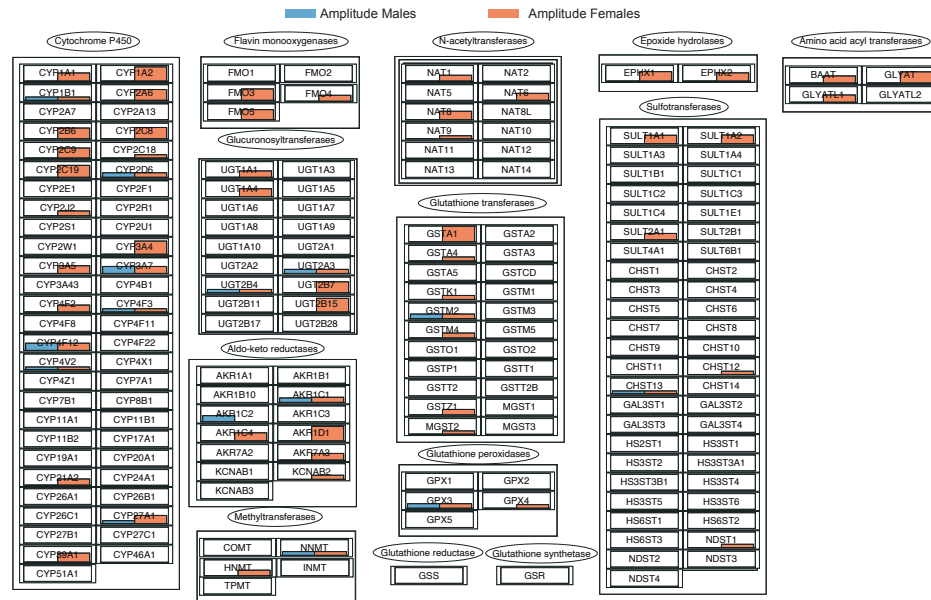
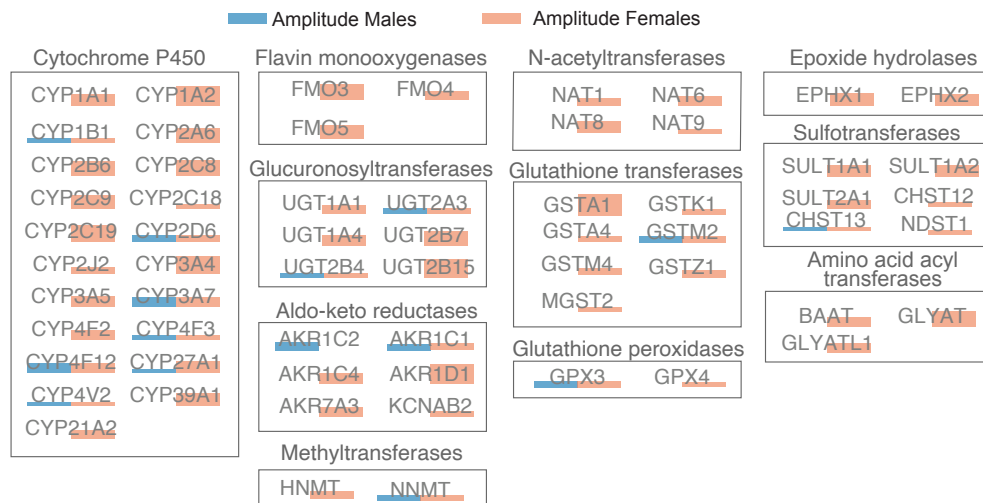


Figure 4.18: **Sex-dimorphism in cholesterol biosynthesis.** Cholesterol biosynthesis pathway (WP197) colored according to the oscillatory amplitudes of genes in males (blue) and females (red) in the liver, shows that all enzymes but one are rhythmic in females.

Chapter 4. Sex-dimorphic and age-dependent organization of 24-hour gene expression rhythms in human



(a) Overview of sex-dimorphism in detoxification



(b) Zoom on sex-dimorphic genes in detoxification

Figure 4.19: **Sex-dimorphism in detoxification.** Liver: (a) visualization of biotransformation meta pathway (WP702) colored according to amplitude of genes in males (blue) and females (red) exemplifies the strong female biased rhythmicity in both phase I and phase II enzymes. (b) Zoom on the rhythmic genes in the biotransformation meta pathway (WP702) colored according to amplitude of genes in males (blue) and females (red).

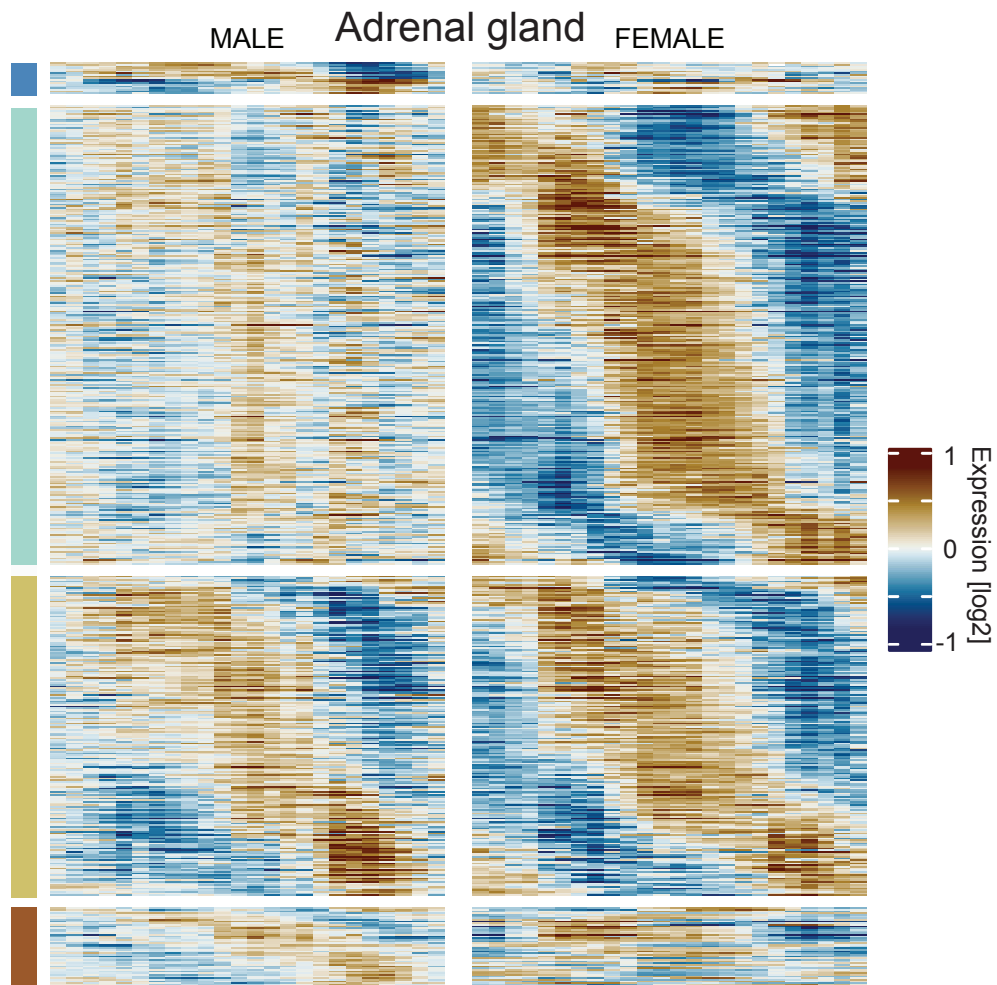


Figure 4.20: **Sex-dimorphism in glucocorticoid signaling.** Genes only rhythmic in the female adrenal gland displayed as in figure 4.17. Note the enrichment of GR signaling genes (blue text).

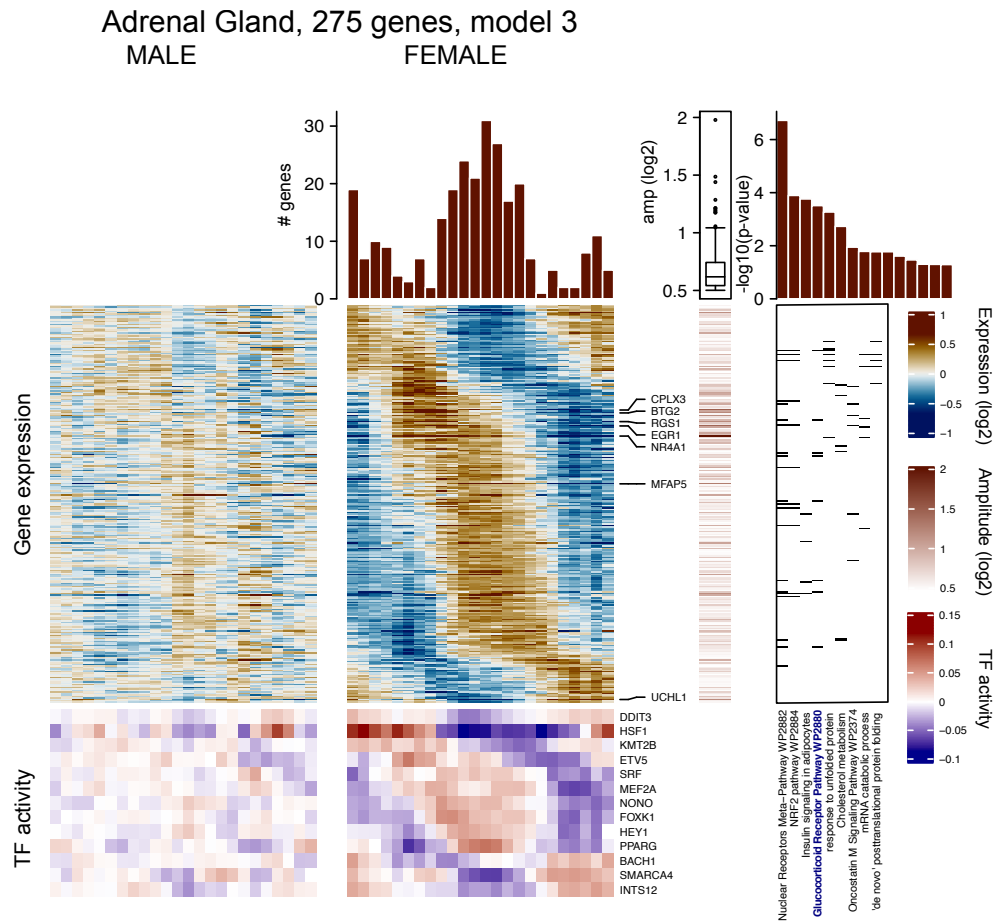


Figure 4.21: **Female-biased mRNA rhythms in the adrenal gland.** Left part: Heatmaps of genes classified as being only rhythmic in female liver (model 3) (left heatmap: males; right heatmap: females, blue/brown scale). mRNAs with the top 3% amplitudes are explicitly labeled. The heatmaps show log2 mean centered expression, low/blue to high/brown, 1h bins plotted with a 4h window moving average. Corresponding transcription factor activities (MARA, section s 4.7) are shown below (blue/red scale). Right part: GO and WikiPathway enrichments. P-values, TF activities, amplitudes of gene fits are also reported (section 4.7). Fatty acid and cholesterol metabolism are written in red and blue, respectively.

4.5 Age dependent circadian reprogramming of human gene expression

We analyzed how aging reprograms daily rhythmic gene expression across the human body. Donors were divided in two age groups: less than fifty (38 ± 9 years), and over sixty (65 ± 3 years). The clock was conserved across age groups (first cSVD component: 91%, figure 4.22a), with core clock genes having mostly identical rhythms (model 4, Table S3). Rhythmic transcripts showed two waves (figure 4.22b), and were overall damped in the older donors (figure 4.22b) [254]. The latter showed one third of rhythmic genes with amplitudes above 4-fold; such loss of rhythmicity with aging occurred in the majority of but not all tissues (figure 4.23). For instance, adipose tissues, esophagus, and skeletal muscle showed conserved rhythmicity across age, with most genes exhibiting statistically identical rhythms in the two groups (model 4, mustard, figure 4.23). This is illustrated in subcutaneous adipose tissue, where the morning and evening waves are pronounced in both younger and older donors (figure 4.24), represented by a majority of conserved mRNA rhythms (figure 4.24 and 4.26). We next focused on coronary arteries, a tissue that strongly loses rhythms with age. Although morning and evening transcript waves were observed in both groups (figure 4.25), the number of rhythmic mRNAs in older donors was about half that in younger donors, across all amplitudes (figure 4.25, figure 4.27). Programs that lost rhythmicity include cholesterol biosynthesis, FA synthesis, and the regulation of glycolysis (figure 4.28 and 4.29), processes known to be deregulated in vascular smooth muscle cells in cardiovascular diseases [255]. Most enzymes in the cholesterol biosynthesis pathway, including HMGCR, were rhythmically expressed in young coronary arteries but lost this feature with age (figure 4.28). Comparing ovaries between pre- and post-menopausal women revealed both lost and gained mRNA rhythms (figure 4.30). Although rhythmicity in lipid and cholesterol biosynthesis was suppressed in older donors, as in the coronary arteries (figure 4.31), stress, and in particular heat shock response genes, became rhythmic, as supported by predicted HSF1 transcription factor activity (figure 4.31 and 4.32). This signature of a thermal stress response in post-menopausal women may reflect circadian patterns of temperature control [256]. In some tissues, genes switched from a 24-hour to a 12-hour ultradian periodicity with age. In the pituitary gland, the liver, and the colon, 12-hour rhythms arose in 30 to 50% of genes classified as only rhythmic in young (figure 4.33). In humans, these tissues regulate rhythms of temperature, energy metabolism and absorption. Such destabilization of 24-hour periodicity in favor of an ultradian state as a result of aging might reflect differences

Chapter 4. Sex-dimorphic and age-dependent organization of 24-hour gene expression rhythms in human

in reception of external cues in older individuals [257].

4.5 Age dependent circadian reprogramming of human gene expression

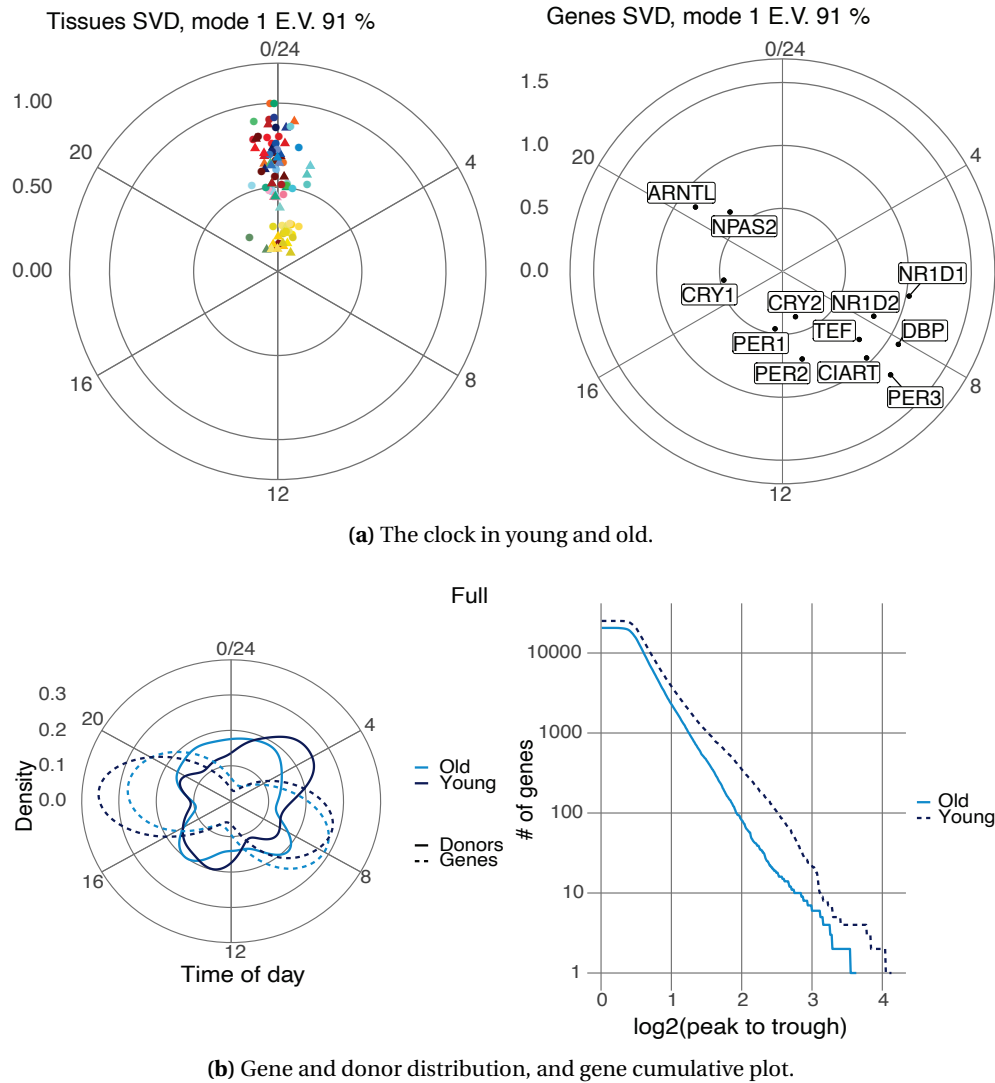


Figure 4.22: **Overview of clock, DIP, and rhythms in young and old.** (a) First gene (left) and tissue (right) vectors of cSVD performed on CRGs indicate invariant clocks between younger (triangle) and older (circle). Young and old samples for one tissue were treated as two tissues. The first module captures 91% of the 24-hour variance, E.V.. (b) (right) Polar densities of DIPs (solid line) and gene peak phases (dashed line) in all tissues combined exhibit substantially stronger 24-hour mRNA rhythmicity in younger (dark blue) compared to older (light blue), especially for higher amplitudes. (left) Number of 24-hour-rhythmic genes with amplitude higher than a threshold plotted as a function of the threshold in all tissues combined show a globally higher rhythmicity in females (solid line) than males (dashed line).

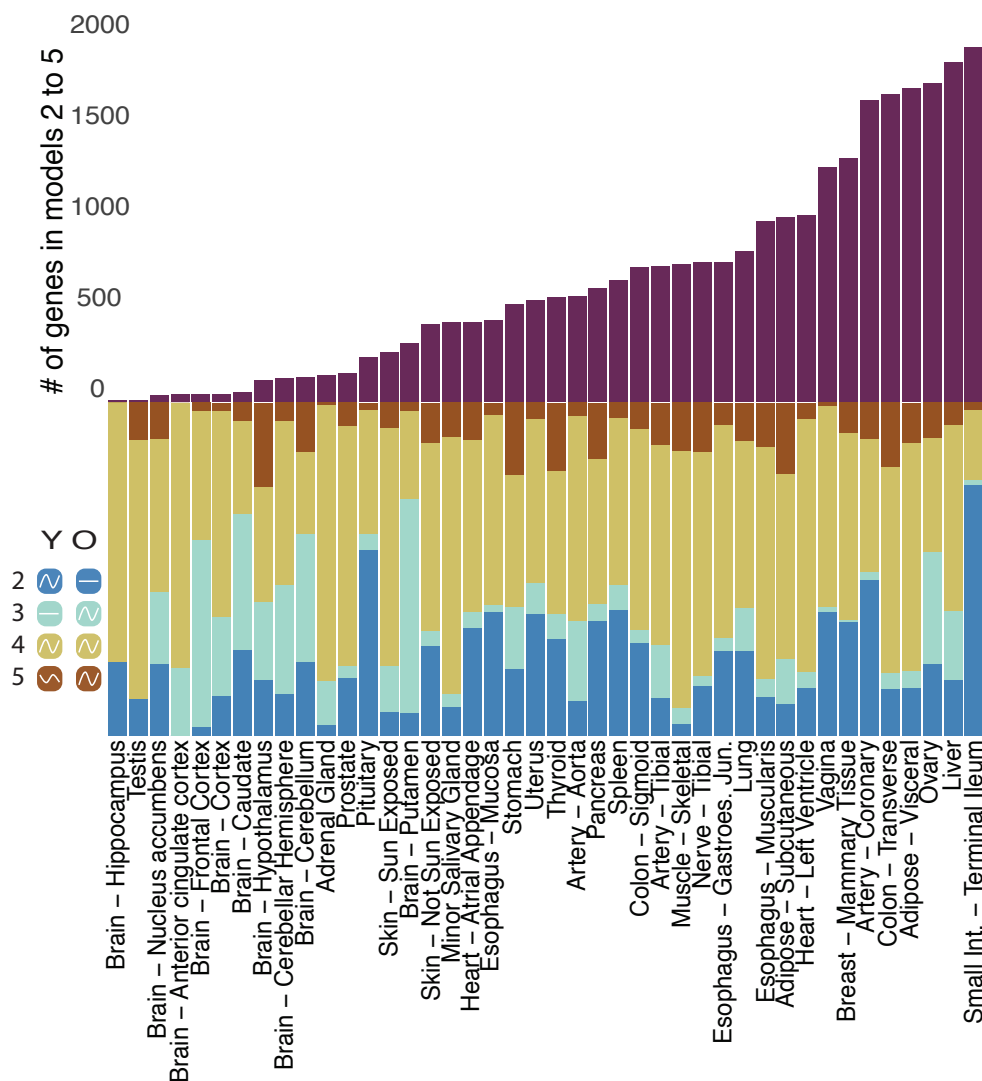


Figure 4.23: **Distribution of rhythmicity between ages.** Summary of the number of rhythmic genes in each tissue divided according to the model selection approach recapitulates the global behavior of tissues (metabolic tissues are more rhythmic, brain less) and the general trend of loss of rhythmicity as a result of aging. Models: 2/blue, 3/cyan, 4/mustard, 5/brick. Models: 2/blue, 3/cyan, 4/mustard, 5/brick

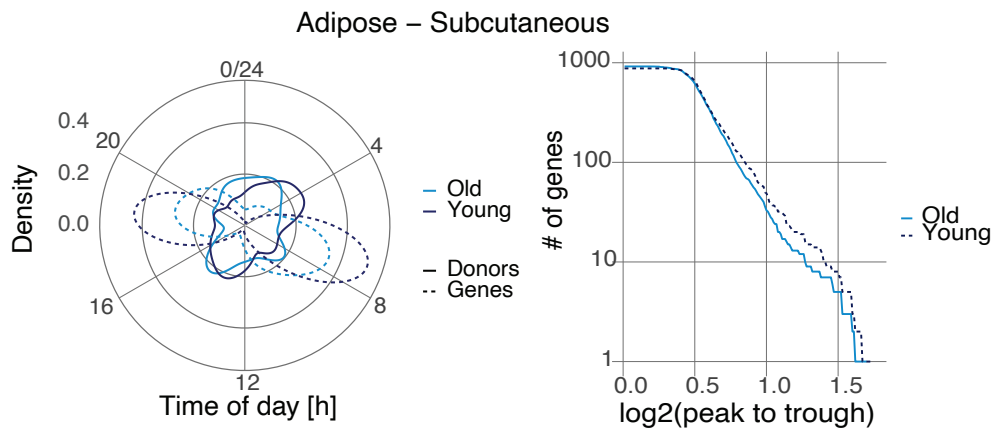


Figure 4.24: **Close up on age-dependent rhythms in adipose subcutaneous.** (left) Adipose subcutaneous: polar densities of DIPs (solid line) and peak phase of rhythmic genes (dashed line) in adipose subcutaneous for younger (dark blue) and older (light blue) show a decay in the biphasic rhythmic programming as age progresses. (right) Adipose subcutaneous: number of genes with amplitude higher than a threshold for young (dark blue) and old (light blue) shows highly conserved rhythmic programs with only a mild dampening of high amplitude rhythms in elderly.

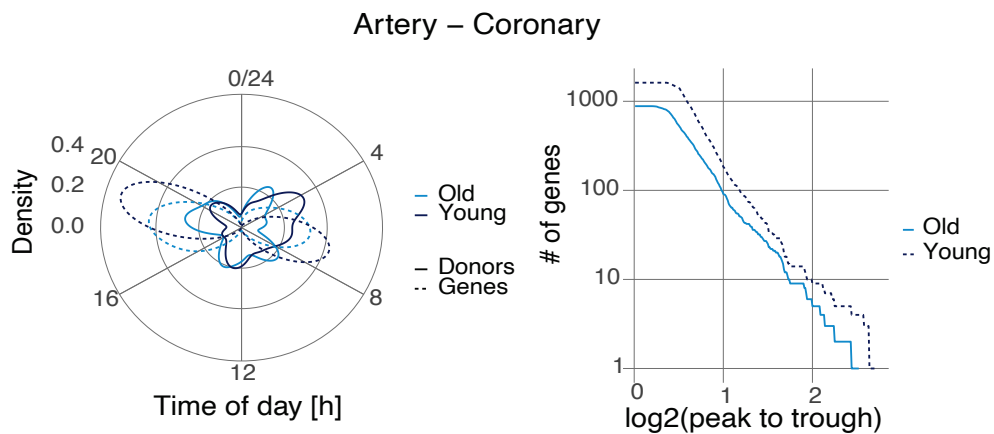


Figure 4.25: **Close up on age-dependent rhythms in the coronary arteries.** (left) Coronary arteries: polar density plot of DIPs (solid line) and peak phase of rhythmic genes (dashed line) for young (dark blue) and old (light blue) shows morning and evening waves of gene expression that decay with age. (right) Coronary artery: number of genes with amplitude higher than a threshold plotted as a function of the threshold for young (dark blue) and old (light blue) emphasizes more abundant rhythmic programming in younger donors across all amplitudes.

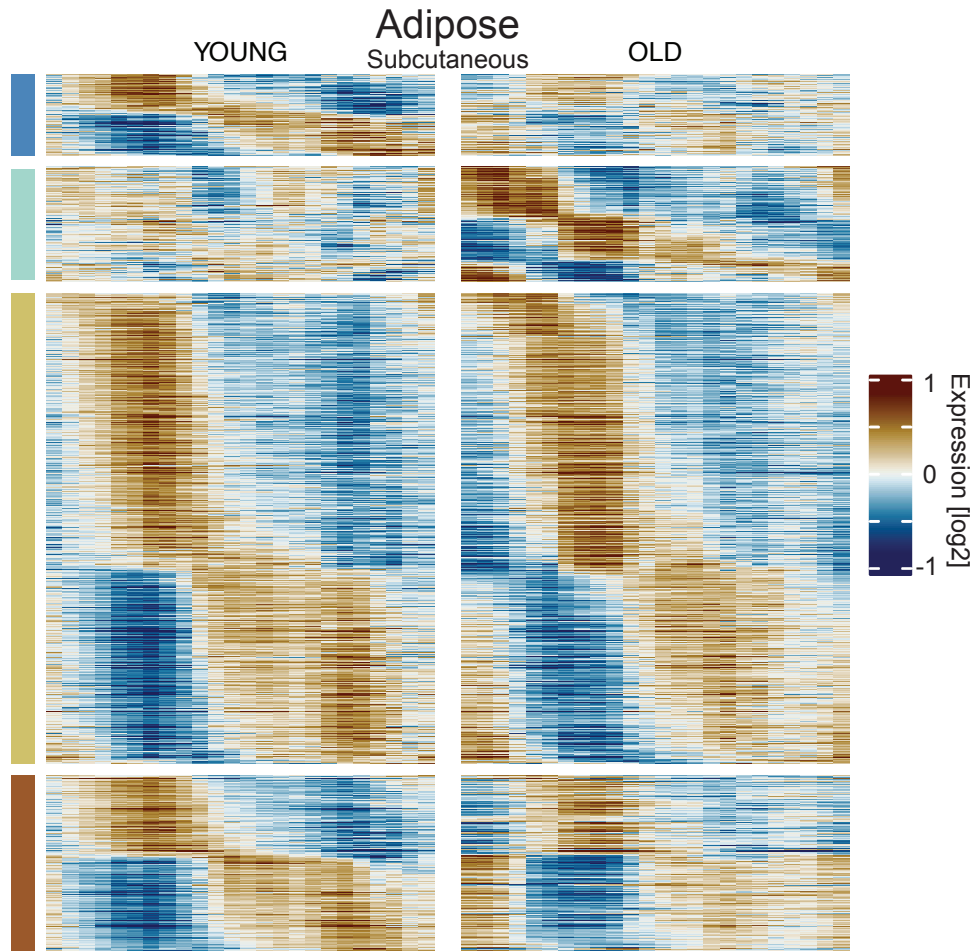


Figure 4.26: **Heatmap of age-dependent rhythms for adipose subcutaneous.** Adipose subcutaneous: Heatmap of genes in models 2 to 5 highlights the mainly conserved rhythmic programming (mustard) as well fewer losses (blue) and gains (cyan) with aging. Log2 mean centered expression, low/blue to high/brown, 1h bins plotted with a 4h window moving average.

