UNIVERSITÀ
DEGLI STUDI
DI PADOVA

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA INDUSTRIALE
SCUOLA DI DOTTORATO DI RICERCA IN INGEGNERIA INDUSTRIALE
CURRICULUM INGEGNERIA CHIMICA ED AMBIENTALE
CICLO XXX

Tesi di Dottorato in
Ingegneria Industriale

CIRCADIAN CLOCK STUDY THROUGH FREQUENCY-ENCODED ENTRAINMENT STIMULATIONS

Direttore della Scuola: Ch.mo Prof. Paolo Colombo
Coordinatore di Curriculum: Ch.mo Ing. Matteo Strumendo
Supervisore: Ch.mo Prof. Nicola Elvassore

Dottoranda: SILVIA GALVANIN

Foreword

This Ph.D. Dissertation is written in partial fulfillment of requirements for the Ph.D. Course of Industrial Engineering, Curriculum Chemical and Environmental Engineering, at University of Padova. The work of this Ph.D. program was carried out at "Dipartimento di Ingegneria Industriale, sede M - Via Marzolo" of "Università degli Studi di Padova" and at "Venetian Institute of Molecular Medicine" of "Fondazione per la Ricerca Biomedica Avanzata ONLUS" in Padova, from November 1st, 2014 to October 31st, 2017. Part of the research work has been carried out at "Shanghai Institute of Advanced Immunochemical Studies" at ShanghaiTech University, Shanghai.

The research work has been supervised by Professor Nicola Elvassore, "Dipartimento di Ingegneria Industriale", "Università degli Studi di Padova".

The author would like to thank Professor Nicola Elvassore for his supervision throughout the entire research work and his enlightening discussions. My special thanks go to all present and past members of BioERA lab and to Dr. Pierantonio Facco, for their unconditioned contributions to my personal and professional growth.

All the material reported in this dissertation is original unless explicit references to studies carried out by other people are indicated.

During this Ph.D. program the following publication has been produced:

- Scattolini, V., Luni, C., Zambon, A., Galvanin, S., Gagliano, O., Ciubotaru, C. D., Avogaro, A., Mammano, F., Elvassore, N., Fadini, G. P. (2016). Simvastatin Rapidly and Reversibly Inhibits Insulin Secretion in Intact Single-Islet Cultures. Diabetes Therapy, 7(4), 679–693.

The following publication is under submission:

- Gagliano, O., Galvanin, S., Takahashi, J. S., Elvassore, N. Periodic Caloric Restriction Induces Anti-phase Circadian Resetting.

Part of this work has been presented at the following conferences:

- Galvanin, S., Torchio E., Gagliano, O., Elvassore, N. Development of a patient-specific chronotherapy model. Venetian Institute of Molecular Medicine 15th Annual Retreat. Preganziol (TV), Italy, February 18th-19th 2017.
- Gagliano, O., Galvanin, S., Elvassore, N. Interactions between metabolism and circadian rhythm on peripheral clocks. SAB visit of Venetian Institute of Molecular Medicine, Padova (PD), Italy, May 29th-31st 2016.
- Gagliano, O., Galvanin, S., Elvassore, N. Effects of cyclic dynamic perturbations on peripheral clocks. Venetian Institute of Molecular Medicine 14th Annual Retreat. Ponzano Veneto (TV), Italy, February 19th-20th 2016.
- Gagliano, O., Galvanin, S., Elvassore, N. Altering circadian rhythms in peripheral tissues by metabolic stimulations. Venetian Institute of Molecular Medicine 13th Annual Retreat. Preganziol (TV), Italy, February 6th-7th 2015.

Padova, October 31st 2017

Silvia Galvanin

Abstract

Circadian clock are intrinsic, time-tracking systems that enable organisms to maintain their physiological state and their synchrony with the 24-hour rotation of the Earth, by partitioning behavioural and metabolic processes according to time of day within each tissue. They are entrained to the external environment by light/dark cycles and by food timing, which act as clock synchronizers. Emerging evidence suggests that circadian regulation is intimately linked to metabolic homeostasis and that dysregulation of circadian rhythms can contribute to disease. Conversely, metabolic signals also feed back into the circadian system, modulating circadian gene expression and behaviour.