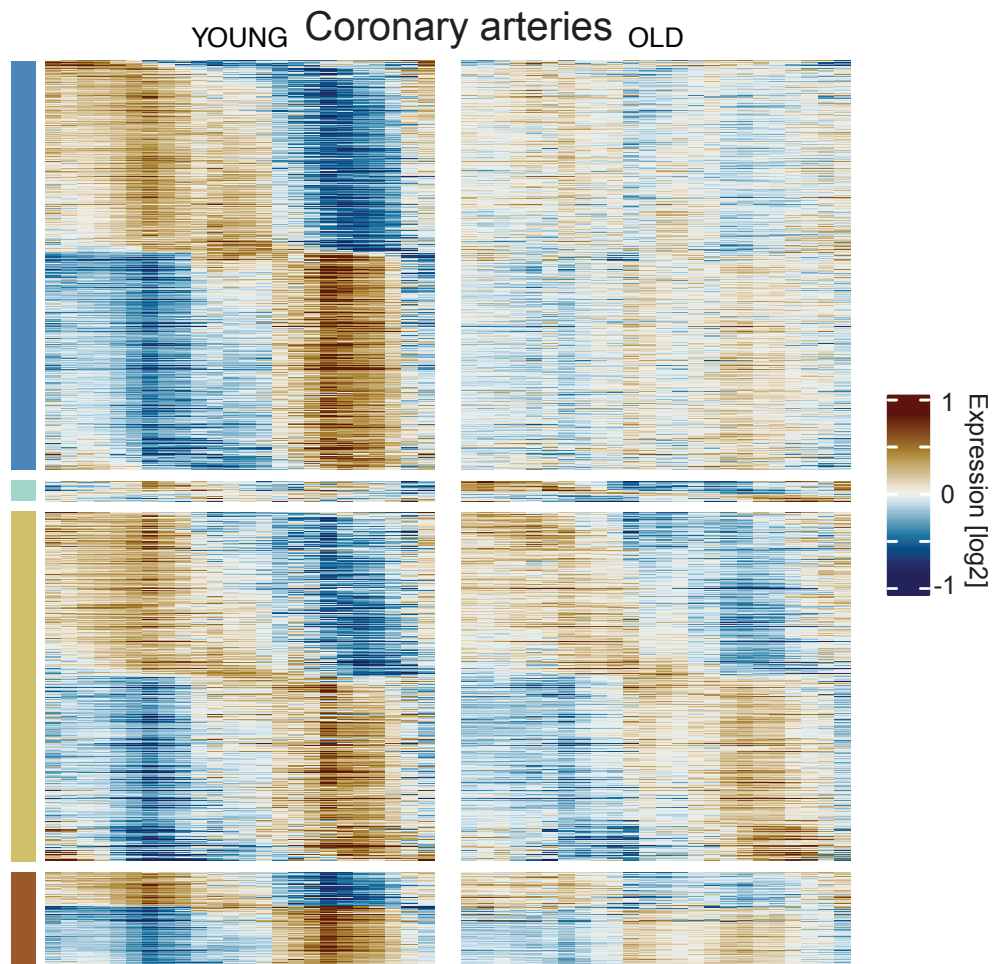


Figure 4.27: **Heatmap of age-dependent rhythms for coronary arteries.** Coronary arteries: Heatmap of genes in models 2 to 5 illustrates strong morning and evening waves of gene expression and exemplifies instances of loss of functional rhythmic programs as a result of ageing. Log2 mean centered expression, low/blue to high-/brown, 1h bins plotted with a 4h window moving average.

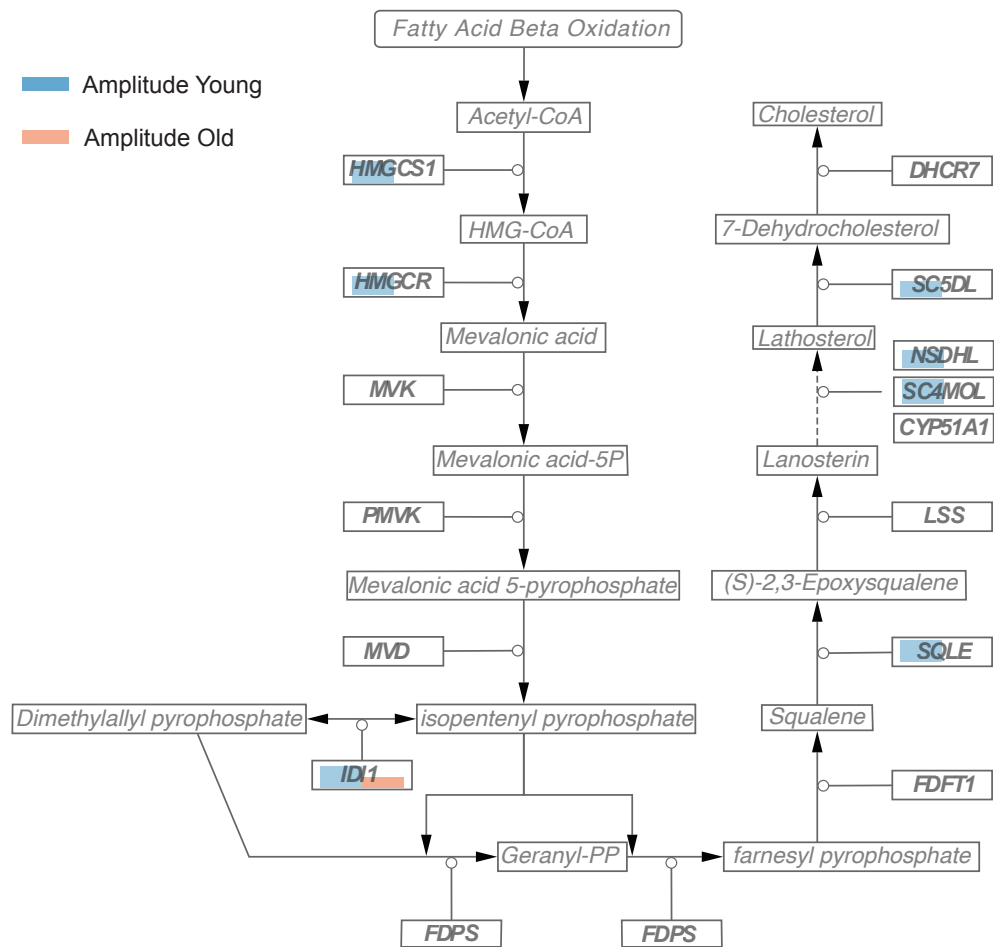
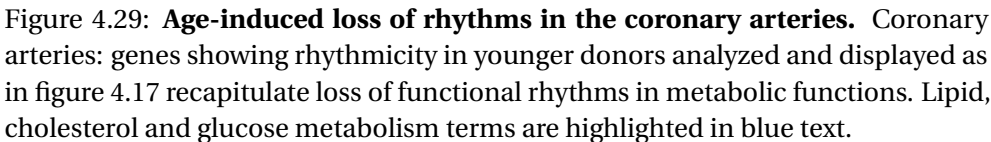


Figure 4.28: **Age-dependent cholesterol biosynthesis.** Artery coronary: visualization of cholesterol biosynthesis pathway (WP197) colored according to amplitude of genes in young (blue) and old (red).



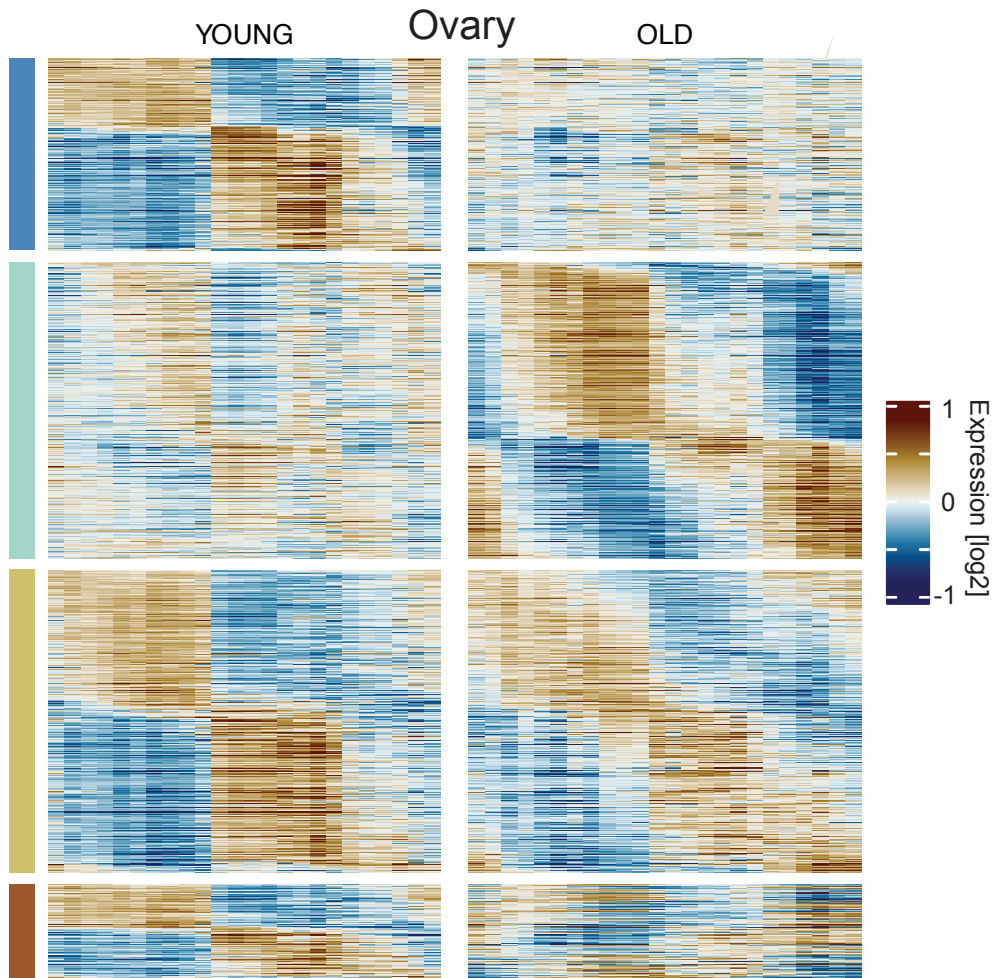
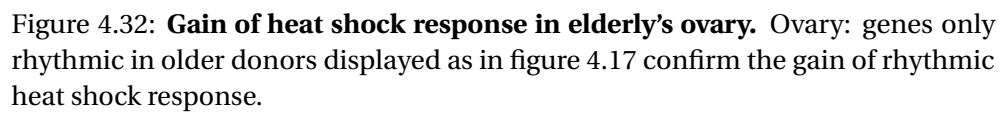


Figure 4.30: **Age-related change of physiology in the ovary.** Ovary: heatmap of genes in models 2 to 5 depicts differential rhythmic programming of gene expression as a function of age.



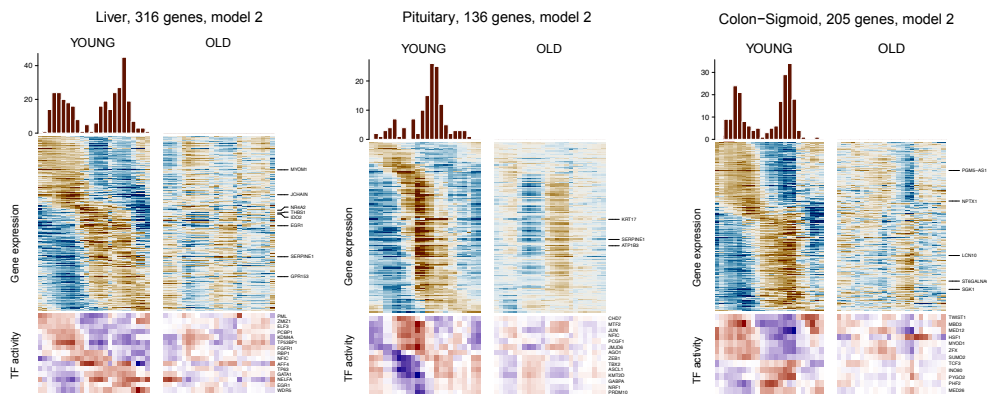


Figure 4.33: **Emergence of ultradian rhythms with age.** Genes only rhythmic in young liver, pituitary and colon displayed as in figure 4.17. Note that high amplitude 24-hour patterns in the younger group showed the emergence of 12h rhythms in older donors.

4.6 Discussion

We developed an algorithm to temporally order GTEx samples that could overcome several limitations of postmortem data through pre-processing, controlled statistics and formulation in terms of a population-level phase model (DIP). Nevertheless, sensitivity to data quality, complex covariate structures or sampling bias cannot be fully ruled out. Clocks were largely in phase in 46 analyzed tissues, with the adrenal gland peaking earliest. The concomitant signature of a sizable wave of negative GR targets in the afternoon suggests that released adrenal glucocorticoids play a crucial role in human body-wide circadian synchronization, including overall increased rhythmicity in females. HSF1 targets contributed a significant portion of the body's 24-hour rhythms and showed sex- and age-dependency. The observation that rhythmic liver transcript levels, particularly in xenobiotic detoxification, was prevalent in females may reflect sex-dimorphic incidence of liver diseases [258]. Similarly, the loss of mRNA rhythms with age in coronary arteries correlates with age-dependent incidence rates of cardiovascular diseases [259]. The identified differences in 24-hour rhythmic processes across sexes and ages may help improve patient-specific chronopharmacology [260].

4.7 Methodological details

This section has been adapted as little as possible from the preprint of the supplementary material of preprint of [219] written by Lorenzo Talamanca, Cedric Gobet and Felix Naef. The research was developed along the way by interactions of the three; the original research question envisioned by Felix Naef has been addressed in chapter 3. This work in its entirety had been supervised and checked by Felix Naef; the analysis, coding, writing and illustrations were made by Lorenzo Talamanca and Cedric Gobet in equal parts.

RNA-seq data preprocessing

Gene mRNA read count data from GTEx V8 (2017-06-05_v8_RNASeQCv1.1.9) were downloaded. For each tissue, genes with less than 10 counts on average across the samples were discarded. Counts were normalized by library size and scaled using the TMM method from edgeR [261], and then converted to counts per million (CPM). Normalized counts were then log-transformed (\log_2) after addition of a pseudo-count of 0.25 (called hereafter LCPM). We proceeded in two filtering steps: the first that is a more stringent one which is used to assign the DIPs to all individuals; once these are fixed, a second step where we add back more samples and genes to increase coverage. In the first step (stringent selection) of the analysis, to control for sample quality, we selected samples with RIN numbers larger than 6, total number of reads larger than 40 millions, proportion of uniquely mapped read larger than 0.8, a positive ischemic time and autolysis score of 1 or 0; genes with on average less than 3 LCPM per tissue were discarded. Tissues with less than 48 samples were discarded. The initial filterings led to 35 tissues comprising 10646 samples (set-1) that were used for the TIP/DIP assignments. In the second filtering step, once the DIPs had been assigned, the criteria for sample inclusion were relaxed and we kept all samples with RIN numbers larger than 4 and whose donor had at least one sample in set-1, we also included genes with more than 0 LCPM. Tissues with less than 48 samples were discarded once again. This allowed us to include additional samples and genes covering in total 15'745 samples across 46 tissues (set-2). Using principal component analysis (PCA) for data exploration, we noticed that several covariates available as metadata explained significant portions of the variance (Fig. S1D). Therefore, we regressed out ischemic time, sex, age and type of death as categorical variables with a linear regression model on the log-transformed normalized count data (LCPM) [262]. The regression was done on each gene in each tissue independently in both set-1 and set-2. The residuals of

these fits were then used for all subsequent analyses. Outlier data points distant more than thirty two-fold ($5 \log_2$ units) from the mean were not considered.

Estimation of TIPs with CHIRAL

First, we assigned a TIP for each sample in each tissue using CHIRAL (details in chapter 3). For the purpose of inferring the Tissue Internal Phases, we modeled the LCPM matrix (E_{gc}) in each tissue using a multivariate harmonic regression model $E_{gc} = \mu_g + a_g \cos(\varphi_c) + b_g \sin(\varphi_c) + \varepsilon$ for condition c and gene g with Gaussian noise ε . If a gene did not carry time information in that tissue, the corresponding gene Fourier parameters (a_g, b_g) can be very small or even zero. We treated the model probabilistically with prior Gaussian distributions over the gene parameters and used a Bayesian calculation to marginalize over the gene parameters distribution and derived a posterior distribution over the unknown phases φ_c . We then inferred maximum a posteriori estimators using an expectation maximization (EM) algorithm that had a similar structure as probabilistic PCA [235]. Our EM procedure enforced one additional constraint to ensure that each 2D cosine and sine vector has norm 1, which we imposed exactly using Lagrange multipliers. We considered a two state mixture model in the EM to account for genes that do not carry temporal information (flat genes): the two state model weighs each gene contribution to the phase estimate according to the posterior probability of that gene carrying information. The initial phases were seeded using a mean field approximation, which yielded a so-called XY spin glass [84], and which was solved using an iterative scheme (chapter 3). Although CHIRAL was constructed to take as input an arbitrary number of genes, even the full genome, the different sources of variance in addition to the internal phase external time renders this approach ineffective in human data. Thus, we applied the algorithm only to a subset of genes which are good candidates for carrying temporal information. We picked the 12 Clock Reference Genes (CRGs) from the literature (ARNTL, CIART, CRY1, CRY2, DBP, NR1D1, NR1D2, NPAS2, PER1, PER2, PER3, TEF), and standardized the measurements (LCPM) for each gene. Applying CHIRAL on this gene subset yields very good reconstruction of circadian phases in human time-series of muscle biopsies [236] (Fig. S1B). In this task, CHIRAL outperformed CYCLOPS (11), the most widely used unsupervised method for circadian phase inference (Table S1). The phase ordering from CHIRAL (but also any other unsupervised method) has two intrinsic symmetries: a time direction symmetry and a rotational symmetry. We broke the first by considering the known temporal order of clock genes. For the rotation symmetry we referenced the TIPs to the peak of PER3 mRNA. We broke the invariances in a tissue specific

way.

Inference of DIPs from TIPs

Due to the nature of the GTEx data, the TIPs for one donor were not independent but had additional structure. At the level of TIPs we had a set of internal phases $\{\varphi_{td}\}$ that depended on the donor (d) and tissue (t). Our main assumption consisted in expressing the TIPs as a sum of two contributions: $\varphi_{t,d} = \varphi_d + \delta\varphi_t$, where φ_d was the donor specific phase (DIP) and $\delta\varphi_t$ was the tissue specific shift which was assumed to be common to all donors (as a convention we used skeletal-muscle as the reference, i.e. $\delta\varphi_{t=\text{skeletal muscle}} = 0$). In essence, this assumption considers a population level model for the tissue specific shifts, and neglects individual deviations with respect to the mean behavior of the population. Moreover, the number of parameters is reduced from $|d| \times |t|$ to $|d| + |t|$. To compute the DIPs from the TIPs we then applied a heuristic approach based on the probabilistic model and the following observations (using the samples in set-1). The distributions of t,d across tissues for one donor were typically fairly peaked (Fig. S1G-H), indicating that the majority of tissue harbored very similarly phased clocks. Outliers in those distributions tended to be caused by an inherent instability of phase ordering algorithms, linked to the collinearity of clock transcripts (most transcripts peak in the ARNTL or almost exactly opposite PER3 phase), and which was particularly present in tissues with weak or confounded clocks. We thus defined the robustly estimated mode of the distribution of TIPs for each individual as the DIP. Specifically, for each donor, after identifying the primary mode we computed the circular mean of samples within a 2h interval of the model. In this approach tissues with weak clocks, or tissues which contain significant sources of non temporal gene expression variability did not bias the donor phase estimates. Moreover, unlike time of death (TODs), DIPs did not need to be corrected for chronotype as they directly measured the internal clock phase. The DIP assignment based on samples in set-1 was directly applied to set-2 since the donors of samples in set-1 and set-2 were the same. DIPs allowed us to expand the analysis to tissues and samples of lower data quality from which a robust circadian phase could not have been inferred alone. Overall, this allowed us to study the daily oscillations in 46 tissues and 16k bulk RNA-seq samples from GTEx (set-2). In other words, the DIPs allow us to transfer clock phase information from tissues with strong clock signatures to ones with weaker weaker or biased signatures. Because only one global reference was needed (PER3 peak expression in muscle was fixed to 9am as in labeled muscle data (12)), all gene and tissue phases could be compared across tissues.

Harmonic regression

Once we assigned the DIPs d to all donors, we assessed rhythmicity and inferred gene coefficients in each tissues using multivariate harmonic regression for gene g in donor d for all the LCPM (Egd) of set-2: $E_{gc} = \mu_g + a_g \cos(\varphi_c) + b_g \sin(\varphi_c) + \varepsilon$. Rhythmicity was assessed using a likelihood ratio test between a rhythmic and a flat ($a_g = b_g = 0$) model. P-values were computed from a χ -squared distribution. Adjusted p-values were computed with the Benjamini-Hochberg (BH) correction. The complex number representation of a gene used in the cSVD is $a_g + ib_g$.

Clock gene correlation matrix

We computed a clock correlation matrix following [84]. For tissue X , the clock correlation matrix is the matrix of correlation of the CRGs across samples of tissue X . We then performed a weighted average of the correlation matrices across tissues using as weights the number of samples in each tissue shown in figure 4.1c. The distance from a reference clock correlation matrix is a proposed metric for clock integrity in tissues without the need of time stamped data.

Model selection

To study differential mRNA rhythms after stratifying the samples into two groups (male vs. female and young vs. old), we first selected the union of possible rhythmic genes, namely those which had a q-value (Benjamini Hochberg, BH) < 0.2 and a \log_2 peak-to-trough > 0.5 when performing harmonic regression with either all samples, or only samples in one of the two conditions. Then, on this set of genes, we used model selection similarly to [43, 102] using the Akaike Information Criterion (AIC). This allowed us to classify genes into 5 different models. Model 1 (never shown and highly infrequent due to the selection) represented genes flat in both conditions, male-female or young-old. Model 2 (blue) represented genes only rhythmic in the first condition, male/young, and model 3 (cyan) those in the second condition, female/old. Model 4 (mustard) is composed of genes rhythmic in both conditions with the same parameters (phase and amplitude) while in model 5 (brick), genes are rhythmic in both conditions with different parameters. As the number of samples in the two compared groups might differ, we subsampled the group with more samples to match the sample size of the smaller group, therefore avoiding sample size biases.

Complex-valued Singular Value Decomposition (cSVD)

We used the complex-valued singular value decomposition (cSVD) to decompose and represent rhythmic gene expression across multiple tissues as in [19]. In particular, gene complex Fourier component for the 24-hour oscillation are fitted tissue-by-tissue. Then we apply the cSVD factorization to the matrix of gene complex Fourier components across tissues for relevant sets of genes. The algebra of complex numbers and linear mathematics of cSVD is optimally suited to decompose oscillatory gene expression matrices into rank 1 modules consisting of complex-valued gene and condition vectors allowing to display the relative phases and amplitudes in genes space, and how these genes modules are scale- and phase-shifted in tissues (visualized in two separate polar plots). We mostly displayed the rank-one approximations and indicated the fraction of the variance (explained variance) explained by those. In the representations, we set as a convention the maximum amplitude of the tissue vector to 1 (so that the gene amplitudes could be read directly on the gene graphs) and the mean phase of the tissue vector to zero (so that the gene phases correspond to daily time). Thus, this allowed the gene representation to correctly show both amplitude and phase of the genes. For each relevant gene set (CRGs, GO gene sets, TF targets), we constructed $N_{\text{gene}} \times N_{\text{tissue}}$ matrices containing the Fourier coefficients ($a_g + ib_g$) in complex notation and applied the SVD decomposition. The first left (gene) and right (tissue) complex vectors of the cSVD were shown in the figures and used to calculate the metrics described below.

Gene Ontology analysis with cSVD

We considered all sets of genes in each Gene Ontology (GO) category (biological process, molecular function, and cellular component) as well as Wikipathways [263] and KEGG pathways and performed the cSVD. We identified common (Fig. S2A) and tissue-specific (Fig. S2B) rhythmic functions. To prevent highly varying genes with poor sinusoidal fits from dominating the cSVD, for this analysis, we renormalized the complex matrix entries such that the amplitudes (norms of the complex numbers) reflected the fraction of explained variance by the harmonic fit for each gene in each tissue. This analysis using the cSVD allowed us to calculate a number of useful metrics for each GO term (see below for a detailed list) and used to make selections. Commonly rhythmic functions in (Fig. S2A) have been selected due to either a high variance explained (variance) by the first SVD component, a high phase coherence of the mean peak time across tissues (tissue_phase_similarity), or a high

coherence of the peak times of the genes across tissues (gene_phase_similarity). In particular, we ordered the GO terms according to these three metrics and extracted the top 50 from each. We then removed redundancies and eliminated general GO terms to reach the list in Fig. S2A. Finally, by inspecting all genes in these functions we coarse-grained the terms into functional themes for easier interpretation. These gene functions are displayed as ordered by (mean) time of day. To identify more tissue-specific circadian mRNA programs, we filtered functions according to either a high entropy of the first tissue vector (tix_entropy), or a high standard deviation of the mean peak phase across tissues (tissue_phase_similarity) (Fig. S2B). As before, we ordered the GO terms according to these three metrics and extracted the top 50 from each. Then we removed redundancies and eliminated general GO terms to remain with the list in Fig. S2B. The coarse-grained themes often overlap between Figs. S2A and S2B, however, the specific pathways or genes involved in these functions are often not exactly the same.

Metrics calculated for all GO or Wikipathway gene sets

The metrics indicated in bold are the columns of Table S5. Variance: variance explained by the 1st cSVD component. Mean_gene_time: argument of the sum of complex valued gene parameters of 1st cSVD component. Mean_ampl: mean of absolute values of complex valued gene parameters of 1st svd components. Tix_entropy: entropy of the probability distribution given by the modulus squared of the complex valued tissue parameters of 1st cSVD component. Gene_phase_similarity: modulus of the sum of complex valued gene parameters of 1st cSVD component. Tissue_phase_similarity: modulus of the sum of complex valued tissue parameters of 1st cSVD component.

GO enrichment analysis

For the functional analysis of the stratified analysis (sex and age groups), we used the “enrichR” R package providing an interface to the Enrichr database [264]. For each model, we performed enrichment analysis for three databases: GO_Biological_Process_2021, KEGG_2021_Human, WikiPathway_2021_Human. Terms with an adjusted p-values smaller than 0.05 and a combined score larger than 20 were selected. Duplicated terms with a similar list of genes were discarded. The top fifteen terms (i.e. smallest adjusted p-values) across databases were reported in the heatmaps (e.g. Fig. S3A). Selected Wikipathways were visualized using Cytoscape [265] using the “RCy3” R package [266].

Transcription factor analysis with the MARA model

This approach was used for the stratified analyses (sex and age groups). ChIP-Seq target gene data were downloaded from ChIP-Atlas for the H. Sapiens (hg38) genome [?]. Peak-calls overlapping a 5kb window around the TSS of a gene were assigned to it. For each sample, MACS2 score was normalized to its maximum value. For each transcription factor (TF), the normalized MACS2 score was averaged across all samples. To infer TF activity, we adapted a penalized regression model [267] as previously [19]. In the stratified analysis, we inferred sample-specific TF activities using mean-centered mRNA expression level (without subsampling) and the ChIP-Seq matrices described above. As for the mRNA patterns, we performed model selection on the inferred TF activities and selected TFs classified in each model with high confidence (AICW > 0.5). The top 10 TFs in terms of amplitude or explained variance (z-score) were shown in the heatmaps. Gene expression, GO term analysis and TF activity inference were visualized using the complexheatmap R package [268].

Transcription factor (TF) analysis with cSVD

We performed cSVD analysis on sets of transcription factor (TF) target genes in a similar manner to the analysis of GO terms. This allowed us to visualize the predicted peak phase activities of each TF throughout the day (Fig. S1N). In order to build a matrix of genes times conditions, we identified sets of transcription factor targets from human ChIP-Atlas (46). We used the same process and normalization as described above and selected the top 100 target genes for each TF. We performed cSVD on the resulting complex matrix (Supplementary Table ST5). This allowed us to calculate the same metrics as for the GO term analysis for each TFs. As for the GO terms, we used the renormalized Fourier coefficients to account for explained variance.

Dynamically consistent RNA velocity

Part IV

5 RNA velocity 2.0

5.1 Introduction

This work to improve RNA velocity [180] is underway with a collaboration with the creator of RNA velocity, Gioele La Manno, as well as Felix Naef, Alex Lederer, Maxine Leonardi; Colas Droin also contributed to a first analysis and implementation. The mathematical part described in this chapter has been developed by Lorenzo Talamanca under the supervision of Felix Naef and Gioele La Manno. This is still a work in progress and will evolve until publication.

5.1.1 The general problem

RNA velocity has brought to single cells the idea that we can predict the future state of a cell by exploiting unspliced mRNA counts. We can draw arrows connecting the present location of the cell to its future one. The set of these arrows for all cells is a velocity field on the low dimensional manifold. This technique has been favourably adopted by the community and has allowed to infer developmental and differentiation trajectories from single cell omics snapshots. The inner consistency of the velocity field, both geometrical and derived from a dynamical systems perspective have not yet been exploited. We want to introduce these constraints which would also allow us to integrate the velocity field and find the trajectories of the cells. The possibility of inferring time intervals between different states of cells from a single snapshot could be used to greatly improve our knowledge of cell differentiation and development. A good process to develop and test a more precise formulation of RNA velocity is the cell cycle, due to its known low dimensional manifold and experimentally measurable period.

5.1.2 Estimation of cell cycle period

The cell cycle is a ubiquitous biological process, so a more in depth knowledge of how exactly embryonic cells progress through it could be a great tool to better understand the first steps of life. In particular, we want to infer a phase dependent speed, and from it the length of the cell cycle. In addition, we aim to be able to compare the shape of the speed of cells undergoing the cell cycle across cell types and species.

5.2 The general model

We describe the general mathematical formulation that accounts for geometrical and dynamical constraints. Let's assume that our process of interest can be projected onto a low-dimensional manifold, \mathcal{M} . For example $\mathcal{M} = \mathbb{R}^2, S^1$. All the variance in the data not explained by changes in the position on \mathcal{M} is assumed to be noise. The right¹ inverse link function from the low dimensional manifold to the high dimensional gene space comprised of spliced and unspliced reads is defined as $\mathcal{L}^{-1} : \mathcal{M} \rightarrow \mathcal{O}$, where \mathcal{O} is the high dimensional gene space. Thus, each measured cell (c) provides an $|\mathcal{O}|$ -dimensional realisation $Y_c = (U_c, S_c) \in \mathcal{O}$ of $y_c = (u_c, s_c) = (u(x_c), s(x_c)) = \mathcal{L}^{-1}(x_c)$, and $x_c \in \mathcal{M}$ is the position of cell c on the manifold \mathcal{M} . As a notation capital letters represent the result of measurements while lower-case letters the respective theoretical noiseless values. The projection from the gene space to \mathcal{M} can be both probabilistic, with a probability distribution $\mathcal{P}(x_c | Y_c)$, or deterministic inferred using one of the methods described in section 1.2.1. In the high-dimensional space, we expect some dynamical equation governing the time-evolution of the spliced mRNA depending both on spliced and unspliced mRNA:

$$\frac{dS}{dt} = F(S, U). \quad (5.1)$$

We do not expect to be able to infer F . Instead we assume F to be the one described in [180]. For each gene we have:

$$\dot{S} = \beta U - \gamma S \quad (5.2)$$

with β and γ being the splicing and degradation rate. This equation cannot make predictions (for example we can't integrate it), since it's not an autonomous equa-

¹ Given $f : A \rightarrow B$ the right inverse function is the function $f^{-1} : B \rightarrow A$ such that $f \circ f^{-1} \equiv I : B \rightarrow B$, but not necessarily $f^{-1} \circ f \equiv I : A \rightarrow A$

tion for S . We also remind the assumed dynamics for the unspliced counts:

$$\dot{U} = \alpha - \beta U, \quad (5.3)$$

where α is the transcription rate. All the parameters governing gene dynamics (α , β , γ) could in principle be dependent on $x \in \mathcal{M}$ or time, but won't in our formulation. Our crucial assumption is that there exists an autonomous and deterministic equation for x :

$$\frac{dx}{dt} = V(x), \quad (5.4)$$

which provides a low-dimensional approximation of the full dynamics of equation (5.1). Thus, on \mathcal{M} , we expect a velocity field $V(x)$ which could also be explicitly time-dependent but will not due to the assumptions on α , β , and γ . $V(x)$ is the deterministic field describing the temporal evolution on the low dimensional manifold, the actual velocity field in low dimensions, and is the quantity of interest. We can now link the geometry of the problem, with its dynamical formulation in terms of biological parameters. Assume $x(t)$ is a deterministic trajectory, *id est* it satisfies $\frac{dx}{dt} = V(x(t))$, then

$$\frac{ds_g(x(t))}{dt} = (\nabla_x s_g) \cdot V(x(t)) = \beta_g u_g(x(t)) - \gamma_g s_g(x(t)). \quad (5.5)$$

Here again β_g and γ_g are the gene specific splicing and degradation rates. Equation (5.5) is the base of all our theory as it connects the topology of the low dimensional manifold on the left with the biology on the right. Of note, from now on we assume that \mathcal{M} can be fully determined by the spliced counts S alone. Biological observables, such as time, are independent of the parametrization² of \mathcal{M} . This is not self-evident in our formulation and should therefore be verified.

5.2.1 Properties of the velocity field

As time is an observable, the time between two points (connected by a trajectory allowed by the dynamical system) on \mathcal{M} shouldn't change depending on the chosen parameterization of the low dimensional manifold. We check that this property is

²If unclear, the parametrization of a manifold is the system of coordinates, the base, used to describe points on it. For example we can think of a segment connecting the points A and B , $A, B \in \mathbb{R}^N$. It can be parametrized in many ways such as $\{Y|Y = A + (B - A)x, x \in (0, 1)\}$, or $\{Y|Y = (A + B)/2 + (B - A)/2x, x \in (-1, 1)\}$, or even $\{Y|Y = A + (B - A)x^2, x \in (0, 1)\}$

verified with our formulation. In general, we have:

$$\Delta t_{x_0, x_1} = \int_{\Gamma_{x_0}^{x_1}} \frac{1}{v(x)} dx. \quad (5.6)$$

$\Gamma_{x_0}^{x_1}$ is the trajectory $x(t)$ that connects the two points. In particular it is useful to notice that $x_i = x(S_i)$. In addition, we remind:

$$\frac{\partial S}{\partial t} = \frac{\partial S}{\partial x} \frac{\partial x}{\partial t} \quad \text{and} \quad v(x) = \frac{\partial x}{\partial S} \frac{\partial S}{\partial t}. \quad (5.7)$$

For any parameterization of \mathcal{M} we can apply a change of variable and obtain:

$$\Delta t_{x_0, x_1} = \int_{\Gamma_{x_0}^{x_1}} \frac{\partial t}{\partial x} dx = \int_{\Gamma_{x_0}^{x_1}} \frac{\partial t}{\partial S} \frac{\partial S}{\partial x} dx = \int_{\Gamma_{S_0}^{S_1}} \frac{\partial t}{\partial S} dS = \int_{\Gamma_{S_0}^{S_1}} \frac{1}{\dot{S}} dS. \quad (5.8)$$

Thus, time intervals are, as they should be, a property of the original high dimensional space which is not changed by different possible parametrizations of \mathcal{M} . This means that the details of the \mathcal{L} function are not relevant as long as it links the same low and high dimensional manifold. This is a very important property allowing us not to worry about how we parameterize the low dimensional manifold. We now simplify our theory and reach an optimization function focusing on the cell cycle.

5.3 Measuring cell cycle period

In this section we will develop the same theory as before, focusing on the cell cycle. In this case we know the low dimensional manifold and we can build a probabilistic formulation to infer the parameters of interest. To link it with the previous formulation: we are assuming that \mathcal{M} is the one dimensional circle $\mathcal{M} \equiv S^1$ and the low dimensional coordinate is a phase $x \equiv \varphi$. We develop a general theory, but we will use in our inference only genes known to oscillate along the cell cycle to reduce noise. The precise list of genes that we will use for the practical inference is not yet fully determined, but comprises about 100 markers of different cell cycle phases. The equation of the dynamics (5.5) becomes:

$$\dot{s} = \partial_\varphi s(\varphi) \omega(\varphi) = \beta u - \gamma s, \quad (5.9)$$

where we again assume that the β s and the γ s are not cell cycle dependent, i.e. $\beta(\varphi) = \beta$ and $\gamma(\varphi) = \gamma$. The parameters of this equation are strongly constrained by their biological function, as detailed in appendix B.1. However, these constraints

will allow us to translate into hour the period of the cell cycle we obtain, in a very direct way explained in appendix B.2. Our coordinate (phase on the cell cycle) follows the equation:

$$\dot{\varphi} = \omega(\varphi). \quad (5.10)$$

We notice that the spliced counts exhibit a "sinusoidal" behaviour in the log transformed space, and this dictates our choice of basis. We consider up to n Fourier components in our expansion (from a practical perspective we will use $n \sim 3$). This means that in the linear space we can write:

$$s_g(\varphi_c) = e^{\sum_f v_{gf} \zeta_{fc}}, \quad (5.11)$$

with

$$v_g = \begin{pmatrix} a_g^0 \\ a_g^1 \\ b_g^1 \\ \vdots \\ a_g^n \\ b_g^n \end{pmatrix} \quad \zeta = \begin{pmatrix} 1 \\ \cos(\varphi) \\ \sin(\varphi) \\ \vdots \\ \cos(n\varphi) \\ \sin(n\varphi) \end{pmatrix} \quad \zeta_c = \begin{pmatrix} 1 \\ \cos(\varphi_c) \\ \sin(\varphi_c) \\ \vdots \\ \cos(n\varphi_c) \\ \sin(n\varphi_c) \end{pmatrix}. \quad (5.12)$$

In this model the v_g is the vector of gene Fourier parameters that identifies the behaviour of gene g . We fix a log normal error model on the spliced and unspliced counts that we write as:

$$\ln(S_{gc}) = \ln(s_g(\varphi_c)) + \varepsilon_{gc} \quad (5.13)$$

$$\ln(U_{gc}) = \ln(u_g(\varphi_c)) + \varepsilon'_{gc}, \quad (5.14)$$

where ε is the Gaussian noise. The predicted spliced are defined in equation (5.11) and the predicted unspliced u are a non trivial function of $\{v\}$, $\{\varphi\}$, $\{\beta\}$, $\{\gamma\}$ and ω . We will evaluate this function later, imposing the dynamical equation (5.9). Before that, it is relevant to notice that we have:

$$\mathcal{P}(U, S | \{\varphi\}, \{v\}, \omega(\varphi), \{\beta\}, \{\gamma\}) \quad (5.15)$$

as a joint probability. However, looking a bit more closely at the model we have made, especially at equation (5.11), we notice that there is a hierarchy. For our formulation $s = f(\varphi, v)$, while u is influenced by all the parameters of the model.

Thus translates to formulae as:

$$\mathcal{P}(U, S | \{\varphi\}, \{\nu\}, \omega(\varphi), \{\beta\}, \{\gamma\}) = \quad (5.16)$$

$$\mathcal{P}(U | \{\varphi\}, \{\nu\}, \omega(\varphi), \{\beta\}, \{\gamma\}) \mathcal{P}(S | \{\varphi\}, \{\nu\}). \quad (5.17)$$

This means that we can proceed step wise, greatly simplifying the inference at the cost of small approximations. In particular, we exploit $\mathcal{P}(S | \{\varphi\}, \{\nu\})$ via either maximum likelihood [269], variational inference [270], maximum a posteriori inference methods [271, 272] (such as CHIRAL described in section 3) to infer both the $\{\varphi\}$ and $\{\nu\}$. Then we inject the parameters inferred using the S to condition our expected value of U . We notice that the $\{\varphi\}$ effectively defines the manifold and is inferred by using only the S , in line with the assumption that the low dimensional manifold is determined by the spliced counts alone. We are interested to have a biological meaning in our low dimensional manifold. Therefore, we rotate our definition of the phase to allow $\varphi = 0$ to be mitosis (appendix B.3). Now we write the explicit formulation of u , remembering that $\{\varphi\}$ and $\{\nu\}$ are known. Using the chain rule we can explicitly write the dynamics in our notation:

$$\partial_{\varphi} s_g(\varphi) \big|_{\varphi=\varphi_c} = \sum_{ff'} \nu_{gf} D_{ff'} \zeta_{f'c} \omega(\varphi_c) s_g(\varphi_c), \quad (5.18)$$

with $D_{ff'}$ defined as:

$$D_{ff'} = \left(\delta_{\left\lceil \frac{f}{2} \right\rceil, \left\lceil \frac{f'}{2} \right\rceil} - \delta_{f, f'} \right) \left\lceil \frac{f}{2} \right\rceil (-1)^f. \quad (5.19)$$

Thus we have an expression for the expected unspliced, u :

$$\begin{aligned} \beta u - \gamma s &= \partial_{\varphi} s(\varphi) \omega(\varphi) \\ \rightarrow u &= \frac{1}{\beta} (\partial_{\varphi} s(\varphi) \omega(\varphi) + \gamma s) \\ \rightarrow u &= \frac{1}{\beta_g} \left(\sum_{ff'} \nu_{gf} D_{ff'} \zeta_{f'c} \omega(\varphi_c) + \gamma_g \right) s_g(\varphi_c). \end{aligned} \quad (5.20)$$

This can be combined with our error model written in equation (5.13) to write

$$\ln(U_{gc}) = \ln \left(\frac{1}{\beta_g} \left(\sum_{ff'} \nu_{gf} D_{ff'} \zeta_{f'c} \omega(\varphi_c) + \gamma_g \right) s_g(\varphi_c) \right) + \varepsilon_{gc}. \quad (5.21)$$

Rewriting all in a probabilistic formulation we obtain

$$\begin{aligned} \mathcal{P}(U|\{\varphi\}, \{\nu\}, \omega(\varphi), \{\beta\}, \{\gamma\}) \sim \\ \exp\left(-\frac{1}{2\sigma^2} \sum_{gc} \left(\ln(U_{gc}) - \ln\left(\frac{1}{\beta_g} \left(\sum_{ff'} \nu_{gf} D_{ff'} \zeta_{f'c} \omega(\varphi_c) + \gamma_g \right) s_g(\varphi_c) \right) \right)^2\right) \sim \\ \exp\left(-\frac{1}{2\sigma^2} \sum_{gc} \left(\ln(U_{gc}) + \ln(\tilde{\beta}_g) - \ln\left(\left(\sum_{ff'} \nu_{gf} D_{ff'} \zeta_{f'c} \tilde{\omega}(\varphi_c) + \tilde{\gamma}_g \right) s_g(\varphi_c) \right) \right)^2\right), \end{aligned} \quad (5.22)$$

where we have introduced the unitless notation: $\tilde{\beta}_g = \beta_g / \bar{\beta}$, $\tilde{\gamma}_g = \gamma_g / \bar{\beta}$, $\tilde{\omega}(\varphi) = \omega(\varphi) / \bar{\beta}$. The geometric mean of the prior probability distribution of β is $\bar{\beta}$. We will soon explicitly write the distribution of β . As we want to marginalise over $\tilde{\beta}$, we want to rewrite $\mathcal{P}(U|\dots)$ as:

$$\mathcal{P}(U|\{\varphi\}, \{\nu\}, \tilde{\omega}(\varphi), \{\tilde{\beta}\}, \{\tilde{\gamma}\}) \sim \exp\left(\sum_g \left(-\frac{1}{2} a_g \ln(\tilde{\beta}_g)^2 + b_g \ln(\tilde{\beta}_g) + c_g \right)\right). \quad (5.23)$$

We need to calculate the gene specific coefficients that appear in the integral: (a, b, c) .

$$\sum_{gc} \left(\ln(U_{gc}) + \ln(\tilde{\beta}_g) - \ln\left(\left(\sum_{ff'} \nu_{gf} D_{ff'} \zeta_{f'c} \tilde{\omega}(\varphi_c) + \tilde{\gamma}_g \right) s_g(\varphi_c) \right) \right)^2 = \quad (5.24)$$

$$\sum_{gc} \left(\ln(\tilde{\beta}_g) - \ln\left(\frac{(\sum_{ff'} \nu_{gf} D_{ff'} \zeta_{f'c} \tilde{\omega}(\varphi_c) + \tilde{\gamma}_g) s_g(\varphi_c)}{U_{gc}} \right) \right)^2 = \quad (5.25)$$

$$\sum_g |C| \ln(\tilde{\beta}_g)^2 - \sum_{gc} 2 \ln(\tilde{\beta}_g) \ln\left(\frac{(\sum_{ff'} \nu_{gf} D_{ff'} \zeta_{f'c} \tilde{\omega}(\varphi_c) + \tilde{\gamma}_g) s_g(\varphi_c)}{U_{gc}} \right) + \quad (5.26)$$

$$\sum_{gc} \ln\left(\frac{(\sum_{ff'} \nu_{gf} D_{ff'} \zeta_{f'c} \tilde{\omega}(\varphi_c) + \tilde{\gamma}_g) s_g(\varphi_c)}{U_{gc}} \right)^2. \quad (5.27)$$

This directly means:

$$a_g = \frac{|C|}{\sigma^2} \quad (5.28)$$

$$b_g = \frac{1}{\sigma^2} \sum_c \ln\left(\frac{(\sum_{ff'} \nu_{gf} D_{ff'} \zeta_{f'c} \tilde{\omega}(\varphi_c) + \tilde{\gamma}_g) s_g(\varphi_c)}{U_{gc}} \right) \quad (5.29)$$

$$c_g = \frac{-1}{2\sigma^2} \sum_c \ln\left(\frac{(\sum_{ff'} \nu_{gf} D_{ff'} \zeta_{f'c} \tilde{\omega}(\varphi_c) + \tilde{\gamma}_g) s_g(\varphi_c)}{U_{gc}} \right)^2. \quad (5.30)$$

For the moment we can forget about c_g as it is only a normalisation constant for the $\tilde{\beta}$ integral. The integral we want to solve to marginalise over $\tilde{\beta}$ is:

$$\int \mathcal{P}(U|\{\varphi\}, \{\nu\}, \tilde{\omega}(\varphi), \{\tilde{\beta}\}, \{\tilde{\gamma}\}) \mathcal{P}(\{\tilde{\beta}\}|\{\varphi\}, \{\nu\}, \tilde{\omega}(\varphi), \{\tilde{\gamma}\}) d\tilde{\beta}. \quad (5.31)$$

We need a (conditioned) prior on β to perform this integral. For ease of analytics we use a log-normal distribution. The geometric mean of this prior should set a time scale for all the other parameters in play. Given the intrinsic scale invariance of the system we will set $\bar{\beta} = 1$, for simplicity. However, the spread that we expect is important, thus we need to carefully treat the τ parameter. As the β s are susceptible to a great deal of technical noise we will assume $\tau = 2$, thus leaving great variability to the splicing rate. This parameter will determine the impact of the prior on the future optimization. In formulae:

$$\mathcal{P}(\{\beta\}|\{\varphi\}, \{\nu\}, \omega(\varphi), \{\gamma\}) \sim \prod_g \frac{1}{\tilde{\beta}_g} e^{-\frac{1}{2\tau^2} \ln(\tilde{\beta}_g)^2} \quad (5.32)$$

So, considering all the definitions we have provided up to now, we can explicitly write our integral:

$$\int \mathcal{P}(U|\{\varphi\}, \{\nu\}, \tilde{\omega}(\varphi), \{\tilde{\beta}\}, \{\tilde{\gamma}\}) \mathcal{P}(\{\tilde{\beta}\}|\{\varphi\}, \{\nu\}, \tilde{\omega}(\varphi), \{\tilde{\gamma}\}) d\tilde{\beta} \sim \quad (5.33)$$

$$\prod_g \int e^{(-a_g \ln(\tilde{\beta}_g)^2 + b_g \ln(\tilde{\beta}_g) + c_g)} e^{-\frac{1}{2\tau^2} \ln(\tilde{\beta}_g)^2} \frac{1}{\tilde{\beta}_g} d\tilde{\beta}_g \sim \quad (5.34)$$

$$\prod_g \int e^{(-a_g \ln(\tilde{\beta}_g)^2 + b_g \ln(\tilde{\beta}_g))} e^{-\frac{1}{2\tau^2} \ln(\tilde{\beta}_g)^2} d\ln(\tilde{\beta}_g). \quad (5.35)$$

Ideally it could be interesting to also integrate out $\tilde{\gamma}$. However, it would be a bit more tricky and would require some approximations. At this moment we have:

$$\mathcal{P}(U|\{\varphi\}, \{\nu\}, \tilde{\omega}(\varphi), \{\tilde{\gamma}\}, \tilde{\beta}, \tau) \sim \prod_g e^{\frac{(b_g)^2}{2(a_g + \frac{1}{\tau^2})} + c_g}. \quad (5.36)$$

This does not really give us any intuition, so we should explicitly write the coefficients. Before that it is useful to introduce:

$$\tilde{u}_g = u_g(\beta_g = \tilde{\beta}) = \frac{\beta_g}{\tilde{\beta}} u_g = \left(\sum_{f,f'} v_{gf} D_{ff'} \zeta_{f'c} \tilde{\omega}(\varphi_c) + \tilde{\gamma}_g \right) s_g(\varphi_c) \quad (5.37)$$

By referring to the appendix in section A.3 we can solve the integral and write:

$$\mathcal{P}(U|\{\varphi\}, \{\nu\}, \tilde{\omega}(\varphi), \{\tilde{\gamma}\}, \tilde{\beta}, \tau) \sim \prod_g e^{-\frac{1}{2\sigma^2} \sum_c \left(\ln\left(\frac{\tilde{u}_{gc}}{U_{gc}}\right) - \langle \ln(\tilde{\beta}_g) \rangle \right)^2} \quad (5.38)$$

with

$$\langle \ln(\tilde{\beta}_g) \rangle = \frac{\frac{1}{\sigma^2} \sum_c \ln\left(\frac{\tilde{u}_{gc}}{U_{gc}}\right) + \frac{1}{\tau^2} \ln(\tilde{\beta})}{\left(\frac{|C|}{\sigma^2} + \frac{1}{\tau^2}\right)} = \frac{\sum_c \ln\left(\frac{\tilde{u}_{gc}}{U_{gc}}\right)}{\sigma^2 \left(\frac{|C|}{\sigma^2} + \frac{1}{\tau^2}\right)} \quad (5.39)$$

as $\tilde{\beta} = 1$. However, care is still needed as $\tilde{u}_{gc} = \tilde{u}_{gc}(\{\varphi\}, \{\nu\}, \tilde{\omega}(\varphi), \{\tilde{\gamma}\}, \tilde{\beta}, \tau)$. In addition, as we have take the derivative of the s to obtain the u , we need to be careful to predict plausible shapes for the unspliced and not venture into un-biological territory allowed by the geometry of the system, as detailed in appendix B.4. We can now solve the problem of fitting the best parameters for (5.38) either with simple maximum likelihood, or set some priors on the parameters of interest and infer the maximum or expected value of their posterior distribution.

A note on differential geometry

We just want to remind how the properties of manifold and velocity field pervade our formulation. In particular, by definition a velocity field constructed according to our method lies in the tangent bundle of a manifold. This geometrical constraint is automatically inserted in equation (5.38). In fact the combination of the expected values of spliced and unspliced, s and u , is always such that the velocity lies in the tangent bundle of \mathcal{M} in its high dimensional parameterization.

5.4 Implementation

We describe here the computational implementation of the models described above. The computational implementation has been implemented by Alex Lederer, Maxine Leonardi, Giole La Manno, and Felix Naef. It is briefly reported here to show the development of the project. The mathematical models needed for the implementation have been calculated by Lorenzo Talamanca.

5.4.1 Probabilistic programming with pyro

We implemented our model in a flexible probabilistic programming language, pyro [273]. Pyro, contrary to the majority of programming languages, allows one to analytically write all the probability distributions in play. After the optimization it

returns a posterior probability density for the variables of interest while marginalising over the hidden variables. So, we have access to the full posterior distributions on the variables of interest and not only a point estimator. However, to comply with the fully probabilistic model we need priors on all the parameters of the system. In addition, Pyro needs as input normalised probability distributions. Fortunately, for many known distributions it is able to calculate the normalisation coefficients. However, the distributions must be given as the correct set of parameters. In particular, for Gaussian and log-normal distributions we must provide the mean and covariance or precision matrix. So, we calculate the pyro-approved parametrization of $\mathcal{P}(U|\{\varphi\}, \{v\}, \omega(\varphi), \{\gamma\}, \tilde{\beta}, \tau) = \mathcal{P}(U|u)$ in detail in appendix B.5. Once this complies with the requirements of pyro we fix the priors on all our parameters and proceed with the inference.

Prior distributions

We have set a series of priors on the parameters of interest. The choices were mainly based in literature and to allow biological meaning of the parameters. We chose:

$$\begin{aligned}\mathcal{P}(\beta) &\sim \frac{1}{\beta} e^{\frac{1}{2*2^2} (\ln(\beta))^2} \\ \mathcal{P}(\gamma) &\sim \frac{1}{\gamma} e^{\frac{1}{2*0.5^2} (\ln(\gamma)-0.25)^2} \\ \mathcal{P}(v_\omega) &\sim e^{\frac{1}{2} v_\omega^T A v_\omega}\end{aligned}\tag{5.40}$$

with

$$A_{ij} = \frac{20}{2 * \pi} \delta_{ij} e^{-[\frac{i}{2}]}\tag{5.41}$$

All these prior probabilities are already formulated in an acceptable way for pyro. Now we will re write $\mathcal{P}(U|u)$ to comply with pyro's input requirements.

5.4.2 Preliminary results

We applied our method to a developmental mice brain atlas [274]. We focused on inferring cell cycle speed for four different brain regions at embryonic day 10.5 (conception is embryonic day 0). The regions analysed are the forebrain, midbrain, hindbrain, and midbrain-hindbrain boundary. To compare datasets and account for different overall expression levels a small dose of care is needed (appendix B.6). We have inferred the v s and φ s simultaneously on all four regions. Then, we have inferred the γ s across datasets while the $\omega(\varphi)$ on each one separately. We

re-scaled all the velocities $\omega(\varphi)$, dividing them by the geometric mean of the γ s, γ_0 . The different constant velocities, therefore periods, for the various regions, are displayed in figure 5.1. We use the inference with a flat omega to infer the period; we set the mean³ inferred half life for the genes considered to be 45 minutes to be able to measure the period of the cell cycle in hours (appendix B.2). We notice that the cell cycle is faster, with a period of about 21h, in the forebrain. In the boundary between the midbrain and hindbrain the cell cycle period is 40h, shorter in a puzzling way than the period of the midbrain which is 55h. In the hindbrain the period is about 100h, which suggests an almost stopped cell cycle. This is almost in accordance with the general idea that neurons in the forebrain are still undergoing cell cycle while as the mature and differentiate they stop [274–278]. Although the standard deviation looks large compared to the mean it is important to notice that it is the standard deviation of the posterior distribution, not the standard deviation of the mean of the posterior. However, for the midbrain, hindbrain and the boundary between the two, the sample standard deviation makes the effect size of the differences in mean relatively small. This is especially important when comparing the midbrain with either the boundary or the hindbrain. The difference in period could be also due to the different proportion of cycling and non-cycling cells in the various regions. We are currently working to solve this problem with a mixture model consider only cycling cells for our velocity estimation. We also inferred $\omega(\varphi)$ up to the first Fourier harmonic, shown in figure 5.2. The velocity is more dependent on the position along the cycle for the midbrain. For the hindbrain $\omega(\varphi)$ becomes negative at certain cell cycle phases probably signalling the fact that these cells no longer cycle. The biological meaning of these findings is still unclear. These results, like the mice brain dataset we exploited, are still in an embryonic stage, however they look promising for relevant future developments.

5.4.3 Next steps

We need to expand our method, verify its robustness and wide applicability. We can start by including more data from the mouse atlas. We then need to validate the approach, to ensure that our estimated cell-cycle periods are biologically relevant. To do this, we could measure the period experimentally, for example with the FUCCI system [197], in various conditions and match the experimental results to the computational inference. In a more distant future it could be interesting to cross⁴ the circular manifold of the cell cycle with any developmental manifold. Information

³In the geometric sense

⁴In the geometrical sense of making a cross product between two spaces

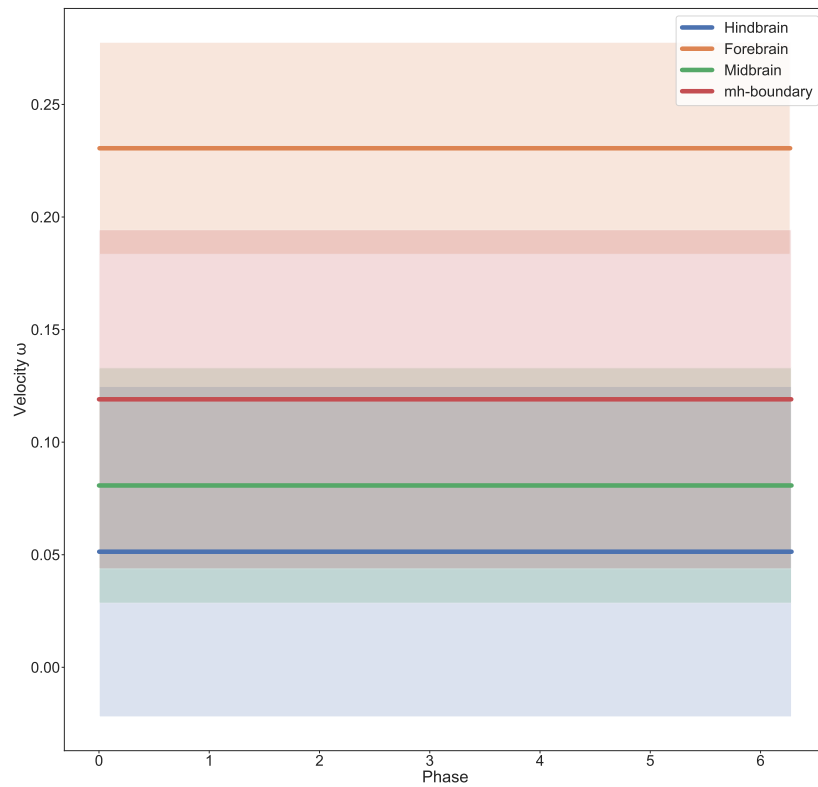


Figure 5.1: **Inference of a flat $\omega(\varphi)$.** Cell cycle speed of cells across different developmental brain regions recapitulates faster cell cycle in the forebrain and a stopped cycle in the hindbrain. Bright colour is the mean, veiled is the standard deviation of the posterior distribution. Cell cycle phase on the x-axis between 0 and 2π , unitless $\omega(\varphi)/\gamma_0$ on the y-axis.