Conventional experimental approach of circadian clock *in vitro* studies is based on single-pulse stimulation of only one metabolite or hormone, while *in vivo* peripheral tissues are exposed to periodic oscillating stimuli of a large number of metabolites and hormones, whose variations are in most cases interconnected, as for example glucose and insulin. Moreover, only one or few clock genes are generally considered, while it is known that a large number of genes, thus biological processes, are under circadian regulation.

Therefore, this Ph.D. research work is aimed at the development of technologies and data analysis tools to investigate the entrainment of peripheral mammalian circadian clock to frequency-encoded metabolic stimuli, which well mimic physiological oscillations at which peripheral tissues are exposed *in vivo*. Technologies, and, more specifically, microtechnologies have been developed to investigate the effects of periodic metabolic entrainment, showing that in murine fibroblasts oscillatory periodic metabolic stimulations entrain the expression of *Per2*, one of the core genes of the circadian molecular mechanism. Moreover, it has been proven that only by metabolic oscillations it is possible to completely reset the phase of cell-autonomous clocks. In order to develop a physiological and pathological *in vitro* model, achieving a high spatio-temporal control of cell culture microenvironment, frequency-encoded perturbations have been automated in a newly designed microfluidic platform for circadian applications.

Finally, to broaden the description of genes expressed with a circadian temporal pattern, a new data analysis method has been proposed and characterized, that allows to identify circadian genes in whole transcriptome data, to group genes based on the phase of their expression, to visualize transcriptome data at a glance and clearly identifying modifications at the transcriptome level from one biological condition to another one.

Summary

Circadian clocks are intrinsic, self-sustained time-tracking systems that enable organisms to anticipate environmental changes and allow them to adapt their behaviour and physiology to the appropriate time of day, such as feeding behaviour, sleep-wake cycles, hormone levels and body temperature. Their key biological function is to maintain the physiological state of an organism and its alignment with the external environment.

In mammals, biological rhythms are established and maintained by a central pacemaker in the suprachiasmatic nucleus (SCN). SCN neurons are entrained by light, the most powerful zeitgeber (time-giver). The central SCN clock directs rhythms in a number of peripheral tissues, such as liver, heart and muscles, using several output cues, and thereby helps to synchronize the clock system. However, peripheral clocks are also entrained by extrinsic cues, specifically food intake, which operates as a powerful zeitgeber.

Recent findings are reinforcing the idea that there is a tightly coupled relationship between metabolic state and the clock and that metabolic entrainment is sufficient for synchronizing the clock of peripheral tissues, without affecting the phase of the SCN. *In vivo*, it has been shown also that peripheral circadian clocks can be entrained by cyclic oscillations of endocrine factors and metabolic cues associated with dietary regimens of food intake and timing. However, the mechanisms through which metabolites regulate peripheral circadian clocks and how peripheral circadian clocks feedback to the SCN are poorly understood.

In vitro circadian clock studies are characterized by single-pulse stimulations of endocrine factors or metabolites (for example glucose), but these *in vitro* models are not able to mimic physiological oscillations at which peripheral tissues are exposed *in vivo* because they consider the stimulus derived from a single pulse, as opposed to the oscillatory periodic intrinsic nature of environmental stimuli. Moreover, published works consider the effects of either insulin or glucose, while their variations are tightly interconnected in biological organisms.

Conventional techniques for *in vitro* circadian clock studies are bioluminescence imaging and quantitative gene expression analysis; the first one allows for continuous detection of only one of the core clock genes at the transcriptional level over time for around 4 days, while with the latter quantification of

few genes is possible with discrete sampling for 1 or 2 days. On the one hand, precise information about the considered genes is obtained with these techniques, on the other hand they do not allow to obtain information regarding other unknown biological processes that may be involved in that phenomena, preventing the possibility of discovering new correlations between biological pathways. This is particularly important in the case of circadian studies, since a very large number of biological processes are under circadian control and researchers continuously identify new clock-controlled genes.