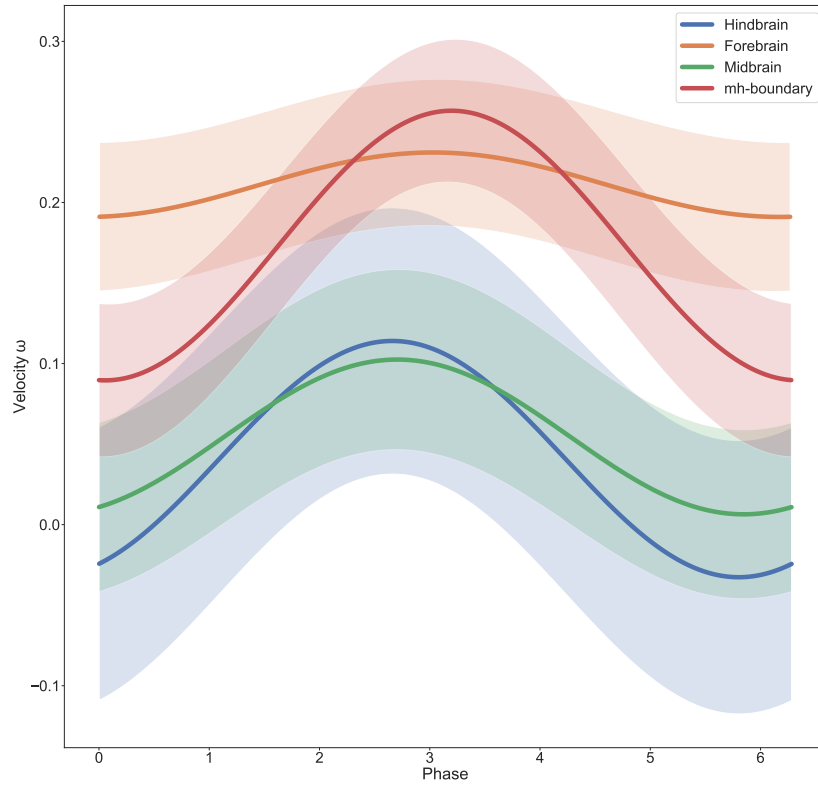


Figure 5.2: **Inference of a phase dependent $\omega(\varphi)$.** Oscillatory behaviour of $\omega(\varphi)$ shows stronger speed differences as a function of cell cycle phase in the midbrain. Bright colour is the mean, veiled is the standard deviation of the posterior distribution. Cell cycle phase on the x-axis between 0 and 2π , unitless $\omega(\varphi)/\gamma_0$ on the y-axis.

on cell cycle phase and speed in each part of the developmental manifold could reveal novel biological properties.

Conclusion & outlook

Part V

6 Conclusions and perspectives

6.1 Contributions of this thesis

We are now at the end, so we will summarise all the relevant contributions to science encapsulated in this thesis. The objective results will be presented in plural form, while personal, and sometimes more philosophical, reflections and speculations are written in the singular form.

6.1.1 Circadian phase inference

The first problem we tackled was the one of circadian phase inference, given the high abundance of unlabelled samples from a variety of different studies and our interest in chronobiology. I think this problem is deeply relevant as it opens the opportunity to reduce the need for animal experiments while studying the circadian clock. Experiments for chronobiology require time points around the clock as well as biological replicates to ensure a robust signal to noise ratio, resulting in at least 8, but more often around 50, animals sacrificed per experiment. Thus, I think it is paramount to widely exploit computational methods of circadian phase inference across the circadian field. We have seen that computational methods are already available, however the community has not yet agreed on a golden standard.

We developed a rigorous, fully probabilistic approach to the problem. It has shown its high level performance on mice and human labelled data outperforming the only other existing unsupervised method, CYCLOPS. I think an interesting part of CHIRAL is its deeply rooted analytical formulation. In fact, we pushed the analytical calculations as much as we could, even when most would have resorted

to computational optimization. This is shown both in the recursive solution of the spin model, in the introduction of Lagrange multipliers, and in the analytical maximisation step in the EM. CHIRAL also differs from CYCLOPS for the mixture model we implemented that is able to discard genes that do not carry temporal information. Overall, the improvement we bring is not only from a results perspective, but also from a conceptual point of view. I believe that all these little details have in the end paid off, allowing CHIRAL to outperform existing methods. We also developed an R package for CHIRAL, making it easily accessible and applicable to project any omics data onto relevant circular manifolds. The main and most crucial application we found for CHIRAL was ordering about 16000 samples from the GTEx project. I hope this set of ordered samples will be used as a circadian atlas by the community fuelling future studies on human chronobiology.

6.1.2 Human chronobiology

We have discussed the inner complexities of studying chronobiology in humans, especially across tissues. I think the majority of these difficulties are born from little communication among scientists and between the clinic and researchers. We are very grateful to the GTEx project for allowing us to access a vast collection of human samples spanning the whole body. We exploited the data, after cleansing it from biases and assigning a circadian phase to each donor, to discover a rich picture of human 24h rhythms. This was the first time that the clock could be studied so in depth in humans in such a variety of tissues. In particular, we are also able for the first time to compare the timing of tissues, although only in an average sense.

We have studied the clock across the body, its structure is conserved while the amplitude of oscillations of core clock gene expression is tissue dependent. In addition, downstream rhythms vary in intensity and breadth throughout the body. In accordance with the expectations from the field, the oscillations in the brain are less frequent and have lower amplitude than those in metabolic tissues, such as the liver. Nonetheless, the heat shock response was highly rhythmic in brain tissues, peaking at the time of highest body temperature, and not as marked in the rest of the body. We have shown an ensemble of rhythmic functions occurring throughout the body every day, from carbohydrate metabolism in the morning to lipid homeostasis in the afternoon and immunity in the night. I think it could be relevant to make these rhythms more known to the general population, as respecting them could increase general well-being. We then stratified our analysis to compare rhythmic oscillations between sexes.

We showed similar behaviour of the circadian clock for males and females, while downstream rhythms were overall more female biased. The liver was a very interesting example of sex-dimorphic tissue. It showed more prevalent mRNA oscillations in the females, with almost ten times more rhythmic high amplitude genes. In fact, in the liver more than 1000 mRNA, some driving physiological changes, lose their rhythms in males. In particular, we found a loss of functional oscillations in the male liver for detoxification pathways; both the xenobiotics and cholesterol metabolism pathways are in fact rhythmically regulated only in females. I believe this should remind the medical community to take more care of drug delivery time. In addition, the optimal drug timing might differ between men and women. We also discovered the oscillations of glucocorticoids in the female adrenal gland which could be a putative mechanistic reason for the difference in overall rhythmicity. I think it is very relevant to notice how known stronger female behavioural rhythms are indeed correlated with higher oscillations at the molecular level. We then looked at the differences in circadian rhythmicity occurring with age.

There is the general idea in the field that the clock and its downstream rhythms decay with age. We confirmed a strong decline of circadian oscillations across the body. However, we noticed an intact circadian clock conserved between young and old. In particular, we exposed a general loss of evening peaked mRNA rhythms as age progresses. I believe the notion that the core clock does not deteriorate with age will either deeply change our understanding of the human circadian rhythms, or will be refuted by the community. The coronary arteries showed a relevant pattern of rhythmicity loss as a function of ageing. They revealed reduced oscillation in the cholesterol biosynthesis pathway along with other 800 mRNAs. We also uncovered a radical functional change of rhythmic expression in the ovary. Rhythmic processes switched from being related to biosynthetic and metabolic processes to heat shock and stress response. I think it's quite unexpected to see a functional change rather than a simple reduction of oscillations as a possible result of menopause. Lastly, we showed how rhythmic processes can change phase locking with the environment as a result of ageing. Pituitary, liver, and colon, important regulators of physiology, switched to a 1 : 2 phase locking with earth's period exhibiting 12h ultradian rhythms in elderly, for many processes with circadian oscillations in young.

Overall, we gave a comprehensive analysis of the circadian clock in humans and revealed a rich picture of rhythms specific only to sub populations. We recapitulated some known or expected behaviours of the clock while detailing its downstream

effects on putative physiological rhythms in humans. We also showed some unexpected behaviour of the circadian pacemaker that I hope will shape the way the clock is perceived in the field. While we did not focus strictly on any pathway and its physiological and clinical aspects, I think we provided a great resource for more specialised scientists to explore. I am glad we were able to open the way for, hopefully, many future studies about the impact of sex and age on human circadian rhythms.

6.1.3 Dynamically consistent RNA velocity

We wrote a general theory to include differential geometry and dynamics in the RNA velocity framework. We then derived a more mathematically precise formulation for the special case of the cell cycle. Knowing the low dimensional manifold of interest allowed us to explicitly formulate our theory and greatly progress with analytical calculations. We reached a prediction of unspliced measurement based on the characteristics of the low dimensional manifold, gene parameters and cell cycle speed. Exploiting the prediction we optimised the gene parameters and the cell cycle speed with probabilistic programming. We managed to apply our method for cell cycle speed inference to mouse brain developmental atlas and found interesting results. Although not perfectly in line with biological expectations, the cell cycle velocity showed a relevant decrease for cells outside the forebrain. Still much work is ahead, but I believe that there is promise in this mathematical formulation coupled with the strategy of inference via probabilistic programming. Independently on the practical results of the inference, I think the theoretical formulation we developed is quite elegant and is able to include the beautiful constraints of tangent bundles to a complex biological problem in a graceful way.

7 Future directions

7.1 Cancer and circadian clock

Here we will describe the work that we performed to study the effects of cancer on the clock, and how it should be continued as it is still in a preliminary stage.

Current developments

Many connections have been made between cancer presence and circadian clock malfunction [81–89]. However, still much is to be discovered about the interplay between the clock and cancer. To study more in depth these links and contribute to the field we started analyzing data from The Cancer Genome Atlas (TCGA) program [279]. We applied our methodology to test whether we could exploit the TCGA data to gain new insights about the interplay of cancer and circadian clock. The TCGA dataset contains about 11000 RNA-seq samples from 33 distinct cancers across the body. Samples are often taken when patients undergo surgery; in about 5% of the cases cancerous samples are matched with a sample from healthy tissue surrounding the cancer. We hypothesised that, although the cancerous tissue might harbour a broken clock, healthy tissues surrounding the cancer have an intact circadian pacemaker. Therefore, we transferred circadian phase information from the surrounding healthy tissue to the tumor; we are transmitting time information from robust clocks to presumably broken ones where the inference would conceptually not be possible. We selected tissues from the TCGA data where matched samples were > 36 . The selection left us with 516 matched samples from 12 tissues. We then applied CHIRAL on the healthy samples and assigned the inferred phase also to the cancer sample. Performing our cSVD approach we see how the clock retains its

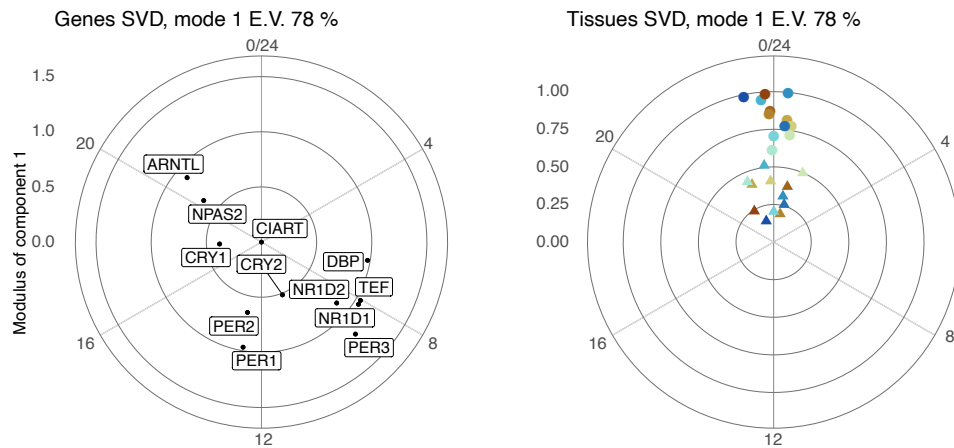


Figure 7.1: **Clock structure in healthy and cancerous tissues.** First gene (left) and tissue (right) vectors of cSVD performed on CRGs indicate highly reduced clock reference genes oscillations in cancerous samples (triangle) compared to healthy tissues (circle). The first module captures 78% of the 24h variance, E.V..

properties both in the healthy and cancerous set of samples, while the amplitude of oscillations is greatly decreased in cancer (figure 7.1). The clock appears still intact with the correct structure in the healthy tissues next to cancerous growth in all cancers studied. This is not an artefact as CHIRAL is an unsupervised algorithm. However, we can see a strong decrease in amplitude of clock oscillations in the various associated tumors that should be investigated further. Also the relatively low percentage of variance explained could be studied more in depth. In particular, relevant contributions to deformation of the clock either in healthy or tumorous tissue could come from further cSVD components.

Future directions

From the circadian ordering of the TCGA data we can learn much in terms of which oscillatory functions are lost when cancer invades healthy tissues. This dataset has already been used to study the connection of various cancer types with the expression of clock genes, but has not been fully exploited. Studies have zoomed on specific cancers and have not assigned any time stamp to either healthy or diseases samples whilst focusing on the overall gene expression differences and not considering circadian oscillations [280–283]. Assigning time circadian phases to samples in TCGA allows us to include the temporal aspect when comparing healthy tissues and cancers. In addition, we can also study the time oscillations of mRNAs

across cancers to look for potential pan-cancer rhythmic expression. This newly ordered data allows us to study possible differences in rhythmic activity between TCGA and GTEx data from the same tissues. Thus, we would have three stages per tissue: healthy, healthy neighbouring cancer, and tumorous. This stratification could bring to light a stepwise decline of relevant circadian functions; the loss of rhythmicity in such functions could be included in the diagnosis and prognosis of various cancer types, as proposed for breast cancer [284–286]. Finally, it can also be relevant to investigate, in a patient-specific way, if and how rhythmic functions are correlated with survival rate or response to treatment. However, to advance in many of these directions we first need to have solid metrics to evaluate the state of the circadian clock.

Introduction of metrics

For many of the challenges that lie ahead the introduction of metrics to evaluate either the strength of the clock in a set of samples or the probability that one sample has a working clock seems paramount. Some metrics have been proposed, but they lack solid mathematical foundations as they are mostly based on empirical observations [84, 287]. The cSVD could be exploited to introduce a metric on collective clock robustness, as it recapitulates well the oscillatory properties of clock reference genes. The sample-specific metric is more complicated to introduce. In fact, for this kind of metric a reference with which to compare our sample of interest is needed. One option would be to use a probabilistic projection onto a reference clock behaviour, possibly in a tissue specific way, exploiting the ordered GTEx data. This would mean to evaluate the probability that a sample is on the high dimensional trajectory of the clock, thus separating (in a probabilistic way) samples with a functioning clock from samples with a broken one.

7.2 Timed treatments

We explore possible future studies, spurring from the findings in chapter 4, related to the need of introducing time in medical, clinical, and pharmaceutical research.

Fatty liver diseases

Behaviour of non alcoholic fatty liver diseases has recently been strongly correlated with different states of the clock [288–291]. The role of the clock as a prevention

against such diseases has been proposed [292–294]. On the other hand, malfunction of the clock has been strictly linked with the insurgence of liver pathologies [295–297]. It could be relevant to combine the time labels we inferred with the sample-specific metric described in section 7.1 and histology slices from the donor’s liver. In fact, we could verify if also in humans liver cell size shows circadian oscillations, as it does in mice ?? In addition, we could infer novel properties and hopefully treatment strategies for liver diseases. Lastly, combining the GTEx data with the TCGA atlas could provide further knowledge on which and how circadian controls breaks in hepatocellular carcinoma [298]. This could provide new important insight on the interplay of the circadian clock and liver diseases.

Sex-dimorphism in liver diseases

Non alcoholic fatty liver disease have also been strongly characterized as sex-dimorphic[299–304]. In fact, the incidence is significantly higher in males than females across world regions and races [305–313]. The striking observation, from the analysis of the GTEx data, that rhythmic liver gene expression was significantly more prevalent in females may bring new insights into the staggering sexual dimorphism of fatty liver diseases. While it would need to be established that mRNA rhythms propagate to rhythmic enzymatic activities, we may speculate that enhanced rhythmicity in metabolic pathways may be an attenuating risk factor for such diseases. The interplay of sex-dimorphic liver mRNA rhythms, disrupted clock, and liver diseases should be investigated more in depth, possibly exploiting alongside GTEx other available datasets of RNA-seq from patient with various degrees of fatty liver diseases, possibly like [314–318]. Lastly, the pervasive differential gene expression rhythmicity, in particular in xenobiotic detoxification genes, may also provide opportunities to develop sex-specific chronopharmacology [260].

Cardiovascular diseases in ageing

Cardiovascular diseases show known age-dependent incidence rates, in fact enough to characterise age as a direct risk factor [319, 320]. In addition, age has also been indirectly correlated with cardiovascular disease risk trough the inflammation [321], menopause [322], and overall changes in health [323]. The loss of nearly 800 mRNA rhythms in coronary arteries in the older age group could be exploited to discover possible correlations of loss of circadian function and increased risk of cardiovascular disease. The strong difference in rhythmic processes could also be relevant to explain the uneven timing of sudden cardiac death across ages

and possibly open new avenues for prevention [324] and effectiveness of timed medication [325]. Overall, a study focused on the cardiovascular tissues is needed to understand in detail the interplay of circadian rhythms and cardiovascular disease incidence across ages and sexes [326, 327].

Oscillations of *SERPINE1*

From our analysis Serpin Family E Member 1 (*SERPINE1*) showed the frequent loss of high amplitude circadian rhythms between population subgroups. *SERPINE1* is needed for controlled blood clot degradation [328, 329] and has been related to myocardial infarction [330, 331]. We found it to be arrhythmic in elderly or in male cardiovascular tissues. On the other hand, in females and young its amplitude was in the top 3%. Therefore, we have possibly identified a physiologically relevant circadian process which is highly sex-dimorphic and age dependent. As cardiovascular diseases show lower incidence in young and in females, the precise role of *SERPINE1* oscillations in such diseases should be studied more in depth. Physiological studies on the effect of this gene and especially its oscillations could open unexpected ways for therapy. Hopefully, strategies based on restoring *SERPINE1* rhythms may be attempted. In addition, *SERPINE1* has been proposed as a biomarker for various cancers [332–334] and a null mutation in this gene has been associated with protection against ageing [335]. The circadian direction of these findings has not yet been addressed and makes *SERPINE1* a great candidate for an in-depth study across the body. Such study could pave the way to exploit the different rhythmic behaviour of *SERPINE1* across tissue as both a diagnostic and therapeutic tool for a diverse set of diseases.

Chronopharmacology

Chronopharmacology studies the effect of rhythms on drugs, and suggests that some drugs should be administered at precise times during the day. This concept has been introduced fifty years ago [336–339], but has generally been disregarded by the majority of physicians. Only recently circadian researchers are pushing again for clinical applications of chronopharmacology [340–342]. Our results might open the way to optimize timing and dosing of drugs, notably according to sex and age, both for improving the efficacy or reduce side effects [100, 211, 325]. A screen of the interaction between drug targets and highly rhythmic mRNA could give relevant insights about which treatments should be timed, especially when taking into account the age and sex of a patient.

A Gaussian integration

We remind some known Gaussian integrals, with the help of [166].

A.1 Univariate Gaussian integrals

Here we assume $x, a, b, c \in \mathbf{R}$ and $a > 0$.

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}ax^2 \pm bx + c} = \sqrt{\frac{2\pi}{a}} e^{\frac{b^2}{2a} + c} \quad (\text{A.1})$$

$$\int_{-\infty}^{\infty} x e^{-\frac{1}{2}ax^2 \pm bx + c} = \frac{b}{a} \sqrt{\frac{2\pi}{a}} e^{\frac{b^2}{2a} + c} \quad (\text{A.2})$$

$$\int_{-\infty}^{\infty} x^2 e^{-\frac{1}{2}ax^2 \pm bx + c} = \frac{(b^2 + a)}{a^2} \sqrt{\frac{2\pi}{a}} e^{\frac{b^2}{2a} + c} \quad (\text{A.3})$$

$$\int_{-\infty}^{\infty} f(x) e^{-\frac{1}{2}ax^2 \pm bx + c} = \left(e^{\frac{b}{a} \frac{\partial}{\partial x} + \frac{1}{2a} \frac{\partial}{\partial x} \frac{\partial}{\partial x}} f(x) \right) \Big|_0 \sqrt{\frac{2\pi}{a}} e^{\frac{b^2}{2a} + c} \quad (\text{A.4})$$

A.2 Multivariate gaussian integrals

Here we assume $X, B \in \mathbb{R}^N$ and $A \in \mathbb{R}^N \times \mathbb{R}^N$, $C \in \mathbb{R}$ and that the matrix A is positive definite, although these constraints can be relaxed¹.

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}X^T A X \pm B X + C} = \frac{(2\pi)^{\frac{N}{2}}}{\det(A)} e^{\frac{B^T A^{-1} B}{2} + C} \quad (\text{A.5})$$

$$\int_{-\infty}^{\infty} X e^{-\frac{1}{2}X^T A X \pm B X + C} = A^{-1} B \frac{(2\pi)^{\frac{N}{2}}}{\det(A)} e^{\frac{B^T A^{-1} B}{2} + C} \quad (\text{A.6})$$

¹Lecture on probability, Giorgio Parisi

Appendix A. Gaussian integration

$$\int_{-\infty}^{\infty} X^T X e^{-\frac{1}{2} X^T A X + B X + C} = \left(B^T (A^{-1})^2 B + A^{-1} \right) \frac{(2\pi)^{\frac{N}{2}}}{\det(A)} e^{\frac{B^T A^{-1} B}{2} + C} \quad (\text{A.7})$$

$$\int_{-\infty}^{\infty} f(X) e^{-\frac{1}{2} X^T A X + B X + C} = \left(e^{B_i A_{ij}^{-1} \frac{\partial}{\partial x_j} + \frac{1}{2} A_{ij}^{-1} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j}} f(X) \right) \Big|_0 \frac{(2\pi)^{\frac{N}{2}}}{\det(A)} e^{\frac{B^T A^{-1} B}{2} + C} \quad (\text{A.8})$$

where x_i is the i th component of the vector X .

A.3 Particular case

We develop more in detail this calculation as it is useful in our derivation of the marginalized probability in chapter 5.

$$\begin{aligned} & \int_{-\infty}^{\infty} e^{-\frac{\sum_i (y_i - \mu)^2}{2\sigma^2}} e^{-\frac{(\mu - \hat{\mu})^2}{2\tau^2}} d\mu = \\ & \int_{-\infty}^{\infty} e^{-\frac{1}{2} \mu^2 \left(\frac{|I|}{\sigma^2} + \frac{1}{\tau^2} \right) + \mu \left(\frac{\sum_i y_i}{\sigma^2} + \frac{\hat{\mu}}{\tau^2} \right) - \left(\frac{\sum_i y_i^2}{2\sigma^2} + \frac{\hat{\mu}^2}{2\tau^2} \right)} d\mu = \\ & \sqrt{\frac{2\pi}{\frac{|I|}{\sigma^2} + \frac{1}{\tau^2}}} e^{\frac{\left(\frac{\sum_i y_i}{\sigma^2} + \frac{\hat{\mu}}{\tau^2} \right)^2}{2 \left(\frac{|I|}{\sigma^2} + \frac{1}{\tau^2} \right)} - \left(\frac{\sum_i y_i^2}{2\sigma^2} + \frac{\hat{\mu}^2}{2\tau^2} \right)} \end{aligned} \quad (\text{A.9})$$

We would like to rewrite the exponent of A.9 in a more intuitive way:

$$\frac{\left(\frac{\sum_i y_i}{\sigma^2} + \frac{\hat{\mu}}{\tau^2} \right)^2}{2 \left(\frac{|I|}{\sigma^2} + \frac{1}{\tau^2} \right)} - \left(\frac{\sum_i y_i^2}{2\sigma^2} + \frac{\hat{\mu}^2}{2\tau^2} \right) = \frac{a}{2} \sum_i (y_i - \bar{y})^2 \quad (\text{A.10})$$

for some a and \bar{y} . In particular we would like to have $\bar{y} = \hat{\mu}$ where, however, $\hat{\mu} = f(\sum_i y_i)$ so thngs might get a bit more tricky. Let's get into the calculation, keeping

in mind that we are not interested in the constant term.

$$\begin{aligned}
& \frac{\left(\frac{\sum_i y_i}{\sigma^2} + \frac{\bar{\mu}}{\tau^2}\right)^2}{2\left(\frac{|I|}{\sigma^2} + \frac{1}{\tau^2}\right)} - \left(\frac{\sum_i y_i^2}{2\sigma^2} + \frac{\bar{\mu}^2}{2\tau^2}\right) = \\
& \frac{1}{2\left(\frac{|I|}{\sigma^2} + \frac{1}{\tau^2}\right)} \left(\frac{\sum_i y_i}{\sigma^2} + \frac{\bar{\mu}}{\tau^2}\right) \left(\frac{\sum_i y_i}{\sigma^2} + \frac{\bar{\mu}}{\tau^2}\right) - \left(\frac{\sum_i y_i^2}{2\sigma^2} + \frac{\bar{\mu}^2}{2\tau^2}\right) = \\
& - \sum_i \left(\frac{1}{2\sigma^2} y_i^2 + \frac{1}{\sigma^2} y_i \frac{1}{\left(\frac{|I|}{\sigma^2} + \frac{1}{\tau^2}\right)} \left(\frac{\sum_j y_j}{\sigma^2} + \frac{\bar{\mu}}{\tau^2}\right) \right) + c \rightarrow \\
& - \frac{1}{2\sigma^2} \sum_i \left(y_i - \frac{1}{\left(\frac{|I|}{\sigma^2} + \frac{1}{\tau^2}\right)} \left(\frac{\sum_j y_j}{\sigma^2} + \frac{\bar{\mu}}{\tau^2}\right) \right)^2 = \\
& - \frac{1}{2\sigma^2} \sum_i (y_i - \hat{\mu})^2
\end{aligned} \tag{A.11}$$

where of course:

$$\hat{\mu} = \frac{1}{\left(\frac{|I|}{\sigma^2} + \frac{1}{\tau^2}\right)} \left(\frac{\sum_j y_j}{\sigma^2} + \frac{\bar{\mu}}{\tau^2}\right) = \frac{\frac{|I|}{\sigma^2} \langle y \rangle + \frac{1}{\tau^2} \bar{\mu}}{\frac{|I|}{\sigma^2} + \frac{1}{\tau^2}} \tag{A.12}$$

which is only a weighted mean of the data and prior contributions, where the weights are the variances scaled by the observations. We then write concisely:

$$\int_{-\infty}^{\infty} e^{-\frac{\sum_i (y_i - \mu)^2}{2\sigma^2}} e^{-\frac{(\mu - \bar{\mu})^2}{2\tau^2}} d\mu = \sqrt{\frac{2\pi}{\frac{|I|}{\sigma^2} + \frac{1}{\tau^2}}} e^{-\frac{1}{2\sigma^2} \sum_i (y_i - \hat{\mu})^2} \tag{A.13}$$

B Technical details for the improvement of RNA velocity

Here we collect and discuss a series of small technical details that are however crucial to a correct interpretation of the system and its numerical implementation.

B.1 Biological constraints β and γ

The biological parameters governing the dynamics of spliced mRNA, S are β and γ . For this reason their values are highly constrained by the biology. In particular, for genes oscillating during the cell cycle:

$$\begin{aligned}\beta_0 &= \bar{\beta} = 1 \\ \beta_0^{-1} &\in [10, 20] \text{ minutes, estimate is unsure due to technical artefacts} \\ \gamma_0^{-1} &\in [0.5, 1.5] \text{ hours} \\ 2\pi\langle\omega_0^{-1}\rangle &\in [6, 50] \text{ hours} \sim 12\gamma_0^{-1} \\ \gamma_0 &< \beta_0 \quad \gamma_0 \sim \frac{1}{4}\beta_0 \\ \omega_0 &\sim \frac{2\pi}{12\gamma_0^{-1}} \sim \frac{\gamma_0}{2} \sim \frac{\beta_0}{8}\end{aligned}\tag{B.1}$$

with γ_0 the (geometric) mean of the γ distribution.

B.2 Velocity and period

One of the main issue is how to interpret the velocities we find. Although we have set a time scale with the $\bar{\beta}$ prior, we have chosen a large $\tau = 2$ parameter that does not effectively constrain the scales of the parameters in play. However, the posterior

distribution of the γ s can be used to infer the scale in hours of the cell cycle period. For this reason, we re-scaled the cell cycle velocity we infer by the geometric mean (or the median) of the inferred γ s, γ_0 . If we have a strong prior for the γ s this is equivalent to setting the median of the prior equal to γ_0 ; however if the prior is uninformative we must use the posterior distribution of γ s across genes. We assume, from literature or experiments, the mean half life of spliced mRNAs to be a hours:

$$\gamma_0^{-1} \simeq ah. \quad (\text{B.2})$$

So, taking advantage of (5.6) we write:

$$T = \int_0^{2\pi} \frac{1}{\omega(\varphi)} d\varphi = \frac{2\pi}{\bar{\omega} \sqrt{1 - |v_\omega|^2}} \quad (\text{B.3})$$

for ω with mean $\bar{\omega}$ and up to one harmonic described by the coefficients v_ω . So, the units of measurement can be included via γ_0 :

$$T = \frac{2\pi}{\bar{\omega} \sqrt{1 - |v_\omega|^2}} = K \gamma_0^{-1} \quad (\text{B.4})$$

which leads to:

$$T[\gamma_0^{-1}] = \frac{2\pi}{\frac{\bar{\omega}}{\gamma_0} \sqrt{1 - |v_\omega|^2}} = K \quad (\text{B.5})$$

and finally

$$T[\text{h}] = \frac{2a\pi}{\frac{\bar{\omega}}{\gamma_0} \sqrt{1 - |v_\omega|^2}}. \quad (\text{B.6})$$

In our case we take:

$$a = \frac{3}{4}. \quad (\text{B.7})$$

B.3 Rotation in the Fourier space

As we know the cell cycle phases of cells are defined apart from a rotation. To have a clearer biological interpretability we would like the phases to be such that $\varphi = 0$ correspond to mitosis.

$$\varphi' = \varphi - \varphi_{\text{mitosis}} = \varphi - \delta. \quad (\text{B.8})$$

So we have in general for gene g after we rotate the phases:

$$\log s_g = v_g \zeta = v'_g \zeta(\varphi') = v'_g \zeta'. \quad (\text{B.9})$$

For each harmonic the rotation is quite straightforward:

$$\begin{aligned} a'_n &= a_n \cos(n\delta) - b_n \sin(n\delta) \\ b'_n &= b_n \cos(n\delta) + a_n \sin(n\delta). \end{aligned} \quad (\text{B.10})$$

However, using the notation $\zeta(\delta) = \zeta_\delta$, we can see how this translates into a matrix multiplication:

$$\mathbf{v}' = \Delta \zeta_\delta \quad (\text{B.11})$$

with

$$\Delta_{ij} = a_{\lfloor \frac{i}{2} \rfloor} \delta_{ij} + b_{\lfloor \frac{i}{2} \rfloor} \left(\delta_{\lfloor \frac{i}{2} \rfloor, \lfloor \frac{j}{2} \rfloor} - \delta_{ij} \right) (-1)^j \quad (\text{B.12})$$

remembering that the notation accounts for the constant mean term, a_0 .

B.4 Constraints on γ

We have approached the system in the "opposite" way. Instead of integrating the levels of u to get the s , we are taking the derivative of s to have a predicted u . This could lead to non-biological situations, such as a predicted negative values of the unspliced counts in the linear space:

$$u_{gc} = \left(\sum_{ff'} v_{gf} D_{ff'} \zeta_{f'c} \tilde{\omega}(\varphi_c) + \tilde{\gamma}_g \right) s_g(\varphi_c) < 0. \quad (\text{B.13})$$

These scenarios can be avoided with correct constraints on γ_g . We notice an important relation:

$$\min(\partial_t s) \geq -\gamma s. \quad (\text{B.14})$$

So in our model it should be included that:

$$\begin{aligned} \min_c \left(\sum_{ff'} v_{gf} D_{ff'} \zeta_{f'c} \tilde{\omega}(\varphi_c) s_g \right) &\geq -\gamma_g s_g \\ \rightarrow \min_c \left(\sum_{ff'} v_{gf} D_{ff'} \zeta_{f'c} \tilde{\omega}(\varphi_c) \right) &\geq -\gamma_g, \end{aligned} \quad (\text{B.15})$$

which clearly means:

$$\min_c \left(\sum_{ff'} v_{gf} D_{ff'} \zeta_{f'c} \tilde{\omega}(\varphi_c) + \gamma_g \right) \geq 0 \quad \rightarrow \quad u \geq 0 \quad \text{as} \quad s \geq 0. \quad (\text{B.16})$$

Appendix B. Technical details for the improvement of RNA velocity

It seems easier to freely chose v and then constrain γ ; so we calculate:

$$\begin{aligned} \min_c \left(\sum_{ff'} v_{gf} D_{ff'} \zeta_{f'c} \tilde{\omega}(\varphi_c) \right) &\geq \\ \min_{\varphi} \left(\sum_{ff'} v_{gf} D_{ff'} \zeta_{f'c} \tilde{\omega}(\varphi) \right) &\geq \\ -\max_{\varphi}(\omega(\varphi)) \sum_{f \text{ odd}} \left(\frac{f+1}{2} \sqrt{v_{g,f}^2 + v_{g,f+1}^2} \right) \end{aligned} \quad (\text{B.17})$$

thus:

$$\gamma_g \geq \max_{\varphi}(\omega(\varphi)) \sum_{f \text{ odd}} \left(\frac{f+1}{2} \sqrt{v_{g,f}^2 + v_{g,f+1}^2} \right). \quad (\text{B.18})$$

B.5 Pyro-accepted probability distributions

We want to explicitly write $\mathcal{P}(U|u)$ as a function of its location and precision matrix in order to include it in the pyro framework. Let us remind the object in question:

$$\mathcal{P}(U|u) = \prod_g e^{-\frac{1}{2\sigma^2} \Sigma_c \left(\ln \left(\frac{u_{gc}}{U_{gc}} \right) - \langle \ln(\tilde{\beta}_g) \rangle \right)^2} d\vec{U}_g. \quad (\text{B.19})$$

We simplify the notation to not carry useless weights. We start by dropping the gene index, as it would only burden us, and define the vectors l and L as:

$$l = \{l_c\} = \{\ln(u_c)\} = \overrightarrow{\ln(\vec{u})}, \quad L = \{L_c\} = \{\ln(U_c)\} = \overrightarrow{\ln(U)} \quad (\text{B.20})$$

So we can rewrite (B.19) using (5.39) as

$$\mathcal{P}(U|u) = e^{-\frac{1}{2\sigma^2} \Sigma_c (l_c - L_c - \kappa \Sigma_{c'} (l_{c'} - L_{c'}))^2} d\vec{U}. \quad (\text{B.21})$$

with

$$\kappa^{-1} = \sigma^2 \left(\frac{|C|}{\sigma^2} + \frac{1}{\tau^2} \right). \quad (\text{B.22})$$

The last ingredient before integrating is a change of variable for the differential

$$d\vec{U} = \prod_c dU_c = \prod_c dU_c \frac{\prod_c U_c}{\prod_c U_c} = \prod_c \frac{dU_c}{U_c} \prod_c U_c = dL \prod_c U_c = e^{\Sigma_c L_c} dL. \quad (\text{B.23})$$

Finally, we write

$$\mathcal{P}(U|u) = e^{-\frac{1}{2\sigma^2} \Sigma_c (l_c - L_c - \kappa \Sigma_{c'} (l_{c'} - L_{c'}))^2} e^{\Sigma_c L_c} dL. \quad (\text{B.24})$$

We have a closer look and manipulate our exponential

$$\begin{aligned}
 & \sum_c \left(l_c - L_c - \kappa \sum_{c'} (l_{c'} - L_{c'}) \right)^2 = \\
 & \sum_c \left(l_c^2 + L_c^2 + \kappa^2 \sum_{c'} \sum_{c''} (l_{c'} - L_{c'}) (l_{c''} - L_{c''}) + \right. \\
 & \quad \left. - 2l_c L_c - 2l_c \kappa \sum_{c'} (l_{c'} - L_{c'}) + 2L_c \kappa \sum_{c'} (l_{c'} - L_{c'}) \right) = \\
 & lll + Lll + |C|\kappa^2 (l1l + L1L - 2l1L) + \\
 & - 2lil + 4\kappa l1L - 2\kappa l1l - 2\kappa L1L = \\
 & L(I + |C|\kappa^2 1 - 2\kappa 1)L + \\
 & l(-2|C|\kappa^2 1 - 2I + 4\kappa 1)L + \\
 & l(I + |C|\kappa^2 1 - 2\kappa 1)l = \\
 & \sigma^2(LAL + -2lAL + lAl)
 \end{aligned} \tag{B.25}$$

with $A = (I - 2\kappa 1 + |C|\kappa^2 1)/\sigma^2$, I the identity matrix, and 1 a matrix or vector of entries always equal to one, in $|C|$ dimensions. Definig the vector $B = -lA + 1$ we can quite easily write (B.19) and obtain

$$\mathcal{P}(U|u) \sim e^{-\frac{1}{2}LAL + BL} dL \tag{B.26}$$

following the definitions above. Thus the A and B parameters are the correct parametrization for pyro to deal with this probability.

B.6 Scales across datasets

After applying all this machinery to real data we found that sometimes data doesn't behave exactly as expected. In particular, some genes are constitutively more or less expressed, although the oscillations are quite similar. In our framework this would mean a "dataset" specific v_{g0} , or discarding the genes with this behaviour. In general, changing the v s between datasets would mean obtaining non-comparable velocities after the optimization. However, if we remember that the only way in which the v s enter the optimization is through \bar{u}_{gc} . We rewrite the definition of \bar{u}_{gc} :

$$\bar{u}_g = u_g(\beta_g = \bar{\beta}) = \frac{\beta_g}{\bar{\beta}} u_g = \left(\sum_{ff'} v_{gf} D_{ff'} \zeta_{f'c} \tilde{\omega}(\varphi_c) + \tilde{\gamma}_g \right) s_g(\varphi_c) \tag{B.27}$$

Appendix B. Technical details for the improvement of RNA velocity

and of the differential operator $D_{ff'}$:

$$D_{ff'} = \left(\delta_{\left\lceil \frac{f}{2} \right\rceil, \left\lceil \frac{f'}{2} \right\rceil} - \delta_{f, f'} \right) \left\lceil \frac{f}{2} \right\rceil (-1)^f \quad (\text{B.28})$$

and notice that

$$D_{0i} = D_{i0} = 0 \quad \forall i. \quad (\text{B.29})$$

This directly means that that v_{g0} does not influence the inference, and is just an adjustemnt for the fits of the exon reads. Therefore, we can easily adjust the mean of the gene fits given by the v s to match the right one. In particular, we can do this before even inferring the phases for a new dataset. This batch correction can improve the overall phase estimate and allow for the inference of $\{v\}$ simultaneously across dataset.

Bibliography

- [1] Barry N Taylor. The International System of Units (SI). page 77.
- [2] C. S. Pittendrigh. Circadian Rhythms and the Circadian Organization of Living Systems. *Cold Spring Harbor Symposia on Quantitative Biology*, 25(0):159–184, January 1960.
- [3] Leon Kreitzman and Russell Foster. *The Rhythms Of Life: The Biological Clocks That Control the Daily Lives of Every Living Thing*. Profile Books, September 2011. Google-Books-ID: XIUyNPDMY8sC.
- [4] R. J. Konopka and S. Benzer. Clock mutants of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*, 68(9):2112–2116, September 1971.
- [5] Ethan D. Buhr and Joseph S. Takahashi. Molecular Components of the Mammalian Circadian Clock. In Achim Kramer and Martha Merrow, editors, *Circadian Clocks*, Handbook of Experimental Pharmacology, pages 3–27. Springer, Berlin, Heidelberg, 2013.
- [6] Hugues Dardente and Nicolas Cermakian. Molecular Circadian Rhythms in Central and Peripheral Clocks in Mammals. *Chronobiology International*, 24(2):195–213, January 2007.
- [7] Urs Albrecht and Gregor Eichele. The mammalian circadian clock. *Current Opinion in Genetics & Development*, 13(3):271–277, June 2003. Number: 3 ZSCC: 0000287.
- [8] L. P. Shearman. Interacting Molecular Loops in the Mammalian Circadian Clock. *Science*, 288(5468):1013–1019, May 2000.
- [9] Amita Sehgal. *Molecular Biology of Circadian Rhythms*. John Wiley & Sons, April 2004. Google-Books-ID: cny7RTlODiAC.

Bibliography

- [10] Caroline H. Ko and Joseph S. Takahashi. Molecular components of the mammalian circadian clock. *Human Molecular Genetics*, 15(suppl_2):R271–R277, October 2006.
- [11] Jay C. Dunlap. Molecular Bases for Circadian Clocks. *Cell*, 96(2):271–290, January 1999.
- [12] Carrie L. Partch, Carla B. Green, and Joseph S. Takahashi. Molecular architecture of the mammalian circadian clock. *Trends in Cell Biology*, 24(2):90–99, February 2014.
- [13] R. Allada, P. Emery, J. S. Takahashi, and M. Rosbash. Stopping time: the genetics of fly and mouse circadian clocks. *Annual Review of Neuroscience*, 24:1091–1119, 2001.
- [14] Joseph Bass. Circadian topology of metabolism. *Nature*, 491(7424):348–356, November 2012. Number: 7424.
- [15] Forest Agostinelli, Nicholas Ceglia, Babak Shahbaba, Paolo Sassone-Corsi, and Pierre Baldi. What time is it? Deep learning approaches for circadian rhythms. *Bioinformatics*, 32(12):i8–i17, June 2016. ZSCC: 0000030.
- [16] Till Roenneberg, Anna Wirz-Justice, and Martha Merrow. Life between Clocks: Daily Temporal Patterns of Human Chronotypes:. *Journal of Biological Rhythms*, June 2016. Publisher: SAGE Publications.
- [17] T. Roenneberg and M. Merrow. Entrainment of the Human Circadian Clock. *Cold Spring Harbor Symposia on Quantitative Biology*, 72(1):293–299, January 2007.
- [18] Ray Zhang, Nicholas F. Lahens, Heather I. Ballance, Michael E. Hughes, and John 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, November 2014. Number: 45 ZSCC: 0000967.
- [19] Jake Yeung, Jérôme Mermet, Céline Jouffe, Julien Marquis, Aline Charpagne, Frédéric Gachon, and Felix Naef. Transcription factor activity rhythms and tissue-specific chromatin interactions explain circadian gene expression across organs. *Genome Research*, 28(2):182–191, February 2018. Number: 2 ZSCC: 0000030.
- [20] Ludovic S. Mure, Hiep D. Le, Giorgia Benegiamo, Max W. Chang, Luis Rios, Ngalla Jillani, Maina Ngotho, Thomas Kariuki, Ouria Dkhissi-Benyahya,

- Howard M. Cooper, and Satchidananda Panda. Diurnal transcriptome atlas of a primate across major neural and peripheral tissues. *Science*, 359(6381), March 2018. ZSCC: 0000149 Publisher: American Association for the Advancement of Science Section: Research Article.
- [21] Karen Wager-Smith and Steve A. Kay. Circadian rhythm genetics: from flies to mice to humans. *Nature Genetics*, 26(1):23–27, September 2000. Number: 1 Publisher: Nature Publishing Group.
- [22] Ron C. Anafi, Yool Lee, Trey K. Sato, Anand Venkataraman, Chidambaram Ramanathan, Ibrahim H. Kavakli, Michael E. Hughes, Julie E. Baggs, Jacqueline Growe, Andrew C. Liu, Junhyong Kim, and John B. Hogenesch. Machine Learning Helps Identify CHRONO as a Circadian Clock Component. *PLoS Biology*, 12(4):e1001840, April 2014.
- [23] Jacob J. Hughey, Trevor Hastie, and Atul J. Butte. ZeitZeiger: supervised learning for high-dimensional data from an oscillatory system. *Nucleic Acids Research*, 44(8):e80–e80, May 2016. Number: 8 ZSCC: 0000043.
- [24] Jacob J. Hughey. Machine learning identifies a compact gene set for monitoring the circadian clock in human blood. *Genome Medicine*, 9(1):19, February 2017. Number: 1 ZSCC: 0000025.
- [25] Ron C. Anafi, Lauren J. Francey, John B. Hogenesch, and Junhyong Kim. CYCLOPS reveals human transcriptional rhythms in health and disease. *Proceedings of the National Academy of Sciences*, 114(20):5312–5317, May 2017. Number: 20 ZSCC: 0000069.
- [26] Rosemary Braun, William L. Kath, Marta Iwanaszko, Elzbieta Kula-Eversole, Sabra M. Abbott, Kathryn J. Reid, Phyllis C. Zee, and Ravi Allada. Universal method for robust detection of circadian state from gene expression. *Proceedings of the National Academy of Sciences*, 115(39):E9247–E9256, September 2018. Number: 39.
- [27] Denise Vlachou, Georg A. Bjarnason, Sylvie Giacchetti, Francis Lévi, and David A. Rand. TimeTeller: a New Tool for Precision Circadian Medicine and Cancer Prognosis. *bioRxiv*, page 622050, April 2019.
- [28] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, 1933. Place: US Publisher: Warwick & York.

Bibliography

- [29] Charna Dibner, Ueli Schibler, and Urs Albrecht. The Mammalian Circadian Timing System: Organization and Coordination of Central and Peripheral Clocks. *Annual Review of Physiology*, 72(1):517–549, 2010. Number: 1.
- [30] D. C. Klein, Robert Y. Moore, and Steven M. Reppert, editors. *Suprachiasmatic nucleus: the mind's clock*. Oxford University Press, New York, 1991.
- [31] Charles A. Czeisler, Jeanne F. Duffy, Theresa L. Shanahan, Emery N. Brown, Jude F. Mitchell, David W. Rimmer, Joseph M. Ronda, Edward J. Silva, James S. Allan, Jonathan S. Emens, Derk-Jan Dijk, and Richard E. Kronauer. Stability, Precision, and Near-24-Hour Period of the Human Circadian Pacemaker. *Science*, 284(5423):2177–2181, June 1999.
- [32] Jürgen Aschoff. Freerunning and Entrained Circadian Rhythms. In Jürgen Aschoff, editor, *Biological Rhythms*, pages 81–93. Springer US, Boston, MA, 1981.
- [33] Diego A. Golombek and Ruth E. Rosenstein. Physiology of Circadian Entrainment. *Physiological Reviews*, 90(3):1063–1102, July 2010. Publisher: American Physiological Society.
- [34] Robert Y. Moore. Chapter 8 Entrainment pathways and the functional organization of the circadian system. In R. M. Buijs, A. Kalsbeek, H. J. Romijn, C. M. A. Pennartz, and M. Mirmiran, editors, *Progress in Brain Research*, volume 111, pages 103–119. Elsevier, January 1996.
- [35] Jeanne F. Duffy and Kenneth P. Wright. Entrainment of the Human Circadian System by Light. *Journal of Biological Rhythms*, 20(4):326–338, August 2005. Publisher: SAGE Publications Inc.
- [36] Samuel E. Jones, Jacqueline M. Lane, Andrew R. Wood, Vincent T. van Hees, Jessica Tyrrell, Robin N. Beaumont, Aaron R. Jeffries, Hassan S. Dashti, Melvyn Hillsdon, Katherine S. Ruth, Marcus A. Tuke, Hanieh Yaghootkar, Seth A. Sharp, Yingjie Jie, William D. Thompson, Jamie W. Harrison, Amy Dawes, Enda M. Byrne, Henning Tiemeier, Karla V. Allebrandt, Jack Bowden, David W. Ray, Rachel M. Freathy, Anna Murray, Diego R. Mazzotti, Philip R. Gehrman, Debbie A. Lawlor, Timothy M. Frayling, Martin K. Rutter, David A. Hinds, Richa Saxena, and Michael N. Weedon. Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. *Nature Communications*, 10(1):1–11, January 2019. ZSCC: NoCitationData[s0] Number: 1 Publisher: Nature Publishing Group.

- [37] Samuel E. Jones, Jessica Tyrrell, Andrew R. Wood, Robin N. Beaumont, Katherine S. Ruth, Marcus A. Tuke, Hanieh Yaghootkar, Youna Hu, Maris Teder-Laving, Caroline Hayward, Till Roenneberg, James F. Wilson, Fabiola Del Greco, Andrew A. Hicks, Chol Shin, Chang-Ho Yun, Seung Ku Lee, Andres Metspalu, Enda M. Byrne, Philip R. Gehrman, Henning Tiemeier, Karla V. Allebrandt, Rachel M. Freathy, Anna Murray, David A. Hinds, Timothy M. Frayling, and Michael N. Weedon. Genome-Wide Association Analyses in 128,266 Individuals Identifies New Morningness and Sleep Duration Loci. *PLoS genetics*, 12(8):e1006125, August 2016. Number: 8 ZSCC: NoCitationData[s0].
- [38] Youna Hu, Alena Shmygelska, David Tran, Nicholas Eriksson, Joyce Y. Tung, and David A. Hinds. GWAS of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. *Nature Communications*, 7(1):10448, February 2016. ZSCC: 0000260 Number: 1 Publisher: Nature Publishing Group.
- [39] Jacqueline M. Lane, Irma Vlasac, Simon G. Anderson, Simon D. Kyle, William G. Dixon, David A. Bechtold, Shubhroz Gill, Max A. Little, Annemarie Luik, Andrew Loudon, Richard Emsley, Frank A. J. L. Scheer, Deborah A. Lawlor, Susan Redline, David W. Ray, Martin K. Rutter, and Richa Saxena. Genome-wide association analysis identifies novel loci for chronotype in 100,420 individuals from the UK Biobank. *Nature Communications*, 7(1):10889, March 2016. ZSCC: NoCitationData[s0] Number: 1 Publisher: Nature Publishing Group.
- [40] Jake Yeung and Felix Naef. Rhythms of the Genome: Circadian Dynamics from Chromatin Topology, Tissue-Specific Gene Expression, to Behavior. *Trends in Genetics*, 34(12):915–926, December 2018.
- [41] Jennifer A. Mohawk, Carla B. Green, and Joseph S. Takahashi. CENTRAL AND PERIPHERAL CIRCADIAN CLOCKS IN MAMMALS. *Annual review of neuroscience*, 35:445–462, 2012.
- [42] Karl-Arne Stokkan, Shin Yamazaki, Hajime Tei, Yoshiyuki Sakaki, and Michael Menaker. Entrainment of the Circadian Clock in the Liver by Feeding. *Science*, 291(5503):490–493, January 2001.
- [43] Florian Atger, Cédric Gobet, Julien Marquis, Eva Martin, Jingkui Wang, Benjamin Weger, Grégory Lefebvre, Patrick Descombes, Felix Naef, and Frédéric 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, November 2015. Number: 47 ZSCC: 0000132.
- [44] Christopher Vollmers, Shubhroz Gill, Luciano DiTacchio, Sandhya R. Pulivarthy, Hiep D. Le, and Satchidananda Panda. Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression. *Proceedings of the National Academy of Sciences*, 106(50):21453–21458, December 2009. Publisher: Proceedings of the National Academy of Sciences.
- [45] Adi Neufeld-Cohen, Maria S. Robles, Rona Aviram, Gal Manella, Yaarit Adamovich, Benjamin Ladeuix, Dana Nir, Liat Rouso-Noori, Yael Kuperman, Marina Golik, Matthias Mann, and Gad Asher. Circadian control of oscillations in mitochondrial rate-limiting enzymes and nutrient utilization by PERIOD proteins. *Proceedings of the National Academy of Sciences*, 113(12):E1673–E1682, March 2016. Publisher: Proceedings of the National Academy of Sciences.
- [46] Yaarit Adamovich, Liat Rouso-Noori, Ziv Zwihaft, Adi Neufeld-Cohen, Marina Golik, Judith Kraut-Cohen, Miao Wang, Xianlin Han, and Gad Asher. Circadian Clocks and Feeding Time Regulate the Oscillations and Levels of Hepatic Triglycerides. *Cell Metabolism*, 19(2):319–330, February 2014.
- [47] Masanobu Kawai, Carla B. Green, Beata Lecka-Czernik, Nicholas Douris, Misty R. Gilbert, Shihoko Kojima, Cheryl Ackert-Bicknell, Neha Garg, Mark C. Horowitz, Martin L. Adamo, David R. Clemmons, and Clifford J. Rosen. A circadian-regulated gene, Nocturnin, promotes adipogenesis by stimulating PPAR- nuclear translocation. *Proceedings of the National Academy of Sciences*, 107(23):10508–10513, June 2010. Publisher: Proceedings of the National Academy of Sciences.
- [48] Kathryn Moynihan Ramsey and Joseph Bass. Lean gene and the clock machine. *Proceedings of the National Academy of Sciences*, 104(23):9553–9554, June 2007. Publisher: Proceedings of the National Academy of Sciences.
- [49] Masanobu Kawai, Anne M. Delany, Carla B. Green, Martin L. Adamo, and Clifford J. Rosen. Nocturnin Suppresses Igf1 Expression in Bone by Targeting the 3 Untranslated Region of Igf1 mRNA. *Endocrinology*, 151(10):4861–4870, October 2010.
- [50] Takashi Kudo, Toru Tamagawa, Mihoko Kawashima, Natsuko Mito, and Shigenobu Shibata. Attenuating Effect of Clock Mutation on Triglyceride

- Contents in the ICR Mouse Liver under a High-Fat Diet. *Journal of Biological Rhythms*, 22(4):312–323, August 2007. Publisher: SAGE Publications Inc.
- [51] M. Mahmood Hussain and Xiaoyue Pan. Circadian regulators of intestinal lipid absorption. *Journal of Lipid Research*, 56(4):761–770, April 2015. Publisher: Elsevier.
- [52] M. Mahmood Hussain and Xiaoyue Pan. Circadian Regulation of Macronutrient Absorption. *Journal of Biological Rhythms*, 30(6):459–469, December 2015. Publisher: SAGE Publications Inc.
- [53] Krista Kaasik, Saul Kivimäe, Jasmina J. Allen, Robert J. Chalkley, Yong Huang, Kristin Baer, Holger Kissel, Alma L. Burlingame, Kevan M. Shokat, Louis J. Ptáček, and Ying-Hui Fu. Glucose Sensor O-GlcNAcylation Coordinates with Phosphorylation to Regulate Circadian Clock. *Cell Metabolism*, 17(2):291–302, February 2013. Publisher: Elsevier.
- [54] Maria S. Robles, Jürgen Cox, and Matthias Mann. In-Vivo Quantitative Proteomics Reveals a Key Contribution of Post-Transcriptional Mechanisms to the Circadian Regulation of Liver Metabolism. *PLOS Genetics*, 10(1):e1004047, January 2014. Publisher: Public Library of Science.
- [55] Min-Dian Li, Hai-Bin Ruan, Michael E. Hughes, Jeong-Sang Lee, Jay P. Singh, Steven P. Jones, Michael N. Nitabach, and Xiaoyong Yang. O-GlcNAc Signaling Entrain the Circadian Clock by Inhibiting BMAL1/CLOCK Ubiquitination. *Cell Metabolism*, 17(2):303–310, February 2013.
- [56] Eric E. Zhang, Yi Liu, Renaud Dentin, Pagkapol Y. Pongsawakul, Andrew C. Liu, Tsuyoshi Hirota, Dmitri A. Nusinow, Xiujie Sun, Severine Landais, Yuzo Kodama, David A. Brenner, Marc Montminy, and Steve A. Kay. Cryptochrome mediates circadian regulation of cAMP signaling and hepatic gluconeogenesis. *Nature Medicine*, 16(10):1152–1156, October 2010. Number: 10 Publisher: Nature Publishing Group.
- [57] Rajesh Narasimamurthy, Megumi Hatori, Surendra K. Nayak, Fei Liu, Satchidananda Panda, and Inder M. Verma. Circadian clock protein cryptochrome regulates the expression of proinflammatory cytokines. *Proceedings of the National Academy of Sciences*, 109(31):12662–12667, July 2012. Publisher: Proceedings of the National Academy of Sciences.
- [58] Biliana Marcheva, Kathryn Moynihan Ramsey, Ethan D. Buhr, Yumiko Kobayashi, Hong Su, Caroline H. Ko, Ganka Ivanova, Chiaki Omura, Shelley

- Mo, Martha H. Vitaterna, James P. Lopez, Louis H. Philipson, Christopher A. Bradfield, Seth D. Crosby, Lellean JeBailey, Xiaozhong Wang, Joseph S. Takahashi, and Joseph Bass. Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes. *Nature*, 466(7306):627–631, July 2010. Number: 7306 Publisher: Nature Publishing Group.
- [59] Kenneth S. Polonsky, Bruce D. Given, Laurence J. Hirsch, Hartmut Tillil, E. Timothy Shapiro, Christine Beebe, Bruce H. Frank, John A. Galloway, and Eve Van Cauter. Abnormal Patterns of Insulin Secretion in Non-Insulin-Dependent Diabetes Mellitus. *New England Journal of Medicine*, 318(19):1231–1239, May 1988. Publisher: Massachusetts Medical Society _eprint: <https://doi.org/10.1056/NEJM198805123181903>.
- [60] Andries Kalsbeek, Ewout Foppen, Ingrid Schali, Caroline Van Heijningen, Jan van der Vliet, Eric Fliers, and Ruud M. Buijs. Circadian Control of the Daily Plasma Glucose Rhythm: An Interplay of GABA and Glutamate. *PLOS ONE*, 3(9):e3194, September 2008. Publisher: Public Library of Science.
- [61] E. Peschke and D. Peschke. Evidence for a circadian rhythm of insulin release from perfused rat pancreatic islets. *Diabetologia*, 41(9):1085–1092, August 1998.
- [62] Diana E. Ayala, Ramón C. Hermida, Luisa Chayán, Artemio Mojón, María J. Fontao, and José R. Fernández. Circadian Pattern of Ambulatory Blood Pressure in Untreated Hypertensive Patients with and Without Metabolic Syndrome. *Chronobiology International*, 26(6):1189–1205, August 2009. Publisher: Taylor & Francis _eprint: <https://doi.org/10.3109/07420520903206294>.
- [63] Anne M. Curtis, Yan Cheng, Shiv Kapoor, Dermot Reilly, Tom S. Price, and Garret A. FitzGerald. Circadian variation of blood pressure and the vascular response to asynchronous stress. *Proceedings of the National Academy of Sciences*, 104(9):3450–3455, February 2007. Publisher: Proceedings of the National Academy of Sciences.
- [64] Elizabeth J. Westgate, Yan Cheng, Dermot F. Reilly, Tom S. Price, Jacqueline A. Walisser, Christopher A. Bradfield, and Garret A. FitzGerald. Genetic Components of the Circadian Clock Regulate Thrombogenesis In Vivo. *Circulation*, 117(16):2087–2095, April 2008. Publisher: American Heart Association.
- [65] Martial M. Massin, Krystel Maeyns, Nadia Withofs, Françoise Ravet, and Paul Gérard. Circadian rhythm of heart rate and heart rate variability. *Archives of*

- Disease in Childhood*, 83(2):179–182, August 2000. Publisher: BMJ Publishing Group Ltd Section: Article.
- [66] Ram B Singh, Germaine Cornélissen, Andi Weydahl, Othild Schwartzkopff, George Katinas, Kuniaki Otsuka, Yoshihiko Watanabe, Shoki Yano, Hideki Mori, Yuhei Ichimaru, Gen Mitsutake, Daniel Pella, Lu Fanghong, Ziyang Zhao, Reema S Rao, Anna Gvozdjakova, and Franz Halberg. Circadian heart rate and blood pressure variability considered for research and patient care. *International Journal of Cardiology*, 87(1):9–28, January 2003.
- [67] Simon C Malpas and Gordon L Purdie. Circadian variation of heart rate variability. *Cardiovascular Research*, 24(3):210–213, March 1990.
- [68] John A. Schoenhard, Layton H. Smith, Corrie A. Painter, Mesut Eren, Carl H. Johnson, and Douglas E. Vaughan. Regulation of the PAI-1 promoter by circadian clock components: differential activation by BMAL1 and BMAL2. *Journal of Molecular and Cellular Cardiology*, 35(5):473–481, May 2003.
- [69] Julie E. Baggs, Tom S. Price, Luciano DiTacchio, Satchidananda Panda, Garrett A. FitzGerald, and John B. Hogenesch. Network Features of the Mammalian Circadian Clock. *PLOS Biology*, 7(3):e1000052, March 2009. Publisher: Public Library of Science.
- [70] E. Challet. Keeping circadian time with hormones. *Diabetes, Obesity and Metabolism*, 17(S1):76–83, 2015. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/dom.12516>.
- [71] Jacob Richards and Michelle L. Gumz. Advances in understanding the peripheral circadian clocks. *The FASEB Journal*, 26(9):3602–3613, 2012. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1096/fj.12-203554>.
- [72] Urs Albrecht. Timing to Perfection: The Biology of Central and Peripheral Circadian Clocks. *Neuron*, 74(2):246–260, April 2012.
- [73] Yu Tahara and Shigenobu Shibata. Circadian rhythms of liver physiology and disease: experimental and clinical evidence. *Nature Reviews Gastroenterology & Hepatology*, 13(4):217–226, April 2016. Number: 4 Publisher: Nature Publishing Group.
- [74] Dmitri Firsov and Olivier Bonny. Circadian rhythms and the kidney. *Nature Reviews Nephrology*, 14(10):626–635, October 2018. Number: 10 Publisher: Nature Publishing Group.