Thanks to recent technological development and cost reduction in RNA sequencing, a more comprehensive approach for deeply understanding biological responses has started to be feasible. By sequencing, it is possible to quantify the expression of all genes (transcriptome), thus obtaining a global description of cellular responses. When samples are taken over time, it is possible to identify temporal patterns of gene expression, distinguishing, among others, genes that display a circadian expression pattern from genes that are constantly expressed. Several algorithms and software are available to assess if a gene has a circadian expression (Fisher's G-test, Jonckheere-Terpstra-Kendall algorithm, COSOPT, ARSER, etc.); by the use of these algorithms it is possible to assess if a gene is significantly expressed in a circadian way. Some algorithms also estimate the characteristics of the oscillation, i.e. amplitude, phase and period, mostly by fitting temporal profiles to a cosine curve. However, algorithms discriminate between rhythmic and non-rhythmic transcripts and describe the properties of the oscillatory behaviour gene by gene, without globally visualizing transcriptome and showing which genes have the same phase at whole gene expression level.

The limitations of current approaches are particularly important in the case of circadian studies considering that physiological circadian entrainment stimuli are oscillatory and periodic and that circadian clock regulates a great number of biological processes (the actual percentage of circadian genes is different among different tissues). Therefore, appropriate technologies and data analysis methods are required to address these issues.

The aim of the Ph.D. research is the development of technologies and data analysis tools to investigate the entrainment of peripheral mammalian circadian clock to frequency-encoded metabolic stimuli. Technologies, and, more specifically, microtechnologies are developed in order to produce oscillating stimulations on biological systems and monitor their dynamic biological response for long-term. This type of perturbation can well mimic periodic dynamic fluctuations of metabolites and hormones in living organisms; thus, with this technology, quantification of biological responses consequent to physiological and pathological stimuli is possible. Data analysis methods are developed to identify circadian genes in whole transcriptome data, to cluster genes based on the phase of their expression, to visualize globally transcriptomic data and clearly identifying modifications at the transcriptome level from one biological condition to another one, or between organs or species.

In the research activities, microtechnologies have been developed to investigate the effects of periodic metabolic entrainment showing that oscillatory periodic metabolic stimulations entrain *Per2* expression in murine fibroblasts. *Per2* is one of the core clock genes belonging the circadian molecular network, highly expressed during the behaviourally active phase, corresponding to the subjective night for a murine model. Experimental investigation showed that only 12-hour frequency stimuli are able to sustain *Per2* expression *in vitro* and that anti-phase cyclic metabolic stimulations are able to completely reset fibroblasts clock, leading to a 12 h phase shift. This is relevant from a biological point of view, because it means that metabolic oscillations are able to synchronize the phase of peripheral clocks. Moreover, the key role of the starvation pathway has been identified, showing that starvation is necessary for circadian phase resetting.

Then, in order to develop an *in vitro* model that can resemble oscillations (physiological or pathological) at which peripheral tissues are exposed *in vivo*, a new automated microfluidic platform has been designed, tested and validated. This platform allows to perform frequency-encoded experiments with precise timing of the stimuli as well as with precise spatial control of the microenvironment inside culture chambers.

Finally, a data analysis method for transcriptomic time series has been developed to identify circadian genes in whole transcriptome data, to group genes based on the phase of their expression, to visualize globally transcriptome data and clearly identifying modifications at the transcriptome level from one biological condition to another one, that might be used to discover new pathways in the gene regulatory network.

The developed technologies and data analysis method could be applied to chronopharmacology, the branch of pharmacology that investigates efficacy and toxicity of drugs as a function of the administration time. The drug could be administered at different circadian time within the microfluidic platform in a highly controlled environment, and it effects at whole transcriptome level could be evaluated by the developed analysis method, allowing to identify which is the administration time that minimizes side effects and maximizes therapeutic outcome.

Sommario

I ritmi circadiani sono meccanismi biologici di organizzazione temporale intrinseci e autosostenuti, che consentono agli organismi di anticipare i cambiamenti ambientali e permettono loro di adattare il loro comportamento e la loro fisiologia nell'arco della giornata. Alcune funzioni regolate dall'orologio circadiano sono i cicli sonno/veglia, l'alimentazione, i livelli ormonali e la temperatura corporea. La funzione biologica essenziale del ritmo circadiano è mantenere lo stato fisiologico dell'organismo e la sua sincronia comportamentale e metabolica con l'ambiente esterno.

Nei mammiferi, il ritmo circadiano è scandito e mantenuto da un orologio biologico principale, anatomicamente localizzato nel nucleo soprachiasmatico. I neuroni del nucleo soprachiasmatico sono sincronizzati dallo stimolo luminoso, il più forte agente sincronizzante. Il ritmo biologico generato dal nucleo soprachiasmatico regola gli orologi circadiani presenti nei tessuti periferici, come ad esempio nel fegato, nel cuore e nei muscoli, tramite una modulazione neuroumorale; in questo modo l'orologio principale mantiene sincronizzato l'orologio a livello sistemico. Gli orologi periferici, però, sono stimolati anche da agenti sincronizzanti esterni, e in modo particolare dall'apporto di nutrienti.

Studi recenti hanno messo in evidenza la stretta interconnessione tra il metabolismo e l'orologio circadiano e il fatto che lo stimolo metabolico è di per sé sufficiente per sincronizzare il ritmo dei tessuti periferici, senza interferire con la fase dell'orologio principale. *In vivo* è stato dimostrato che gli orologi periferici possono essere sincronizzati con stimolazioni cicliche di fattori endocrini e metabolici, derivanti da diverse diete, in cui vengono variati sia l'apporto calorico sia l'ora dei pasti. Tuttavia, sono stati descritti solo parzialmente i meccanismi molecolari attraverso cui i metaboliti sincronizzano gli orologi periferici e attraverso cui gli orologi periferici mandano segnali all'orologio centrale.

Gli studi *in vitro* dell'orologio circadiano sono condotti con stimolazioni singole di fattori endocrini o metaboliti, come, ad esempio, il glucosio; questo tipo di studi però non simula le oscillazioni metaboliche fisiologiche a cui sono esposti i tessuti periferici *in vivo*: considerano, infatti, una singola stimolazione, mentre gli stimoli ambientali sono per loro stessa natura ciclici e oscillanti. Inoltre, in letteratura gli effetti di glucosio e insulina sono valutati separatamente, quando invece le loro variazioni fisiologiche sono altamente collegate.

Le tecniche convenzionali per lo studio *in vitro* dell'orologio circadiano sono la bioluminescenza

e l'analisi quantitativa dell'espressione genica; la prima permette di misurare in modo continuato nel tempo l'espressione di un unico gene circadiano per circa 4 giorni, mentre con la seconda è possibile quantificare l'espressione di alcuni geni in modo discontinuo nel tempo, per uno o due giorni. Queste tecniche convenzionali permettono di misurare l'espressione dei geni considerati in modo preciso, tuttavia non consentono di ottenere informazioni su altri processi biologici il cui ruolo non è ad oggi noto. Un approccio interamente basato sulle tecniche convenzionali limita quindi la possibilità di scoprire nuove correlazioni tra pathway o geni. Questo aspetto è particolarmente rilevante nel caso degli studi sull'orologio circadiano, dal momento che molti processi biologici sono controllati in modo circadiano e il numero di geni, già noti, di cui viene scoperta l'espressione circadiana è in continua crescita.