Bibliography

- [75] Ryan W. Logan and Colleen A. McClung. Rhythms of life: circadian disruption and brain disorders across the lifespan. *Nature Reviews Neuroscience*, 20(1):49–65, January 2019. Number: 1 Publisher: Nature Publishing Group.
- [76] Sandra Crnko, Bastiaan C. Du Pré, Joost P. G. Sluijter, and Linda W. Van Laake. Circadian rhythms and the molecular clock in cardiovascular biology and disease. *Nature Reviews Cardiology*, 16(7):437–447, July 2019. Number: 7 Publisher: Nature Publishing Group.
- [77] Gabriele Sulli, Michael Tun Yin Lam, and Satchidananda Panda. Interplay between Circadian Clock and Cancer: New Frontiers for Cancer Treatment. *Trends in Cancer*, 5(8):475–494, August 2019.
- [78] Gabriele Sulli, Emily N. C. Manoogian, Pam R. Taub, and Satchidananda Panda. Training the Circadian Clock, Clocking the Drugs, and Drugging the Clock to Prevent, Manage, and Treat Chronic Diseases. *Trends in Pharmacological Sciences*, 39(9):812–827, September 2018.
- [79] M. C. Mormont and F. Lévi. Circadian-system alterations during cancer processes: A review. *International Journal of Cancer*, 70(2):241–247, 1997.
- [80] Loning Fu and Cheng Chi Lee. The circadian clock: pacemaker and tumour suppressor. *Nature Reviews Cancer*, 3(5):350–361, May 2003. ZSCC: 0000728 Number: 5 Publisher: Nature Publishing Group.
- [81] Amandine Verlande and Selma Masri. Circadian Clocks and Cancer: Time-keeping Governs Cellular Metabolism. *Trends in Endocrinology & Metabolism*, 30(7):445–458, July 2019.
- [82] Ayesha A. Shafi and Karen E. Knudsen. Cancer and the Circadian Clock. *Cancer Research*, 79(15):3806–3814, August 2019. Publisher: American Association for Cancer Research Section: Reviews.
- [83] Fabiola Hernández-Rosas, Carlos Alberto López-Rosas, and Margarita Virginia Saavedra-Vélez. Disruption of the Molecular Circadian Clock and Cancer: An Epigenetic Link. *Biochemical Genetics*, 58(1):189–209, February 2020.
- [84] Jarrod Shilts, Guanhua Chen, and Jacob J. Hughey. Evidence for widespread dysregulation of circadian clock progression in human cancer. *PeerJ*, 6:e4327, January 2018.

-
- [85] Richard G. Stevens. Light-at-night, circadian disruption and breast cancer: assessment of existing evidence. *International Journal of Epidemiology*, 38(4):963–970, August 2009.
- [86] Sigal Gery, Naoki Komatsu, Lilit Baldjyan, Andrew Yu, Danielle Koo, and H. Phillip Koeffler. The Circadian Gene *Per1* Plays an Important Role in Cell Growth and DNA Damage Control in Human Cancer Cells. *Molecular Cell*, 22(3):375–382, May 2006.
- [87] Richard G. Stevens. Circadian Disruption and Breast Cancer: From Melatonin to Clock Genes. *Epidemiology*, 16(2):254, March 2005.
- [88] Sandra Sephton and David Spiegel. Circadian disruption in cancer: a neuroendocrine-immune pathway from stress to disease? *Brain, Behavior, and Immunity*, 17(5):321–328, October 2003.
- [89] Zandra E. Walton, Brian J. Altman, Rebekah C. Brooks, and Chi V. Dang. Circadian Clock’s Cancer Connections. *Annual Review of Cancer Biology*, 2(1):133–153, March 2018.
- [90] Ron Edgar and Tanya Barrett. NCBI GEO standards and services for microarray data. *Nature Biotechnology*, 24(12):1471, December 2006.
- [91] J. Bass and J. S. Takahashi. Circadian Integration of Metabolism and Energetics. *Science*, 330(6009):1349–1354, December 2010. Number: 6009 ZSCC: 0001549.
- [92] Satchidananda Panda. Circadian physiology of metabolism. *Science*, 354(6315):1008–1015, November 2016.
- [93] Shubhroz Gill, Hiep D. Le, Girish C. Melkani, and Satchidananda Panda. Time-restricted feeding attenuates age-related cardiac decline in *Drosophila*. *Science (New York, N.Y.)*, 347(6227):1265–1269, March 2015.
- [94] Amandine Chaix, Emily N.C. Manoogian, Girish C. Melkani, and Satchidananda Panda. Time-Restricted Eating to Prevent and Manage Chronic Metabolic Diseases. *Annual review of nutrition*, 39:291–315, August 2019.
- [95] Patrick Boyd, Sydney G O’Connor, Brandy M Heckman-Stoddard, and Edward R Sauter. Time-Restricted Feeding Studies and Possible Human Benefit. *JNCI Cancer Spectrum*, 6(3):pkac032, June 2022.

Bibliography

- [96] Rachel Leproult, Gaétane Deliens, Médhi Gilson, and Philippe Peigneux. Beneficial Impact of Sleep Extension on Fasting Insulin Sensitivity in Adults with Habitual Sleep Restriction. *Sleep*, 38(5):707–715, May 2015.
- [97] Rob H. P. Henst, Paula R. Pienaar, Laura C. Roden, and Dale E. Rae. The effects of sleep extension on cardiometabolic risk factors: A systematic review. *Journal of Sleep Research*, 28(6):e12865, 2019. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/jsr.12865>.
- [98] Mariana G Figueiro, Barbara A Plitnick, Anna Lok, Geoffrey E Jones, Patricia Higgins, Thomas R Hornick, and Mark S Rea. Tailored lighting intervention improves measures of sleep, depression, and agitation in persons with Alzheimer’s disease and related dementia living in long-term care facilities. *Clinical Interventions in Aging*, 9:1527–1537, September 2014.
- [99] Mariana G Figueiro. Light, sleep and circadian rhythms in older adults with Alzheimer’s disease and related dementias. *Neurodegenerative Disease Management*, 7(2):119–145, April 2017. Publisher: Future Medicine.
- [100] Pasquale F. Innominato, Annabelle Ballesta, Qi Huang, Christian Focan, Philippe Chollet, Abdoulaye Karaboué, Sylvie Giacchetti, Mohamed Bouchahda, René Adam, Carlo Garufi, and Francis A. Lévi. Sex-dependent least toxic timing of irinotecan combined with chronomodulated chemotherapy for metastatic colorectal cancer: Randomized multicenter EORTC 05011 trial. *Cancer Medicine*, 9(12):4148–4159, 2020. Number: 12 _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/cam4.3056>.
- [101] Seán T. Anderson and Garret A. FitzGerald. Sexual dimorphism in body clocks. *Science*, 369(6508):1164–1165, September 2020. Publisher: American Association for the Advancement of Science Section: Perspective.
- [102] Benjamin D. Weger, Cédric Gobet, Jake Yeung, Eva Martin, Sonia Jimenez, Bertrand Betrisey, Francis Foata, Bernard Berger, Aurélie Balvay, Anne Foussier, Aline Charpagne, Brigitte Boizet-Bonhoure, Chieh Jason Chou, Felix Naef, and Frédéric Gachon. The Mouse Microbiome Is Required for Sex-Specific Diurnal Rhythms of Gene Expression and Metabolism. *Cell Metabolism*, 29(2):362–382.e8, February 2019.
- [103] Laura M Lyall, Cathy A Wyse, Nicholas Graham, Amy Ferguson, Donald M Lyall, Breda Cullen, Carlos A Celis Morales, Stephany M Biello, Daniel Mackay, Joey Ward, Rona J Strawbridge, Jason M R Gill, Mark E S Bailey, Jill P Pell, and

- Daniel J Smith. Association of disrupted circadian rhythmicity with mood disorders, subjective wellbeing, and cognitive function: a cross-sectional study of 91105 participants from the UK Biobank. *The Lancet Psychiatry*, 5(6):507–514, June 2018. Number: 6 ZSCC: NoCitationData[s0].
- [104] Edward O. Bixler, Maria N. Papaliaga, Alexandros N. Vgontzas, Hung-Mo Lin, Slobodanka Pejovic, Maria Karataraki, Antonio Vela-Bueno, and George P. Chrousos. Women sleep objectively better than men and the sleep of young women is more resilient to external stressors: effects of age and menopause. *Journal of Sleep Research*, 18(2):221–228, 2009. Number: 2 _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2869.2008.00713.x>.
- [105] Nayantara Santhi, Alpar S. Lazar, Patrick J. McCabe, June C. Lo, John A. Groeger, and Derk-Jan Dijk. Sex differences in the circadian regulation of sleep and waking cognition in humans. *Proceedings of the National Academy of Sciences*, 113(19):E2730–E2739, May 2016. Publisher: Proceedings of the National Academy of Sciences.
- [106] Sean W. Cain, Christopher F. Dennison, Jamie M. Zeitzer, Aaron M. Guzik, Sat Bir S. Khalsa, Nayantara Santhi, Martin W. Schoen, Charles A. Czeisler, and Jeanne E. Duffy. Sex Differences in Phase Angle of Entrainment and Melatonin Amplitude in Humans. *Journal of Biological Rhythms*, 25(4):288–296, August 2010. Publisher: SAGE Publications Inc.
- [107] Dorothee Fischer, David A. Lombardi, Helen Marucci-Wellman, and Till Roenneberg. Chronotypes in the US – Influence of age and sex. *PLOS ONE*, 12(6):e0178782, June 2017. Publisher: Public Library of Science.
- [108] World Health Organization. *World Report on Ageing and Health*. World Health Organization, October 2015. Google-Books-ID: n180DgAAQBAJ.
- [109] Suzanne Hood and Shimon Amir. The aging clock: circadian rhythms and later life. *The Journal of Clinical Investigation*, 127(2):437–446, February 2017. Number: 2 ZSCC: 0000200 Publisher: American Society for Clinical Investigation.
- [110] Oren Froy. Circadian aspects of energy metabolism and aging. *Ageing Research Reviews*, 12(4):931–940, September 2013. Number: 4 ZSCC: 0000028.
- [111] Shin Yamazaki, Marty Straume, Hajime Tei, Yoshiyuki Sakaki, Michael Menaker, and Gene D. Block. Effects of aging on central and peripheral mam-

Bibliography

- malian clocks. *Proceedings of the National Academy of Sciences*, 99(16):10801–10806, August 2002. Number: 16 ZSCC: 0000283 Publisher: National Academy of Sciences Section: Biological Sciences.
- [112] Hung-Chun Chang and Leonard Guarente. SIRT1 Mediates Central Circadian Control in the SCN by a Mechanism that Decays with Aging. *Cell*, 153(7):1448–1460, June 2013.
- [113] Michael T. Sellix, Jennifer A. Evans, Tanya L. Leise, Oscar Castanon-Cervantes, DiJon D. Hill, Patrick DeLisser, Gene D. Block, Michael Menaker, and Alec J. Davidson. Aging Differentially Affects the Re-entrainment Response of Central and Peripheral Circadian Oscillators. *Journal of Neuroscience*, 32(46):16193–16202, November 2012. Publisher: Society for Neuroscience Section: Articles.
- [114] Yan Zhang, George C Brainard, Phyllis C Zee, Lawrence H Pinto, Joseph S Takahashi, and Fred W Turek. Effects of aging on lens transmittance and retinal input to the suprachiasmatic nucleus in golden hamsters. *Neuroscience Letters*, 258(3):167–170, December 1998.
- [115] Verónica S. Valentinuzzi, Kathryn Scarbrough, Joseph S. Takahashi, and Fred W. Turek. Effects of aging on the circadian rhythm of wheel-running activity in C57BL/6 mice. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 273(6):R1957–R1964, December 1997. Number: 6 ZSCC: 0000273 Publisher: American Physiological Society.
- [116] Yuliya V. Dubrovsky, William E. Samsa, and Roman V. Kondratov. Deficiency of circadian protein CLOCK reduces lifespan and increases age-related cataract development in mice. *Aging (Albany NY)*, 2(12):936–944, December 2010. Number: 12 ZSCC: 0000151.
- [117] Baokun He, Kazunari Nohara, Noheon Park, Yong-Sung Park, Bobby Guillory, Zhaoyang Zhao, Jose M. Garcia, Nobuya Koike, Cheng Chi Lee, Joseph S. Takahashi, Seung-Hee Yoo, and Zheng Chen. The Small Molecule Nobiletin Targets the Molecular Oscillator to Enhance Circadian Rhythms and Protect against Metabolic Syndrome. *Cell Metabolism*, 23(4):610–621, April 2016.
- [118] Subhash D. Katewa, Kazutaka Akagi, Neelanjan Bose, Kuntol Rakshit, Timothy Camarella, Xiangzhong Zheng, David Hall, Sonnet Davis, Christopher S. Nelson, Rachel B. Brem, Arvind Ramanathan, Amita Sehgal, Jadwiga M.

- Giebultowicz, and Pankaj Kapahi. Peripheral Circadian Clocks Mediate Dietary Restriction-Dependent Changes in Lifespan and Fat Metabolism in *Drosophila*. *Cell Metabolism*, 23(1):143–154, January 2016.
- [119] Sergiy Libert, Michael S. Bonkowski, Kelli Pointer, Scott D. Pletcher, and Leonard Guarente. Deviation of innate circadian period from 24 h reduces longevity in mice. *Aging Cell*, 11(5):794–800, 2012. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1474-9726.2012.00846.x>.
- [120] M. W. Hurd, K. A. Zimmer, M. N. Lehman, and M. R. Ralph. Circadian locomotor rhythms in aged hamsters following suprachiasmatic transplant. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 269(5):R958–R968, November 1995. Publisher: American Physiological Society.
- [121] Mark W. Hurd and Martin R. Ralph. The Significance of Circadian Organization for Longevity in the Golden Hamster. *Journal of Biological Rhythms*, 13(5):430–436, October 1998. Publisher: SAGE Publications Inc.
- [122] A. J. Davidson, M. T. Sellix, J. Daniel, S. Yamazaki, M. Menaker, and G. D. Block. Chronic jet-lag increases mortality in aged mice. *Current Biology*, 16(21):R914–R916, November 2006.
- [123] Elizabeth A. Yu and David R. Weaver. Disrupting the circadian clock: Gene-specific effects on aging, cancer, and other phenotypes. *Aging*, May 2011. Accepted: 2022-08-23T16:11:21Z.
- [124] Imre Lakatos. History of Science and Its Rational Reconstructions. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1970:91–136, January 1970. Publisher: The University of Chicago Press.
- [125] Lydia Patton, editor. *Philosophy, Science, and History: A Guide and Reader*. Routledge, New York, May 2014.
- [126] E. Nordenskiöld. *The history of biology*. The history of biology. Knopf, Oxford, England, 1928. Pages: x, 629.
- [127] C. Singer. A history of biology. *A history of biology*, 1950. Publisher: Henry Schuman, New York.
- [128] Gangxu Shen. Campbell biology (edited by Lisa Urry, Michael Cain, Steven Wasserman, Peter Minorsky and Jane Reece). *Journal of Biological Research-Thessaloniki*, 27(1):19, December 2020.

Bibliography

- [129] Michel Morange. *A History of Molecular Biology*. Harvard University Press, 2000. Google-Books-ID: Qe93vz5ZLVsC.
- [130] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Molecular Biology of the Cell*. Garland Science, 4th edition, 2002.
- [131] Edda Klipp, Wolfram Liebermeister, Christoph Wierling, Axel Kowald, Hans Lehrach, and Ralf Herwig. *Systems Biology: A Textbook*. January 2009.
- [132] D F Hunt, J R Yates, J Shabanowitz, S Winston, and C R Hauer. Protein sequencing by tandem mass spectrometry. *Proceedings of the National Academy of Sciences*, 83(17):6233–6237, September 1986. Publisher: Proceedings of the National Academy of Sciences.
- [133] Bryan John Smith. *Protein Sequencing Protocols*. Springer Science & Business Media, February 2008.
- [134] Laura Restrepo-Pérez, Chirlmin Joo, and Cees Dekker. Paving the way to single-molecule protein sequencing. *Nature Nanotechnology*, 13(9):786–796, September 2018. Number: 9 Publisher: Nature Publishing Group.
- [135] Nicholas Callahan, Jennifer Tullman, Zvi Kelman, and John Marino. Strategies for Development of a Next-Generation Protein Sequencing Platform. *Trends in Biochemical Sciences*, 45(1):76–89, January 2020.
- [136] Christof M. Niemeyer and Dietmar Blohm. DNA Microarrays. *Angewandte Chemie International Edition*, 38(19):2865–2869, 1999. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291521-3773%2819991004%2938%3A19%3C2865%3A%3AAID-ANIE2865%3E3.0.CO%3B2-F>
- [137] David J Graves. Powerful tools for genetic analysis come of age. *Trends in Biotechnology*, 17(3):127–134, March 1999.
- [138] Charlie C. Xiang and Yidong Chen. cDNA microarray technology and its applications. *Biotechnology Advances*, 18(1):35–46, March 2000.
- [139] Deirdre Meldrum. Automation for Genomics, Part Two: Sequencers, Microarrays, and Future Trends. *Genome Research*, 10(9):1288–1303, September 2000. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press

Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

- [140] David J. Lockhart and Elizabeth A. Winzeler. Genomics, gene expression and DNA arrays. *Nature*, 405(6788):827–836, June 2000. Number: 6788 Publisher: Nature Publishing Group.
- [141] Richard A. Young. Biomedical Discovery with DNA Arrays. *Cell*, 102(1):9–15, July 2000. Publisher: Elsevier.
- [142] Dietmar H Blohm and Anthony Guiseppi-Elie. New developments in microarray technology. *Current Opinion in Biotechnology*, 12(1):41–47, February 2001.
- [143] Ugrappa Nagalakshmi, Zhong Wang, Karl Waern, Chong Shou, Debasish Raha, Mark Gerstein, and Michael Snyder. The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. *Science*, 320(5881):1344–1349, June 2008. Publisher: American Association for the Advancement of Science.
- [144] Ali Mortazavi, Brian A. Williams, Kenneth McCue, Lorian Schaeffer, and Barbara Wold. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, 5(7):621–628, July 2008. Number: 7 Publisher: Nature Publishing Group.
- [145] Yangming Wang, Scott Baskerville, Archana Shenoy, Joshua E. Babiarz, Lauren Baehner, and Robert Blelloch. Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nature Genetics*, 40(12):1478–1483, December 2008. Number: 12 Publisher: Nature Publishing Group.
- [146] Nicole Cloonan, Alistair R. R. Forrest, Gabriel Kolle, Brooke B. A. Gardiner, Geoffrey J. Faulkner, Mellissa K. Brown, Darrin F. Taylor, Anita L. Steptoe, Shivangi Wani, Graeme Bethel, Alan J. Robertson, Andrew C. Perkins, Stephen J. Bruce, Clarence C. Lee, Swati S. Ranade, Heather E. Peckham, Jonathan M. Manning, Kevin J. McKernan, and Sean M. Grimmond. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature Methods*, 5(7):613–619, July 2008. Number: 7 Publisher: Nature Publishing Group.
- [147] Zhong Wang, Mark Gerstein, and Michael Snyder. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1):57–63, January 2009. Number: 1 Publisher: Nature Publishing Group.

Bibliography

- [148] Fuchou Tang, Catalin Barbacioru, Yangzhou Wang, Ellen Nordman, Clarence Lee, Nanlan Xu, Xiaohui Wang, John Bodeau, Brian B. Tuch, Asim Siddiqui, Kaiqin Lao, and M. Azim Surani. mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*, 6(5):377–382, May 2009. Number: 5 Publisher: Nature Publishing Group.
- [149] Saiful Islam, Una Kjällquist, Annalena Moliner, Pawel Zajac, Jian-Bing Fan, Peter Lönnerberg, and Sten Linnarsson. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Research*, 21(7):1160–1167, July 2011. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.
- [150] Daniel Ramsköld, Shujun Luo, Yu-Chieh Wang, Robin Li, Qiaolin Deng, Omid R. Faridani, Gregory A. Daniels, Irina Khrebtkova, Jeanne F. Loring, Louise C. Laurent, Gary P. Schroth, and Rickard Sandberg. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotechnology*, 30(8):777–782, August 2012. Number: 8 Publisher: Nature Publishing Group.
- [151] Tamar Hashimshony, Florian Wagner, Noa Sher, and Itai Yanai. CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. *Cell Reports*, 2(3):666–673, September 2012.
- [152] Saiful Islam, Amit Zeisel, Simon Joost, Gioele La Manno, Pawel Zajac, Maria Kasper, Peter Lönnerberg, and Sten Linnarsson. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nature Methods*, 11(2):163–166, February 2014.
- [153] Allon M. Klein, Linas Mazutis, Ilke Akartuna, Naren Tallapragada, Adrian Veres, Victor Li, Leonid Peshkin, David A. Weitz, and Marc W. Kirschner. Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell*, 161(5):1187–1201, May 2015.
- [154] Michael Spivak. *A Comprehensive Introduction to Differential Geometry*. 1975.
- [155] J. B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, March 1964.

-
- [156] J.W. Sammon. A Nonlinear Mapping for Data Structure Analysis. *IEEE Transactions on Computers*, C-18(5):401–409, May 1969. Conference Name: IEEE Transactions on Computers.
- [157] Trevor Hastie and Werner Stuetzle. Principal Curves. *Journal of the American Statistical Association*, 84(406):502–516, June 1989.
- [158] Alexis Mignon and Frédéric Jurie. PCCA: A new approach for distance learning from sparse pairwise constraints. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2666–2672, June 2012. ISSN: 1063-6919.
- [159] Sam T. Roweis and Lawrence K. Saul. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, 290(5500):2323–2326, December 2000. Publisher: American Association for the Advancement of Science.
- [160] Mikhail Belkin and Partha Niyogi. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation*, 15(6):1373–1396, June 2003.
- [161] Ronald R. Coifman and Stéphane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, July 2006.
- [162] Laurens van der Maaten and Geoffrey Hinton. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.
- [163] Jian Tang, Jingzhou Liu, Ming Zhang, and Qiaozhu Mei. Visualizing Large-scale and High-dimensional Data. In *Proceedings of the 25th International Conference on World Wide Web, WWW '16*, pages 287–297, Republic and Canton of Geneva, CHE, April 2016. International World Wide Web Conferences Steering Committee.
- [164] Leland McInnes, John Healy, and James Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426 [cs, stat]*, February 2018. arXiv: 1802.03426.
- [165] Colas Droin, Eric R. Paquet, and Felix Naef. Low-dimensional dynamics of two coupled biological oscillators. *Nature Physics*, 15(10):1086–1094, October 2019. Number: 10 Publisher: Nature Publishing Group.
- [166] Giorgio Parisi. Statistical Physics and biology. *Physics World*, 6(9):42, September 1993.