I recenti sviluppi tecnologici e la riduzione dei costi del sequenziamento dell'RNA hanno reso possibile un approccio più generale. Con il sequenziamento dell'RNA è infatti possibile quantificare l'espressione del trascrittoma, cioè di tutti i geni, ottenendo una descrizione globale delle risposte biologiche. Quando i campioni vengono raccolti nel tempo, è possibile inoltre identificare profili temporali di espressione genica, distinguendo, tra gli altri, geni che vengono espressi in modo circadiano da altri che sono invece espressi in modo costante nel tempo. Vari algoritmi e software sono stati sviluppati per stabilire se un gene è espresso in modo circadiano (Fisher's G-test, algoritmo Jonckheere-Terpstra-Kendall, COSOPT, ARSER, etc.); con questi algoritmi è possibile discriminare se un gene è espresso in modo circadiano con significatività statistica. Alcuni algoritmi stimano anche le caratteristiche dell'oscillazione, cioè ampiezza, fase e periodo, soprattutto mediante fitting dei profili temporali con un sinusoide. Gli algoritmi distinguono tra trascritti circadiani e non circadiani e descrivono le proprietà di ogni trascritto oscillante, ma non permettono di visualizzare il trascrittoma in modo comprensivo e globale, indicando quali geni hanno la stessa fase.

L'applicazione di un singolo stimolo e la quantificazione di uno o pochi geni nel tempo rappresentano quindi le due limitazioni principali dell'approccio convenzionale dello studio *in vitro* dell'orologio circadiano, considerando che gli stimoli che gli organismi ricevono dall'ambiente esterno sono oscillanti e periodici e che l'orologio circadiano regola un grande numero di processi biologici. Quindi, tecnologie e metodi di analisi dati sono richiesti per superare le limitazioni delle tecniche convenzionali.

Lo scopo del progetto di dottorato è lo sviluppo di tecnologie e di metodi di analisi dati per analizzare la sincronizzazione di orologi circadiani periferici a seguito di stimoli metabolici in frequenza. Tecnologie, e in modo specifico microtecnologie, sono sviluppate per realizzare stimolazioni oscillanti in frequenza su sistemi biologici e misurare la loro risposta dinamica per lungo tempo. Le stimolazioni in frequenza simulano le fluttuazioni dinamiche e periodiche dei metaboliti e degli ormoni a cui sono soggetti gli organismi; quindi, con questa tecnologia, è possibile quantificare le risposte biologiche a stimoli che simulano comportamenti fisiologici e patologici. Sono stati sviluppati metodi di analisi dati

per identificare geni circadiani nell'intero trascrittoma, per raggruppare geni espressi con la stessa fase, visualizzare globalmente il trascrittoma e identificare modificazioni a livello trascrizionale tra diverse condizioni biologiche, oppure tra diversi organi o specie.

Nell'attività di ricerca sono state sviluppate microtecnologie per analizzare gli effetti di stimolazioni periodiche e sperimentalmente si è dimostrato che stimolazioni metaboliche periodiche e oscillanti sono in grado di sincronizzare l'espressione del gene *Per2* in fibroblasti murini. Il gene *Per2* ha un ruolo centrale nel meccanismo molecolare dell'orologio circadiano; esso è espresso durante la fase attiva, che corrisponde alla notte nel modello murino. Si è inoltre dimostrato che solo stimolazioni con frequenza di 12 ore danno espressione sostenuta di *Per2 in vitro* e che stimolazioni metaboliche cicliche in antifase causano un completo reset dell'orologio periferico dei fibroblasti, mostrando uno spostamento di fase di 12 ore. Questo è estremamente rilevante dal punto di vista biologico perchè significa che gli stimoli metabolici sono in grado di sincronizzare la fase degli orologi periferici. Inoltre, si è dimostrato sperimentalmente il ruolo chiave e necessario della mancanza di nutrienti per avere un cambiamento di fase dell'orologio periferico.

In seguito, per sviluppare un modello *in vitro* per simulare le oscillazioni sia fisiologiche che patologiche a cui i tessuti periferici sono soggetti *in vivo*, una nuova piattaforma microfluidica è stata progettata, testata e validata. La piattaforma permette di realizzare stimolazioni in frequenza con elevata precisione temporale nelle stimolazioni e preciso controllo spaziale della composizione biochimica del microambiente delle camere di coltura.