Bibliography

- [167] Guy Sella and Aaron E. Hirsh. The application of statistical physics to evolutionary biology. *Proceedings of the National Academy of Sciences*, 102(27):9541–9546, July 2005. Publisher: Proceedings of the National Academy of Sciences.
- [168] Stephen Smale. On Gradient Dynamical Systems. *Annals of Mathematics*, 74(1):199–206, 1961. Publisher: Annals of Mathematics.
- [169] K. R. Meyer. Energy Functions for Morse Smale Systems. *American Journal of Mathematics*, 90(4):1031–1040, 1968. Publisher: Johns Hopkins University Press.
- [170] S. Smale. Differentiable dynamical systems. *Bulletin of the American Mathematical Society*, 73(6):747–817, 1967.
- [171] David A. Rand, Archishman Raju, Meritxell Sáez, Francis Corson, and Eric D. Siggia. Geometry of gene regulatory dynamics. *Proceedings of the National Academy of Sciences*, 118(38):e2109729118, September 2021. Publisher: Proceedings of the National Academy of Sciences.
- [172] Nadim Aizarani, Antonio Saviano, Sagar, Laurent Mailly, Sarah Durand, Josip S. Herman, Patrick Pessaux, Thomas F. Baumert, and Dominic Grün. A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature*, 572(7768):199–204, August 2019. Number: 7768 Publisher: Nature Publishing Group.
- [173] Principal investigators, The Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type annotation, Writing group, Supplemental text writing group, Nicholas Schaum, Jim Karkanias, Norma F. Neff, Andrew P. May, Stephen R. Quake, Tony Wyss-Coray, Spyros Darmanis, Joshua Batson, Olga Botvinnik, Michelle B. Chen, Steven Chen, Foad Green, Robert C. Jones, Ashley Maynard, Lolita Penland, Angela Oliveira Pisco, Rene V. Sit, Geoffrey M. Stanley, James T. Webber, Fabio Zanini, Ankit S. Baghel, Isaac Bakerman, Ishita Bansal, Daniela Berdnik, Biter Bilen, Douglas Brownfield, Corey Cain, Min Cho, Giana Cirolia, Stephanie D. Conley, Aaron Demers, Kubilay Demir, Antoine de Morree, Tessa Divita, Haley du Bois, Laughing Bear Torrez Dulgeroff, Hamid Ebadi, F. Hernán Espinoza, Matt Fish, Qiang Gan, Benson M. George, Astrid Gillich, Geraldine Genetiano, Xueying Gu, Gunsagar S. Gulati, Yan Hang, Shayan Hosseinzadeh, Albin Huang, Tal

- Iram, and Taichi Isobe. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*, 562(7727):367–372, October 2018.
- [174] Ilias Angelidis, Lukas M. Simon, Isis E. Fernandez, Maximilian Strunz, Christoph H. Mayr, Flavia R. Greiffo, George Tsitsiridis, Meshal Ansari, Elisabeth Graf, Tim-Matthias Strom, Monica Nagendran, Tushar Desai, Oliver Eickelberg, Matthias Mann, Fabian J. Theis, and Herbert B. Schiller. An atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics. *Nature Communications*, 10(1):963, February 2019. Number: 1 Publisher: Nature Publishing Group.
- [175] Jacob C. Kimmel, Lolita Penland, Nimrod D. Rubinstein, David G. Hendrickson, David R. Kelley, and Adam Z. Rosenthal. Murine single-cell RNA-seq reveals cell-identity- and tissue-specific trajectories of aging. *Genome Research*, 29(12):2088–2103, December 2019. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.
- [176] Sabina Kanton, Michael James Boyle, Zhisong He, Malgorzata Santel, Anne Weigert, Fátima Sanchís-Calleja, Patricia Guijarro, Leila Sidow, Jonas Simon Fleck, Dingding Han, Zhengzong Qian, Michael Heide, Wieland B. Huttner, Philipp Khaitovich, Svante Pääbo, Barbara Treutlein, and J. Gray Camp. Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature*, 574(7778):418–422, October 2019. Number: 7778 Publisher: Nature Publishing Group.
- [177] Kanti V. Mardia and Jochen Voss. Some Fundamental Properties of a Multivariate von Mises Distribution. *Communications in Statistics - Theory and Methods*, 43(6):1132–1144, March 2014.
- [178] Amit Zeisel, Wolfgang J Köstler, Natali Molotski, Jonathan M Tsai, Rita Krauthgamer, Jasmine Jacob-Hirsch, Gideon Rechavi, Yoav Soen, Steffen Jung, Yosef Yarden, and Eytan Domany. Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli. *Molecular Systems Biology*, 7(1):529, January 2011. Publisher: John Wiley & Sons, Ltd.
- [179] Jesse M. Gray, David A. Harmin, Sarah A. Boswell, Nicole Cloonan, Thomas E. Mullen, Joseph J. Ling, Nimrod Miller, Scott Kuersten, Yong-Chao Ma,

- Steven A. McCarroll, Sean M. Grimmond, and Michael Springer. SnapShot-Seq: A Method for Extracting Genome-Wide, In Vivo mRNA Dynamics from a Single Total RNA Sample. *PLoS ONE*, 9(2):e89673, February 2014.
- [180] Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E. Kastriti, Peter Lönnerberg, Alessandro Furlan, Jean Fan, Lars E. Borm, Zehua Liu, David van Bruggen, Jimin Guo, Xiaoling He, Roger Barker, Erik Sundström, Gonçalo Castelo-Branco, Patrick Cramer, Igor Adameyko, Sten Linnarsson, and Peter V. Kharchenko. RNA velocity of single cells. *Nature*, 560(7719):494–498, August 2018. Number: 7719 Publisher: Nature Publishing Group.
- [181] Allon M. Klein. Technique to measure the expression dynamics of each gene in a single cell. *Nature*, 560(7719):434–435, August 2018. Bandiera_abtest: a Cg_type: News And Views Number: 7719 Publisher: Nature Publishing Group Subject_term: Computational biology and bioinformatics, Developmental biology, Transcriptomics.
- [182] Quentin Lo Giudice, Marion Leleu, Gioele La Manno, and Pierre J. Fabre. Single-cell transcriptional logic of cell-fate specification and axon guidance in early-born retinal neurons. *Development*, 146(17):dev178103, September 2019.
- [183] Bo Xia, Yun Yan, Maayan Baron, Florian Wagner, Dalia Barkley, Marta Chiodin, Sang Y. Kim, David L. Keefe, Joseph P. Alukal, Jef D. Boeke, and Itai Yanai. Widespread Transcriptional Scanning in the Testis Modulates Gene Evolution Rates. *Cell*, 180(2):248–262.e21, January 2020.
- [184] Jingtao Guo, Edward J. Grow, Hana Mlcochova, Geoffrey J. Maher, Cecilia Lindskog, Xichen Nie, Yixuan Guo, Yodai Takei, Jina Yun, Long Cai, Robin Kim, Douglas T. Carrell, Anne Goriely, James M. Hotaling, and Bradley R. Cairns. The adult human testis transcriptional cell atlas. *Cell Research*, 28(12):1141–1157, December 2018. Number: 12 Publisher: Nature Publishing Group.
- [185] Vera Zywitza, Aristotelis Misios, Lena Bunatyan, Thomas E. Willnow, and Nikolaus Rajewsky. Single-Cell Transcriptomics Characterizes Cell Types in the Subventricular Zone and Uncovers Molecular Defects Impairing Adult Neurogenesis. *Cell Reports*, 25(9):2457–2469.e8, November 2018.
- [186] Qiming Zhang, Yao He, Nan Luo, Shashank J. Patel, Yanjie Han, Ranran Gao, Madhura Modak, Sebastian Carotta, Christian Haslinger, David Kind,

- Gregory W. Peet, Guojie Zhong, Shuangjia Lu, Weihua Zhu, Yilei Mao, Mengmeng Xiao, Michael Bergmann, Xueda Hu, Sid P. Kerkar, Anne B. Vogt, Stefan Pflanz, Kang Liu, Jirun Peng, Xianwen Ren, and Zemin Zhang. Landscape and Dynamics of Single Immune Cells in Hepatocellular Carcinoma. *Cell*, 179(4):829–845.e20, October 2019.
- [187] Alex R Lederer and Gioele La Manno. The emergence and promise of single-cell temporal-omics approaches. *Current Opinion in Biotechnology*, 63:70–78, June 2020.
- [188] Volker Bergen, Marius Lange, Stefan Peidli, F Alexander Wolf, and Fabian J. Theis. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature Biotechnology*, 38(12):1408–1414, December 2020. Number: 12 Publisher: Nature Publishing Group.
- [189] Valentine Svensson and Lior Pachter. RNA Velocity: Molecular Kinetics from Single-Cell RNA-Seq. *Molecular Cell*, 72(1):7–9, October 2018.
- [190] Gennady Gorin, Valentine Svensson, and Lior Pachter. Protein velocity and acceleration from single-cell multiomics experiments. *Genome Biology*, 21(1):39, February 2020.
- [191] Rudolf Virchow. Notes of the winter semester, taken by a student., 1855.
- [192] Rudolf Ludwig Karl Virchow. *Die Cellularpathologie in ihrer Begründung auf physiologische und pathologische Gewebelehre*. A. Hirschwald, 1859. Google-Books-ID: Wi31tvfv3y4C.
- [193] K. A. Schafer. The Cell Cycle: A Review. *Veterinary Pathology*, 35(6):461–478, November 1998. Publisher: SAGE Publications Inc.
- [194] D. G. Johnson and C. L. Walker. Cyclins and Cell Cycle Checkpoints. *Annual Review of Pharmacology and Toxicology*, 39(1):295–312, 1999. _eprint: <https://doi.org/10.1146/annurev.pharmtox.39.1.295>.
- [195] Kim Nasmyth. Viewpoint: Putting the Cell Cycle in Order. *Science*, 274(5293):1643–1645, December 1996. Publisher: American Association for the Advancement of Science.
- [196] Helen K. Matthews, Cosetta Bertoli, and Robertus A. M. de Bruin. Cell cycle control in cancer. *Nature Reviews Molecular Cell Biology*, 23(1):74–88, January 2022. Number: 1 Publisher: Nature Publishing Group.

Bibliography

- [197] Asako Sakaue-Sawano, Hiroshi Kurokawa, Toshifumi Morimura, Aki Hanyu, Hiroshi Hama, Hatsuki Osawa, Saori Kashiwagi, Kiyoko Fukami, Takaki Miyata, Hiroyuki Miyoshi, Takeshi Imamura, Masaharu Ogawa, Hisao Masai, and Atsushi Miyawaki. Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. *Cell*, 132(3):487–498, February 2008.
- [198] Nico Battich, Joep Beumer, Buys de Barbanson, Lenno Krenning, Chloé S. Baron, Marvin E. Tanenbaum, Hans Clevers, and Alexander van Oudenaarden. Sequencing metabolically labeled transcripts in single cells reveals mRNA turnover strategies. *Science*, 367(6482):1151–1156, March 2020. Publisher: American Association for the Advancement of Science Section: Report.
- [199] Zehua Liu, Huazhe Lou, Kaikun Xie, Hao Wang, Ning Chen, Oscar M. Aparicio, Michael Q. Zhang, Rui Jiang, and Ting Chen. Reconstructing cell cycle pseudo time-series via single-cell transcriptome data. *Nature Communications*, 8(1):22, June 2017. Number: 1 Publisher: Nature Publishing Group.
- [200] Chenglong Xia, Jean Fan, George Emanuel, Junjie Hao, and Xiaowei Zhuang. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proceedings of the National Academy of Sciences*, 116(39):19490–19499, September 2019. Publisher: Proceedings of the National Academy of Sciences.
- [201] Shaoheng Liang, Fang Wang, Jincheng Han, and Ken Chen. Latent periodic process inference from single-cell RNA-seq data. *Nature Communications*, 11(1):1441, March 2020. Number: 1 Publisher: Nature Publishing Group.
- [202] Daniel Schwabe, Sara Formichetti, Jan Philipp Junker, Martin Falcke, and Nikolaus Rajewsky. The transcriptome dynamics of single cells during the cell cycle. *Molecular Systems Biology*, 16(11):e9946, November 2020. Publisher: John Wiley & Sons, Ltd.
- [203] Andrea Riba, Attila Oravecz, Matej Durik, Sara Jiménez, Violaine Alunni, Marie Cerciati, Matthieu Jung, Céline Keime, William M. Keyes, and Nacho Molina. Cell cycle gene regulation dynamics revealed by RNA velocity and deep-learning. *Nature Communications*, 13(1):2865, May 2022. Number: 1 Publisher: Nature Publishing Group.
- [204] Raymond J. Cho, Michael J. Campbell, Elizabeth A. Winzeler, Lars Steinmetz, Andrew Conway, Lisa Wodicka, Tyra G. Wolfsberg, Andrei E. Gabrielian, David

- Landsman, David J. Lockhart, and Ronald W. Davis. A Genome-Wide Transcriptional Analysis of the Mitotic Cell Cycle. *Molecular Cell*, 2(1):65–73, July 1998.
- [205] E.D. Israels and L.G. Israels. The Cell Cycle. *The Oncologist*, 5(6):510–513, December 2000.
- [206] J. Wade Harper and Stephen J. Elledge. Cdk inhibitors in development and cancer. *Current Opinion in Genetics & Development*, 6(1):56–64, February 1996.
- [207] Arnaud Besson, Steven F. Dowdy, and James M. Roberts. CDK Inhibitors: Cell Cycle Regulators and Beyond. *Developmental Cell*, 14(2):159–169, February 2008.
- [208] Potu N. Rao and Joseph Engelberg. HeLa Cells: Effects of Temperature on the Life Cycle. *Science*, 148(3673):1092–1094, May 1965. Publisher: American Association for the Advancement of Science.
- [209] D Francis and P W Barlow. Temperature and the cell cycle. *Symposia of the Society for Experimental Biology*, 42:181–201, January 1988.
- [210] Lorenzo Talamanca and Felix Naef. How to tell time: advances in decoding circadian phase from omics snapshots. *F1000Research*, 9, 2020.
- [211] Robert Dallmann, Steven A. Brown, and Frédéric Gachon. Chronopharmacology: New Insights and Therapeutic Implications. *Annual Review of Pharmacology and Toxicology*, 54(1):339–361, 2014. _eprint: <https://doi.org/10.1146/annurev-pharmtox-011613-135923>.
- [212] Svante Wold, Kim Esbensen, and Paul Geladi. Principal Component Analysis. page 16.
- [213] Hiroki R. Ueda, Wenbin Chen, Yoichi Minami, Sato Honma, Kenichi Honma, Masamitsu Iino, and Seiichi Hashimoto. Molecular-timetable methods for detection of body time and rhythm disorders from single-time-point genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 101(31):11227–11232, August 2004. ZSCC: 0000100.
- [214] Emma E Laing, Carla S Möller-Levet, Norman Poh, Nayantara Santhi, Simon N Archer, and Derk-Jan Dijk. Blood transcriptome based biomarkers for human circadian phase. *eLife*, 6:e20214, February 2017. ZSCC: 0000049.

- [215] Nicole Wittenbrink, Bharath Ananthasubramaniam, Mirjam Münch, Barbara Koller, Bert Maier, Charlotte Weschke, Frederik Bes, Jan de Zeeuw, Claudia Nowozin, Amely Wahnschaffe, Sophia Wisniewski, Mandy Zaleska, Osnat Bartok, Reut Ashwal-Fluss, Hedwig Lammert, Hanspeter Herzel, Michael Hummel, Sebastian Kadener, Dieter Kunz, and Achim Kramer. High-accuracy determination of internal circadian time from a single blood sample. *Journal of Clinical Investigation*, 128(9):3826–3839, August 2018.
- [216] Gang Wu, Marc D. Ruben, Robert E. Schmidt, Lauren J. Francey, David F. Smith, Ron C. Anafi, Jacob J. Hughey, Ryan Tasseff, Joseph D. Sherrill, John E. Oblong, Kevin J. Mills, and John B. Hogenesch. Population-level rhythms in human skin with implications for circadian medicine. *Proceedings of the National Academy of Sciences*, 115(48):12313–12318, November 2018.
- [217] Marc D. Ruben, Gang Wu, David F. Smith, Robert E. Schmidt, Lauren J. Francey, Yin Yeng Lee, Ron C. Anafi, and John B. Hogenesch. A database of tissue-specific rhythmically expressed human genes has potential applications in circadian medicine. *Science Translational Medicine*, 10(458):eaat8806, September 2018. Number: 458.
- [218] Christopher R. Cederroth, Urs Albrecht, Joseph Bass, Steven A. Brown, Jonas Dyhrfeld-Johnsen, Frederic Gachon, Carla B. Green, Michael H. Hastings, Charlotte Helfrich-Förster, John B. Hogenesch, Francis Lévi, Andrew Loudon, Gabriella B. Lundkvist, Johanna H. Meijer, Michael Rosbash, Joseph S. Takahashi, Michael Young, and Barbara Canlon. Medicine in the Fourth Dimension. *Cell Metabolism*, 30(2):238–250, August 2019. Number: 2.
- [219] Lorenzo Talamanca, Cédric Gobet, and Felix Naef. Sex-dimorphic and age-dependent organization of 24 hour gene expression rhythms in human, June 2022. bioRxiv.
- [220] Marc Mezard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin Glass Theory and Beyond*. World Scientific, 1987. Google-Books-ID: ZIF9QgAACAAJ.
- [221] Cirano De Dominicis and Irene Giardina. Random Fields and Spin Glasses: A Field Theory Approach, October 2006. ISBN: 9780521847834 9780511534836 9780521143554 Library Catalog: www.cambridge.org Publisher: Cambridge University Press.
- [222] Joseph D. Bryngelson. An introduction to the theory of spin glasses and neural networks. *Journal of Statistical Physics*, 85(3):519–520, November 1996.

-
- [223] Tommaso Castellani and Andrea Cavagna. Spin-Glass Theory for Pedestrians. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(05):P05012, May 2005. arXiv:cond-mat/0505032.
- [224] David L. Stein and Charles M. Nweman. *Spin Glasses and Complexity*. January 2013.
- [225] Nicolas Boumal. Nonconvex phase synchronization. *SIAM Journal on Optimization*, 26(4):2355–2377, January 2016. Number: 4 arXiv: 1601.06114.
- [226] Nicolas Boumal, P.-A. Absil, and Coralía Cartis. Global rates of convergence for nonconvex optimization on manifolds. *IMA Journal of Numerical Analysis*, 39(1):1–33, January 2019. Number: 1.
- [227] Y. K Gambhir, P Ring, and A Thimet. Relativistic mean field theory for finite nuclei. *Annals of Physics*, 198(1):132–179, February 1990.
- [228] T. Garel and H. Orland. Mean-Field Model for Protein Folding. *Europhysics Letters (EPL)*, 6(4):307–310, June 1988.
- [229] David Boris and Michael Rubinstein. A Self-Consistent Mean Field Model of a Starburst Dendrimer: Dense Core vs Dense Shell. *Macromolecules*, 29(22):7251–7260, January 1996.
- [230] Henrik Flyvbjerg, Kim Sneppen, and Per Bak. Mean field theory for a simple model of evolution. *Physical Review Letters*, 71(24):4087–4090, December 1993.
- [231] Bernd A. Heß, Christel M. Marian, Ulf Wahlgren, and Odd Gropen. A mean-field spin-orbit method applicable to correlated wavefunctions. *Chemical Physics Letters*, 251(5):365–371, March 1996.
- [232] Oleg V. Prezhdo. Mean field approximation for the stochastic Schrödinger equation. *The Journal of Chemical Physics*, 111(18):8366–8377, October 1999.
- [233] Frank Dellaert. The Expectation Maximization Algorithm. Technical Report, Georgia Institute of Technology, 2002.
- [234] T. K. Moon. The expectation-maximization algorithm. *IEEE Signal Processing Magazine*, 13(6):47–60, November 1996.
- [235] Michael E Tipping and Christopher M Bishop. Probabilistic Principal Component Analysis. 1998.

- [236] Laurent Perrin, Ursula Loizides-Mangold, Stéphanie Chanon, Cédric Gobet, Nicolas Hulo, Laura Isenegger, Benjamin D Weger, Eugenia Migliavacca, Aline Charpagne, James A Betts, Jean-Philippe Walhin, Iain Templeman, Keith Stokes, Dylan Thompson, Kostas Tsintzas, Maud Robert, Cedric Howald, Howard Riezman, Jerome N Feige, Leonidas G Karagounis, Jonathan D Johnston, Emmanouil T Dermitzakis, Frédéric Gachon, Etienne Lefai, and Charna Dibner. Transcriptomic analyses reveal rhythmic and CLOCK-driven pathways in human skeletal muscle. *eLife*, 7:e34114, April 2018. Publisher: eLife Sciences Publications, Ltd.
- [237] Jürgen Aschoff. Circadian Rhythms in Man: A self-sustained oscillator with an inherent frequency underlies human 24-hour periodicity. *Science*, 148(3676):1427–1432, June 1965. Number: 3676.
- [238] Till Roenneberg, Tim Kuehnle, Myriam Juda, Thomas Kantermann, Karla Allebrandt, Marijke Gordijn, and Martha Merrow. Epidemiology of the human circadian clock. *Sleep Medicine Reviews*, 11(6):429–438, December 2007. Number: 6.
- [239] Meritxell Oliva, Manuel Muñoz-Aguirre, Sarah Kim-Hellmuth, Valentin Wucher, Ariel D. H. Gewirtz, Daniel J. Cotter, Princy Parsana, Silva Kasela, Brunilda Balliu, Ana Viñuela, Stephane E. Castel, Pejman Mohammadi, François Aguet, Yuxin Zou, Ekaterina A. Khramtsova, Andrew D. Skol, Diego Garrido-Martín, Ferran Reverter, Andrew Brown, Patrick Evans, Eric R. Gamazon, Anthony Payne, Rodrigo Bonazzola, Alvaro N. Barbeira, Andrew R. Hamel, Angel Martinez-Perez, José Manuel Soria, GTEx Consortium\$, Brandon L. Pierce, Matthew Stephens, Eleazar Eskin, Emmanouil T. Dermitzakis, Ayellet V. Segrè, Hae Kyung Im, Barbara E. Engelhardt, Kristin G. Ardlie, Stephen B. Montgomery, Alexis J. Battle, Tuuli Lappalainen, Roderic Guigó, and Barbara E. Stranger. The impact of sex on gene expression across human tissues. *Science*, 369(6509), September 2020. Publisher: American Association for the Advancement of Science Section: Research Article.
- [240] Carole Ober, Dagan A. Loisel, and Yoav Gilad. Sex-specific genetic architecture of human disease. *Nature Reviews Genetics*, 9(12):911–922, December 2008. Number: 12.
- [241] Ryo Yamamoto, Ryan Chung, Juan Manuel Vazquez, Huanjie Sheng, Philippa L. Steinberg, Nilah M. Ioannidis, and Peter H. Sudmant. Tissue-specific impacts of aging and genetics on gene expression patterns in humans. *Nature Communications*, 13(1):5803, October 2022. Number: 1.

- [242] Victoria A. Acosta-Rodríguez, Filipa Rijo-Ferreira, Carla B. Green, and Joseph S. Takahashi. Importance of circadian timing for aging and longevity. *Nature Communications*, 12(1):2862, May 2021. Number: 1 Publisher: Nature Publishing Group.
- [243] NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. GTEx Portal, 2021.
- [244] Pinar Pezük, Jennifer A. Mohawk, Laura A. Wang, and Michael Menaker. Glucocorticoids as Entraining Signals for Peripheral Circadian Oscillators. *Endocrinology*, 153(10):4775–4783, October 2012. Number: 10.
- [245] David Morse, Nicolas Cermakian, Stefano Brancorsini, Martti Parvinen, and Paolo Sassone-Corsi. No circadian rhythms in testis: Period1 expression is clock independent and developmentally regulated in the mouse. *Molecular Endocrinology (Baltimore, Md.)*, 17(1):141–151, January 2003. Number: 1.
- [246] Hans Reinke, Camille Saini, Fabienne Fleury-Olela, Charna Dibner, Ivor J. Benjamin, and Ueli Schibler. Differential display of DNA-binding proteins reveals heat-shock factor 1 as a circadian transcription factor. *Genes & Development*, 22(3):331–345, February 2008. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.
- [247] Jürgen Aschoff. Circadian control of body temperature. *Journal of Thermal Biology*, 8(1-2):143–147, January 1983. Number: 1-2.
- [248] Rebecca San Gil, Lezanne Ooi, Justin J. Yerbury, and Heath Ecroyd. The heat shock response in neurons and astroglia and its role in neurodegenerative diseases. *Molecular Neurodegeneration*, 12(1):65, December 2017. Number: 1.
- [249] H. M. Reichardt, Jan P. Tuckermann, Martin Goettlicher, Maja Vujic, Falk Weih, Peter Angel, Peter Herrlich, and Gunther Schutz. Repression of inflammatory responses in the absence of DNA binding by the glucocorticoid receptor. *The EMBO Journal*, 20(24):7168–7173, December 2001. Number: 24.
- [250] Jingkui Wang, Daniel Mauvoisin, Eva Martin, Florian Atger, Antonio Núñez Galindo, Loïc Dayon, Federico Sizzano, Alessio Palini, Martin Kussmann, Patrice Waridel, Manfredo Quadroni, Vjekoslav Dulić, Felix Naef, and Frédéric Gachon. Nuclear Proteomics Uncovers Diurnal Regulatory Landscapes in Mouse Liver. *Cell Metabolism*, 25(1):102–117, January 2017. Number: 1.

Bibliography

- [251] P. J. Jones and D. A. Schoeller. Evidence for diurnal periodicity in human cholesterol synthesis. *Journal of Lipid Research*, 31(4):667–673, April 1990. Number: 4.
- [252] Sarah C. McLoughlin, Philip Haines, and Garret A. FitzGerald. Clocks and Cardiovascular Function. In *Methods in Enzymology*, volume 552, pages 211–228. Elsevier, 2015.
- [253] Yuan-Fu Lu, Tao Jin, Yasha Xu, Dan Zhang, Qin Wu, Yu-Kun Jennifer Zhang, and Jie Liu. Sex Differences in the Circadian Variation of Cytochrome P450 Genes and Corresponding Nuclear Receptors in Mouse Liver. *Chronobiology international*, 30(9):1135–1143, November 2013. Number: 9.
- [254] Cho-Yi Chen, Ryan W. Logan, Tianzhou Ma, David A. Lewis, George C. Tseng, Etienne Sibille, and Colleen A. McClung. Effects of aging on circadian patterns of gene expression in the human prefrontal cortex. *Proceedings of the National Academy of Sciences*, 113(1):206–211, January 2016. Number: 1.
- [255] Jia Shi, Yi Yang, Anying Cheng, Gang Xu, and Fan He. Metabolism of vascular smooth muscle cells in vascular diseases. *American Journal of Physiology-Heart and Circulatory Physiology*, 319(3):H613–H631, September 2020. Number: 3.
- [256] R R Freedman, D Norton, S Woodward, and G Cornélissen. Core body temperature and circadian rhythm of hot flashes in menopausal women. *The Journal of Clinical Endocrinology & Metabolism*, 80(8):2354–2358, August 1995. Number: 8.
- [257] D. Weinert. Age-Dependent Changes of the Circadian System. *Chronobiology International*, 17(3):261–283, January 2000. Number: 3 Publisher: Taylor & Francis _eprint: <https://doi.org/10.1081/CBI-100101048>.
- [258] Stefano Bellentani, Federica Scaglioni, Mariano Marino, and Giorgio Bedogni. Epidemiology of Non-Alcoholic Fatty Liver Disease. *Digestive Diseases*, 28(1):155–161, 2010. Publisher: Karger Publishers.
- [259] Pekka Jousilahti, Erkki Vartiainen, Jaakko Tuomilehto, and Pekka Puska. Sex, Age, Cardiovascular Risk Factors, and Coronary Heart Disease: A Prospective Follow-Up Study of 14 786 Middle-Aged Men and Women in Finland. *Circulation*, 99(9):1165–1172, March 1999. Number: 9.

- [260] Joana Bicker, Gilberto Alves, Amílcar Falcão, and Ana Fortuna. Timing in drug absorption and disposition: The past, present, and future of chronopharmacokinetics. *British Journal of Pharmacology*, 177(10):2215–2239, May 2020. Number: 10.
- [261] Mark D. Robinson, Davis J. McCarthy, and Gordon K. Smyth. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, 26(1):139–140, January 2010. Number: 1.
- [262] Pedro G. Ferreira, Manuel Muñoz-Aguirre, Ferran Reverter, Caio P. Sá Godinho, Abel Sousa, Alicia Amadoz, Reza Sodaee, Marta R. Hidalgo, Dmitri Pervouchine, Jose Carbonell-Caballero, Ramil Nurtdinov, Alessandra Breschi, Raziel Amador, Patrícia Oliveira, Cankut Çubuk, João Curado, François Aguet, Carla Oliveira, Joaquin Dopazo, Michael Sammeth, Kristin G. Ardlie, and Roderic Guigó. The effects of death and post-mortem cold ischemia on human tissue transcriptomes. *Nature Communications*, 9(1):490, December 2018. Number: 1.
- [263] Marvin Martens, Ammar Ammar, Anders Riutta, Andra Waagmeester, Denise N Slenter, Kristina Hanspers, Ryan A. Miller, Daniela Digles, Elisson N Lopes, Friederike Ehrhart, Lauren J Dupuis, Laurent A Winckers, Susan L Coort, Egon L Willighagen, Chris T Evelo, Alexander R Pico, and Martina Kutmon. WikiPathways: connecting communities. *Nucleic Acids Research*, 49(D1):D613–D621, January 2021. Number: D1.
- [264] Maxim V. Kuleshov, Matthew R. Jones, Andrew D. Rouillard, Nicolas F. Fernandez, Qiaonan Duan, Zichen Wang, Simon Koplev, Sherry L. Jenkins, Kathleen M. Jagodnik, Alexander Lachmann, Michael G. McDermott, Caroline D. Monteiro, Gregory W. Gundersen, and Avi Ma’ayan. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, 44(Web Server issue):W90–W97, July 2016. Number: Web Server issue.
- [265] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13(11):2498–2504, November 2003. Number: 11.

- [266] Julia A. Gustavsen, Shraddha Pai, Ruth Isserlin, Barry Demchak, and Alexander R. Pico. RCy3: Network biology using Cytoscape from within R. *F1000Research*, 8:1774, December 2019.
- [267] Piotr J. Balwierz, Mikhail Pachkov, Phil Arnold, Andreas J. Gruber, Mihaela Zavolan, and Erik van Nimwegen. ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. *Genome Research*, 24(5):869–884, May 2014. Number: 5.
- [268] Zuguang Gu, Roland Eils, and Matthias Schlesner. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32(18):2847–2849, September 2016. Number: 18.
- [269] L. Le Cam. Maximum Likelihood: An Introduction. *International Statistical Review / Revue Internationale de Statistique*, 58(2):153–171, 1990. Publisher: [Wiley, International Statistical Institute (ISI)].
- [270] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association*, 112(518):859–877, April 2017. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/01621459.2017.1285773>.
- [271] Robert Bassett and Julio Deride. Maximum a posteriori estimators as a limit of Bayes estimators. *Mathematical Programming*, 174(1):129–144, March 2019.
- [272] K. J. Friston and W. Penny. Posterior probability maps and SPMs. *NeuroImage*, 19(3):1240–1249, July 2003.
- [273] Eli Bingham. Pyro.
- [274] Gioele La Manno, Kimberly Siletti, Alessandro Furlan, Daniel Gyllborg, Elin Vinsland, Alejandro Mossi Albiach, Christoffer Mattsson Langseth, Irina Khven, Alex R. Lederer, Lisa M. Dratva, Anna Johnsson, Mats Nilsson, Peter Lönnerberg, and Sten Linnarsson. Molecular architecture of the developing mouse brain. *Nature*, 596(7870):92–96, August 2021. Number: 7870 Publisher: Nature Publishing Group.
- [275] Yusuke Hirabayashi and Yukiko Gotoh. Stage-dependent fate determination of neural precursor cells in mouse forebrain. *Neuroscience Research*, 51(4):331–336, April 2005.

- [276] A. Caricasole, A. Bakker, A. Copani, F. Nicoletti, G. Gaviraghi, and G. C. Terstappen. Two Sides of the Same Coin: Wnt Signaling in Neurodegeneration and Neuro-Oncology. *Bioscience Reports*, 25(5-6):309–327, October 2005.
- [277] T. Takahashi, R. S. Nowakowski, and V. S. Caviness. The cell cycle of the pseudostratified ventricular epithelium of the embryonic murine cerebral wall. *Journal of Neuroscience*, 15(9):6046–6057, September 1995. Publisher: Society for Neuroscience Section: Articles.
- [278] J. Gray Camp, Farhath Badsha, Marta Florio, Sabina Kanton, Tobias Gerber, Michaela Wilsch-Bräuninger, Eric Lewitus, Alex Sykes, Wulf Hevers, Madeline Lancaster, Juergen A. Knoblich, Robert Lachmann, Svante Pääbo, Wieland B. Huttner, and Barbara Treutlein. Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proceedings of the National Academy of Sciences*, 112(51):15672–15677, December 2015. Publisher: Proceedings of the National Academy of Sciences.
- [279] NCI and NHL. The Cancer Genome Atlas Program, June 2018. Archive Location: nciglobal,ncicenterprise.
- [280] Melissa H. Madden, Gabriella M. Anic, Reid C. Thompson, L. Burton Nabors, Jeffrey J. Olson, James E. Browning, Alvaro N. Monteiro, and Kathleen M. Egan. Circadian pathway genes in relation to glioma risk and outcome. *Cancer Causes & Control*, 25(1):25–32, January 2014.
- [281] Mengjun Qiu, Yao-bing Chen, Si Jin, Xie-fan Fang, Xiao-xiao He, Zhi-fan Xiong, and Sheng-li Yang. Research on circadian clock genes in non-small-cell lung carcinoma. *Chronobiology International*, 36(6):739–750, June 2019. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/07420528.2018.1509080>.
- [282] Meng-jun Qiu, Li-ping Liu, Si Jin, Xie-fan Fang, Xiao-xiao He, Zhi-fan Xiong, and Sheng-li Yang. Research on circadian clock genes in common abdominal malignant tumors. *Chronobiology International*, 36(7):906–918, July 2019. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/07420528.2018.1477792>.
- [283] Dongying Gu, Shuwei Li, Shuai Ben, Mulong Du, Haiyan Chu, Zhengdong Zhang, Meilin Wang, Zuo-Feng Zhang, and Jinfei Chen. Circadian clock pathway genes associated with colorectal cancer risk and prognosis. *Archives of Toxicology*, 92(8):2681–2689, August 2018.

Bibliography

- [284] E. Reszka and M. Przybek. Chapter Three - Circadian Genes in Breast Cancer. In Gregory S. Makowski, editor, *Advances in Clinical Chemistry*, volume 75, pages 53–70. Elsevier, January 2016.
- [285] Chunhui Yi, Lina Mu, Irene A. Rigault de la Longrais, Olga Sochirca, Riccardo Arisio, Herbert Yu, Aaron E. Hoffman, Yong Zhu, and Dionyssios Katsaro. The circadian gene NPAS2 is a novel prognostic biomarker for breast cancer. *Breast Cancer Research and Treatment*, 120(3):663–669, April 2010.
- [286] Cristina Cadenas, Leonie van de Sandt, Karolina Edlund, Miriam Lohr, Birte Hellwig, Rosemarie Marchan, Marcus Schmidt, Jörg Rahnenführer, Henrik Oster, and Jan G Hengstler. Loss of circadian clock gene expression is associated with tumor progression in breast cancer. *Cell Cycle*, 13(20):3282–3291, October 2014. Publisher: Taylor & Francis _eprint: <https://doi.org/10.4161/15384101.2014.954454>.
- [287] Gang Wu, Lauren J Francey, Marc D Ruben, and John B Hogenesch. Normalized coefficient of variation (nCV): a method to evaluate circadian clock robustness in population scale data. *Bioinformatics*, 37(23):4581–4583, December 2021.
- [288] Leon A. Adams, Paul Angulo, and Keith D. Lindor. Nonalcoholic fatty liver disease. *CMAJ*, 172(7):899–905, March 2005. Publisher: CMAJ Section: Review.
- [289] Jeanne M. Clark, Frederick L. Brancati, and Anna Mae Diehl. Nonalcoholic fatty liver disease. *Gastroenterology*, 122(6):1649–1657, May 2002.
- [290] Paul Angulo. Nonalcoholic Fatty Liver Disease. *New England Journal of Medicine*, 346(16):1221–1231, April 2002. Publisher: Massachusetts Medical Society _eprint: <https://doi.org/10.1056/NEJMra011775>.
- [291] Mary E. Rinella. Nonalcoholic Fatty Liver Disease: A Systematic Review. *JAMA*, 313(22):2263–2273, June 2015.
- [292] Anand R. Saran, Shravan Dave, and Amir Zarrinpar. Circadian Rhythms in the Pathogenesis and Treatment of Fatty Liver Disease. *Gastroenterology*, 158(7):1948–1966.e1, May 2020.
- [293] Davide Gnocchi, Carlo Custodero, Carlo Sabbà, and Antonio Mazzocca. Circadian rhythms: a possible new player in non-alcoholic fatty liver disease pathophysiology. *Journal of Molecular Medicine*, 97(6):741–759, June 2019.

- [294] Hans Reinke and Gad Asher. Circadian Clock Control of Liver Metabolic Functions. *Gastroenterology*, 150(3):574–580, March 2016.
- [295] Akshay Shetty, Jennifer W. Hsu, Paul P. Manka, and Wing-Kin Syn. Role of the Circadian Clock in the Metabolic Syndrome and Nonalcoholic Fatty Liver Disease. *Digestive Diseases and Sciences*, 63(12):3187–3206, December 2018.
- [296] Gianluigi Mazzocchi, Manlio Vinciguerra, Jude Oben, Roberto Tarquini, and Salvatore De Cosmo. Non-alcoholic fatty liver disease: the role of nuclear receptors and circadian rhythmicity. *Liver International*, 34(8):1133–1152, 2014. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/liv.12534>.
- [297] Dongmei Shi, Jie Chen, Jiaofeng Wang, Jianfeng Yao, Yiqin Huang, Gansheng Zhang, and Zhijun Bao. Circadian Clock Genes in the Metabolism of Non-alcoholic Fatty Liver Disease. *Frontiers in Physiology*, 10, 2019.
- [298] Nicole M. Kettner, Horatio Voicu, Milton J. Finegold, Cristian Coarfa, Arun Sreekumar, Nagireddy Putluri, Chinenye A. Katchy, Choogon Lee, David D. Moore, and Loning Fu. Circadian Homeostasis of Liver Metabolism Suppresses Hepatocarcinogenesis. *Cancer Cell*, 30(6):909–924, December 2016.
- [299] Stefano Ballestri, Fabio Nascimbeni, Enrica Baldelli, Alessandra Marrazzo, Dante Romagnoli, and Amedeo Lonardo. NAFLD as a Sexual Dimorphic Disease: Role of Gender and Reproductive Status in the Development and Progression of Nonalcoholic Fatty Liver Disease and Inherent Cardiovascular Risk. *Advances in Therapy*, 34(6):1291–1326, June 2017.
- [300] Zeyneb Kurt, Rio Barrere-Cain, Jonnby LaGuardia, Margarete Mehrabian, Calvin Pan, Simon T. Hui, Frode Norheim, Zhiqiang Zhou, Yehudit Hasin, Aldons J. Lusis, and Xia Yang. Tissue-specific pathways and networks underlying sexual dimorphism in non-alcoholic fatty liver disease. *Biology of Sex Differences*, 9(1):46, October 2018.
- [301] Philippe Lefebvre and Bart Staels. Hepatic sexual dimorphism — implications for non-alcoholic fatty liver disease. *Nature Reviews Endocrinology*, 17(11):662–670, November 2021. Number: 11 Publisher: Nature Publishing Group.
- [302] Patrizia Burra, Debora Bizzaro, Anna Gonta, Sarah Shalaby, Martina Gambato, Maria Cristina Morelli, Silvia Trapani, Annarosa Floreani, Fabio Marra, Maurizia Rossana Brunetto, Gloria Taliani, Erica Villa, and the Special Interest Group Gender in Hepatology of the Italian Association

- for the Study of the Liver (AISF). Clinical impact of sexual dimorphism in non-alcoholic fatty liver disease (NAFLD) and non-alcoholic steatohepatitis (NASH). *Liver International*, 41(8):1713–1733, 2021. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/liv.14943>.
- [303] Matthew M Yeh, Xiuhui Shi, Jingxuan Yang, Min Li, Kar-Ming Fung, and Sayed S. Daoud. Perturbation of Wnt/-catenin signaling and sexual dimorphism in non-alcoholic fatty liver disease. *Hepatology Research*, 52(5):433–448, 2022. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/hepr.13754>.
- [304] Amedeo Lonardo and Ayako Suzuki. Sexual Dimorphism of NAFLD in Adults. Focus on Clinical Aspects and Implications for Practice and Translational Research. *Journal of Clinical Medicine*, 9(5):1278, April 2020.
- [305] Chien-Hua Chen, Min-Ho Huang, Jee-Chun Yang, Chiu-Kue Nien, Chi-Chieh Yang, Yung-Hsiang Yeh, and Sen-Kou Yueh. Prevalence and Risk Factors of Nonalcoholic Fatty Liver Disease in an Adult Population of Taiwan: Metabolic Significance of Nonalcoholic Fatty Liver Disease in Nonobese Adults. *Journal of Clinical Gastroenterology*, 40(8):745–752, September 2006.
- [306] Seung H Park, Woo K Jeon, Sang H Kim, Hong J Kim, Dong I Park, Yong K Cho, In K Sung, Chong I Sohn, Dong K Keum, and Byung I Kim. Prevalence and risk factors of non-alcoholic fatty liver disease among Korean adults. *Journal of Gastroenterology and Hepatology*, 21(1):138–143, 2006. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1440-1746.2005.04086.x>.
- [307] Llorenç Caballería, Guillem Pera, Maria Antònia Auladell, Pere Torán, Laura Muñoz, Dolores Miranda, Alba Alumà, José Dario Casas, Carmen Sánchez, Dolors Gil, Josep Aubà, Albert Tibau, Santiago Canut, Jesús Bernad, and Miren Maite Aizpurua. Prevalence and factors associated with the presence of nonalcoholic fatty liver disease in an adult population in Spain. *European Journal of Gastroenterology & Hepatology*, 22(1):24–32, January 2010.
- [308] Jian-Gao Fan, Jun Zhu, Xing-Jian Li, Lan Chen, Lui Li, Fei Dai, Feng Li, and Shi-Yao Chen. Prevalence of and risk factors for fatty liver in a general population of Shanghai, China. *Journal of Hepatology*, 43(3):508–514, September 2005.
- [309] Chengfu Xu, Chaohui Yu, Han Ma, Lei Xu, Min Miao, and Youming Li. Prevalence and Risk Factors for the Development of Nonalcoholic Fatty Liver Disease in a Nonobese Chinese Population: the Zhejiang Zhenhai Study. *Official*

- journal of the American College of Gastroenterology* | ACG, 108(8):1299–1304, August 2013.
- [310] Yuichiro Eguchi, Hideyuki Hyogo, Masafumi Ono, Toshihiko Mizuta, Naofumi Ono, Kazuma Fujimoto, Kazuaki Chayama, Toshiji Saibara, and JSG-NAFLD. Prevalence and associated metabolic factors of nonalcoholic fatty liver disease in the general population from 2009 to 2010 in Japan: a multicenter large retrospective study. *Journal of Gastroenterology*, 47(5):586–595, May 2012.
 - [311] Zobair M. Younossi, Maria Stepanova, Francesco Negro, Shareh Hallaji, Youssef Younossi, Brian Lam, and Manirath Srishord. Nonalcoholic Fatty Liver Disease in Lean Individuals in the United States. *Medicine*, 91(6):319–327, November 2012.
 - [312] Marisa Chiloiro, Maria Gabriella Caruso, Anna Maria Cisternino, Rosa Inguaggiato, Rosa Reddavid, Caterina Bonfiglio, Vito Guerra, Maria Notarnicola, Giampietro De Michele, Mario Correale, Maria Rosaria Noviello, and Giovanni Misciagna. Ultrasound Evaluation and Correlates of Fatty Liver Disease: A Population Study in a Mediterranean Area. *Metabolic Syndrome and Related Disorders*, 11(5):349–358, October 2013. Publisher: Mary Ann Liebert, Inc., publishers.
 - [313] Temitope Foster, Frank A. Anania, Dong Li, Ronit Katz, and Matthew Budoff. The Prevalence and Clinical Correlates of Nonalcoholic Fatty Liver Disease (NAFLD) in African Americans: The Multiethnic Study of Atherosclerosis (MESA). *Digestive Diseases and Sciences*, 58(8):2392–2398, August 2013.
 - [314] Naoto Fujiwara, Naoto Kubota, Emilie Crouch, Bhuvaneswari Koneru, Cesia A. Marquez, Arun K. Jajoriya, Gayatri Panda, Tongqi Qian, Shijia Zhu, Nicolas Goossens, Xiaochen Wang, Shuang Liang, Zhenyu Zhong, Sara Lewis, Bachir Taouli, Myron E. Schwartz, Maria Isabel Fiel, Amit G. Singal, Jorge A. Marrero, Austin J. Fobar, Neehar D. Parikh, Indu Raman, Quan-Zhen Li, Masataka Taguri, Atsushi Ono, Hiroshi Aikata, Takashi Nakahara, Hayato Nakagawa, Yuki Matsushita, Ryosuke Tateishi, Kazuhiko Koike, Masahiro Kobayashi, Takaaki Higashi, Shigeki Nakagawa, Yo-ichi Yamashita, Toru Beppu, Hideo Baba, Hiromitsu Kumada, Kazuaki Chayama, Thomas F. Baumert, and Yujin Hoshida. Molecular signatures of long-term hepatocellular carcinoma risk in nonalcoholic fatty liver disease. *Science Translational Medicine*, 14(650):eabo4474, June 2022. Publisher: American Association for the Advancement of Science.

- [315] Olivier Govaere, Simon Cockell, Dina Tiniakos, Rachel Queen, Ramy Younes, Michele Vacca, Leigh Alexander, Federico Ravaioli, Jeremy Palmer, Salvatore Petta, Jerome Boursier, Chiara Rosso, Katherine Johnson, Kristy Wonders, Christopher P. Day, Mattias Ekstedt, Matej Orešič, Rebecca Darlay, Heather J. Cordell, Fabio Marra, Antonio Vidal-Puig, Pierre Bedossa, Jörn M. Schattenberg, Karine Clément, Michael Allison, Elisabetta Bugianesi, Vlad Ratziu, Ann K. Daly, and Quentin M. Anstee. Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. *Science Translational Medicine*, 12(572):eaba4448, December 2020. Publisher: American Association for the Advancement of Science.
- [316] Tongqi Qian, Naoto Fujiwara, Bhuvaneswari Koneru, Atsushi Ono, Naoto Kubota, Arun K. Jajoriya, Matthew G. Tung, Emilie Crouchet, Won-Min Song, Cesia Ammi Marquez, Gayatri Panda, Ayaka Hoshida, Indu Raman, Quan-Zhen Li, Cheryl Lewis, Adam Yopp, Nicole E. Rich, Amit G. Singal, Shigeki Nakagawa, Nicolas Goossens, Takaaki Higashi, Anna P. Koh, C. Billie Bian, Hiroki Hoshida, Parissa Tabrizian, Ganesh Gunasekaran, Sander Florman, Myron E. Schwarz, Spiros P. Hiotis, Takashi Nakahara, Hiroshi Aikata, Eisuke Murakami, Toru Beppu, Hideo Baba, rew Warren, Sangeeta Bhatia, Masahiro Kobayashi, Hiromitsu Kumada, Austin J. Fobar, Neehar D. Parikh, Jorge A. Marrero, Steve Hategekimana Rwema, Venugopalan Nair, Manishkumar Patel, Seunghye Kim-Schulze, Kathleen Corey, Jacqueline G. O’Leary, Goran B. Klintmalm, David L. Thomas, Mohammed Dibas, Gerardo Rodriguez, Bin Zhang, Scott L. Friedman, Thomas F. Baumert, Bryan C. Fuchs, Kazuaki Chayama, Shijia Zhu, Raymond T. Chung, and Yujin Hoshida. Molecular Signature Predictive of Long-Term Liver Fibrosis Progression to Inform Antifibrotic Drug Development. *Gastroenterology*, 162(4):1210–1225, April 2022.
- [317] Oleg V. Grinchuk, Surya P. Yenamandra, Ramakrishnan Iyer, Malay Singh, Hwee Kuan Lee, Kiat Hon Lim, Pierce Kah-Hoe Chow, and Vladamir A. Kuznetsov. Tumor-adjacent tissue co-expression profile analysis reveals pro-oncogenic ribosomal gene signature for prognosis of resectable hepatocellular carcinoma. *Molecular Oncology*, 12(1):89–113, 2018. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/1878-0261.12153>.
- [318] Lorena Pantano, George Agyapong, Yang Shen, Zhu Zhuo, Francesc Fernandez-Albert, Werner Rust, Dagmar Knebel, Jon Hill, Carine M. Boustany-Kari, Julia F. Doerner, Jörg F. Rippmann, Raymond T. Chung, Shannan J.

- Ho Sui, Eric Simon, and Kathleen E. Corey. Molecular characterization and cell type composition deconvolution of fibrosis in NAFLD. *Scientific Reports*, 11(1):18045, September 2021. Number: 1 Publisher: Nature Publishing Group.
- [319] Ravi Dhingra and Ramachandran S. Vasan. Age as a Cardiovascular Risk Factor. *The Medical Clinics of North America*, 96(1):87–91, January 2012. Number: 1.
- [320] Allan D. Sniderman and Curt D. Furberg. Age as a modifiable risk factor for cardiovascular disease. *The Lancet*, 371(9623):1547–1549, May 2008. Publisher: Elsevier.
- [321] Luca Liberale, Fabrizio Montecucco, Jean-Claude Tardif, Peter Libby, and Giovanni G Camici. Inflamm-ageing: the role of inflammation in age-dependent cardiovascular disease. *European Heart Journal*, 41(31):2974–2982, August 2020.
- [322] Frank B. Hu, Francine Grodstein, Charles H. Hennekens, Graham A. Colditz, Michelle Johnson, JoAnn E. Manson, Bernard Rosner, and Meir J. Stampfer. Age at Natural Menopause and Risk of Cardiovascular Disease. *Archives of Internal Medicine*, 159(10):1061–1066, May 1999.
- [323] Edward G. Lakatta. Age-associated Cardiovascular Changes in Health: Impact on Cardiovascular Disease in Older Persons. *Heart Failure Reviews*, 7(1):29–49, January 2002.
- [324] Dermot F. Reilly, Elizabeth J. Westgate, and Garret A. FitzGerald. Peripheral Circadian Clocks in the Vasculature. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 27(8):1694–1705, August 2007. Number: 8 Publisher: American Heart Association.
- [325] Björn Lemmer. Chronopharmacology of cardiovascular medications. *Biological Rhythm Research*, 38(3):247–258, June 2007. Number: 3.
- [326] Verónica Guarner-Lans, María Esther Rubio-Ruiz, Israel Pérez-Torres, and Guadalupe Baños de MacCarthy. Relation of aging and sex hormones to metabolic syndrome and cardiovascular disease. *Experimental Gerontology*, 46(7):517–523, July 2011.
- [327] Vera Regitz-Zagrosek and Georgios Kararigas. Mechanistic Pathways of Sex Differences in Cardiovascular Disease. *Physiological Reviews*, 97(1):1–37, January 2017. Publisher: American Physiological Society.

Bibliography

- [328] Myoung H. Lee, Evan Vosburgh, Kristen Anderson, and Jan McDonagh. Deficiency of Plasma Plasminogen Activator Inhibitor 1 Results in Hyperfibrinolytic Bleeding. *Blood*, 81(9):2357–2362, May 1993. Number: 9.
- [329] Roman Szabo, Sarah Netzel-Arnett, John P. Hobson, Toni M. Antalis, and Thomas H. Bugge. Matriptase-3 is a novel phylogenetically preserved membrane-anchored serine protease with broad serpin reactivity. *Biochemical Journal*, 390(Pt 1):231–242, August 2005. Number: Pt 1.
- [330] P.e. Morange, N. Saut, M.c. Alessi, J.s. Yudkin, M. Margaglione, G. Di Minno, A. Hamsten, S.e. Humphries, D.a. Tregouet, and I. Juhan-Vague. Association of Plasminogen Activator Inhibitor (PAI)-1 (SERPINE1) SNPs With Myocardial Infarction, Plasma PAI-1, and Metabolic Parameters. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 27(10):2250–2257, October 2007. Publisher: American Heart Association.
- [331] Werner Koch, Matthias Schrempf, Anna Erl, Jakob C. Mueller, Petra Hoppmann, Albert Schömig, and Adnan Kastrati. 4G/5G polymorphism and haplotypes of SERPINE1 in atherosclerotic diseases of coronary arteries. *Thrombosis and Haemostasis*, 103(6):1170–1180, 2010. Publisher: Schattauer GmbH.
- [332] Gianluigi Mazzocchi, Valerio Paziienza, Anna Panza, Maria Rosa Valvano, Giorgia Benegiamo, Manlio Vinciguerra, Angelo Andriulli, and Ada Piepoli. ARNTL2 and SERPINE1: potential biomarkers for tumor aggressiveness in colorectal cancer. *Journal of Cancer Research and Clinical Oncology*, 138(3):501–511, March 2012.
- [333] Miguel Angel Pavón, Irene Arroyo-Solera, Maria Virtudes Céspedes, Isolda Casanova, Xavier León, and Ramón Mangués. uPA/uPAR and SERPINE1 in head and neck cancer: role in tumor resistance, metastasis, prognosis and therapy. *Oncotarget*, 7(35):57351–57366, June 2016. Publisher: Impact Journals.
- [334] Chunyan Zeng and Youxiang Chen. HTR1D, TIMP1, SERPINE1, MMP3 and CNR2 affect the survival of patients with colon adenocarcinoma. *Oncology Letters*, 18(3):2448–2454, September 2019. Publisher: Spandidos Publications.
- [335] Sadiya S. Khan, Sanjiv J. Shah, Ekaterina Klyachko, Abigail S. Baldrige, Mesut Eren, Aaron T. Place, Abraham Aviv, Eli Puterman, Donald M. Lloyd-Jones, Meadow Heiman, Toshio Miyata, Sweta Gupta, Amy D. Shapiro, and Douglas E. Vaughan. A null mutation in SERPINE1 protects against biological

- aging in humans. *Science Advances*, 3(11):eaao1617, November 2017. Publisher: American Association for the Advancement of Science.
- [336] A E Reinberg. Concepts in Chronopharmacology. *Annual Review of Pharmacology and Toxicology*, 32(1):51–66, 1992. _eprint: <https://doi.org/10.1146/annurev.pa.32.040192.000411>.
- [337] A Reinberg and F Halberg. Circadian Chronopharmacology. *Annual Review of Pharmacology*, 11(1):455–492, 1971. _eprint: <https://doi.org/10.1146/annurev.pa.11.040171.002323>.
- [338] Alain Reinberg. Clinical Chronopharmacology. In Alain Reinberg and Michael H. Smolensky, editors, *Biological Rhythms and Medicine*, Topics in Environmental Physiology and Medicine, pages 211–263, New York, NY, 1983. Springer.
- [339] M. Hassan, G. Öberg, A. N. Bekassy, J. Aschan, H. Ehrsson, P. Ljungman, G. Lönnerholm, B. Smedmyr, A. Taube, I. Wallin, and B. Simonsson. Pharmacokinetics of high-dose busulphan in relation to age and chronopharmacology. *Cancer Chemotherapy and Pharmacology*, 28(2):130–134, March 1991.
- [340] Yu Tahara and Shigenobu Shibata. Chrono-biology, Chrono-pharmacology, and Chrono-nutrition. *Journal of Pharmacological Sciences*, 124(3):320–335, 2014.
- [341] Megerle L. Scherholz, Naomi Schlesinger, and Ioannis P. Androulakis. Chronopharmacology of glucocorticoids. *Advanced Drug Delivery Reviews*, 151-152:245–261, November 2019.
- [342] Björn Lemmer. Chronopharmacology and controlled drug release. *Expert Opinion on Drug Delivery*, 2(4):667–681, July 2005. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1517/17425247.2.4.667>.

CONTACT

✉ lorenzo.talamanca@gmail.com

in Lorenzo Talamanca

☎ +41 78 761 74 11

SKILLS

Statistics 6

Modelling 6

Data analysis 6

Bio-informatics 5.75

PROGRAMMING

R/Rstudio 6

Python 5.5

SQL 5.25

Excel 4.75

SOFT SKILLS

Communication 5.75

Self-starter 5.5

Teamwork 5.25

LANGUAGES

Italian 6

English 5.75

LORENZO TALAMANCA

Theoretical physicist, data scientist, bioinformatician

EDUCATION

Ph. D. - Biophysics & Data Analysis
EPFL - Lausanne, Switzerland

2018 - 2023

Advisors: Felix Naef, Paolo De Los Rios

Quantitative biophysics, Data analysis, Modelling, Programming, Data visualization.

Master degree - Theoretical and statistical physics
La Sapienza - Rome, Italy

2016 - 2018

110 cum laude/110. Advisor: Giorgio Parisi (Physics nobel prize 2021);
Theoretical physics, Probability, Statistics, Modelling, Programming.

Bachelor degree- Physics
La Sapienza - Rome, Italy

2013 - 2016

110 cum laude/110. Advisor: Stefano Giagu;
Physics, Statistics, Data Analysis, Programming.

WORK EXPERIENCE

PhD student
EPFL - Lausanne, Switzerland

Sep 18 - ongoing

Develop my PhD project, structure it and make it relevant for the scientific community. Supervise master and bachelor projects as well as semester projects in the lab

Teaching assistant
EPFL - Lausanne, Switzerland

Sep 18 - Jun 22

Write and correct exercises, exams, and course material, direct a team of student assistants for "Mathematical modelling in biology" and "Analytical mechanics".

PUBLICATIONS

Sex-dimorphic and age-dependent organization of
24 hour gene expression rhythms in human
Re-submitted after minor revisions, *Science*

2022

Authors: L.Talamanca, C. Gobet, and F. Naef

How to tell time: advances in decoding circadian
phase from omics snapshots
F1000Research, 2020

2020

Authors: L.Talamanca and F. Naef

ACHIEVEMENTS

Poster prize

XY models in chronobiology

Theoretical Biophysics summer school, University of Colorado Boulder, 2019

TOEFL

108/120

Obtained in 2018:
reading, listening 30/30;
writing, speaking 24/30.

GRE

327/340

Obtained in 2018:
Verbal reasoning 158/170;
Quantitative reasoning 169/170.

National math olimpiad

Bronze medal winner

2013: Single bronze medal, Team silver medal
2012: Single finalist, Team bronze medal
2009-2011: Single & Team finalist

Premio Cesare Bonacini

Magnus effect in football

Awarded in 2011: nation-wide prize to best combination of experimental and theoretical physics at high school level.

Volleyball

U17 male regional volleyball league winner
Volleyball league in a region, with about 6M inhabitants

Study of longitudinal fluctuations of the Sherrington-Kirkpatrick model
Journal of Statistical Mechanics: Theory and Experiment, 2019

Authors: L.Talamanca, L. Sarra, and **G. Parisi**

PROJECTS

Percorso di eccellenza: magistrale

Master degree excellence path (5% admission rate)

Extra courses and projects on Spin glasses, Quantum field theory, Stochastic processes

Percorso di eccellenza: triennale

Bachelor degree excellence path (2% admission rate)

Extra courses and projects on Differential geometry, Classical electrodynamics, Quantum mechanics, Probability, High energy particles

WORKSHOPS & CONFERENCES

Life Science Symposium
EPFL, Lausanne

Biological oscillators: design, mechanism, function
EMBL, Heidelberg

Single Cell Biology
Virtual Keystone Symposia

Theoretical Biophysics summer school
University of Colorado Boulder

Mathematics Research School on Lattices and Application to Cryptography and Coding Theory
CIMPA-ICTP

PRESENTATIONS

The philosophy of probability
EPFL

Inferring internal time and determining circadian health
EPFL

EXTRACURRICULAR

- I played **volleyball** for many years, at high levels in high school.
- I am an avid **pen collector**.

2019

2017-2018

2014-2016

Mar 2022

Mar 2022

Mar 2021

Aug 2019

Aug 2016

Oct 2021

Oct 2019