Infine, un metodo di analisi dati per serie temporali di dati di trascrittomica è stato sviluppato per identificare geni espressi in modo circadiano, raggruppare i geni in base alla fase della loro espressione, visualizzare globalmente il trascrittoma e identificare modifiche a livello trascrizionale tra diverse condizioni biologiche; queste possono essere utilizzate per scoprire nuovi pathway nella mappa di interazione genica.

Le tecnologie ed il metodo di analisi dati sviluppati possono essere applicati al campo della cronofarmacologia, il ramo della farmacologia che studia l'efficacia e la tossicità dei farmaci in relazione all'ora in cui vengono somministrati. Il farmaco potrebbe essere somministrato a tempi diversi nella piattaforma microfluidica, in un ambiente altamente controllato, e i suoi effetti a livello trascrizionale potrebbero essere valutati con il metodo di analisi sviluppato, permettendo di individuare il momento di somministrazione che minimizza gli effetti collaterali e massimizza l'efficacia terapeutica.

Contents

List of Figures

List of Tables

Nomenclature

Roman Symbols

I	Number of samples in **X**
K	Number of replicates in **X**
N	Number of variables in **X**
X	Dataset

Abbreviations

AR	Armonic Regression
BP	Biological Processes
CC	Cellular Component
CCD	Charge-Coupled Device
CCGs	Clock Controlled Genes
C. I.	Confidence Interval
DD	Dark/dark
DMEM	Dulbecco Modified Eagle Medium
FBS	Fetal Bovine Serum
FDR	False Discovery Rate
FRET	Fluorescence Resonance Energy Transfer
FR	Free Running
GEO	Gene Expression Omnibus

GO	Gene Ontology
H	High Glucose-High Insulin medium
iKO	Inducible knock-out
JT	Jonckheere-Terpstra
JTK	Jonckheere-Terpstra-Kendall
KO	Knock-out
LD	Light/dark
L	Low Glucose-Low Insulin medium
LV	Latent Variable
LVM	Latent Variable Model
MF	Molecular Function
mLSI	Microfluidic large-scale integration
NIPALS	Non-linear iterative partial least-squares algorithm
PCA	Principal Component Analysis
PC	Principal Component
PCR	Polymerase Chain Reaction
PDMS	Polydimethylsiloxane
qrtPCR	Quantitative Real-Time Polymerase Chain Reaction
ROI	Region of Interest
SCN	Suprachiasmatic Nucleus
SNV	Standard Normal Variate
SSR	Sum of Squared Residuals
STAMP	Single-cell Transcriptomes Attached to Microparticle
SVD	Singular Value Decomposition

TTFL	Transcription-Translation Feedback Loops
UMI	Unique Molecular Identifier
WT	Wild Type

Chapter 1

Introduction

This Chapter describes physiology, mechanism and structure of circadian clock, highlighting its biological significance and relevance for maintenance of organism health, as well as current technology to study circadian clock. It focuses on the feedback loop between circadian clock and metabolism and it presents the motivation and the aim of this Dissertation.

1.1 The circadian clock

This Section presents the circadian clock system, describing its significance, its systemic and molecular architecture and consequences of circadian clock misalignments.

1.1.1 The clock physiology

An intrinsic molecular clock coordinates mammalian physiology and behaviour into rhythms that are synchronized with the 24-hour rotation of the Earth. These processes, defined as circadian rhythms, from the Latin *circa diem* meaning about a day, are intrinsic, time-tracking systems that enable organisms to anticipate environmental changes (Bass and Takahashi, 2010; Partch et al., 2014; Sahar and Sassone-Corsi, 2009).

Circadian clock machinery is a very complex and finely tuned system, which can be schematized into three parts (Figure 1.1): an input pathway that includes detectors to receive environmental cues (or zeitgebers) and transmit them to the central oscillator, a central oscillator that keeps circadian time and generates rhythm, and output pathways through which the rhythms are manifested by the control of various metabolic, physiological and behavioural processes (Bellet et al., 2011; Sahar and Sassone-Corsi, 2009). Output signals, such as metabolites or hormones, may feed back to the central oscillator, to close the loop of the regulatory mechanism.

Circadian clocks have distinguishing features; they are:

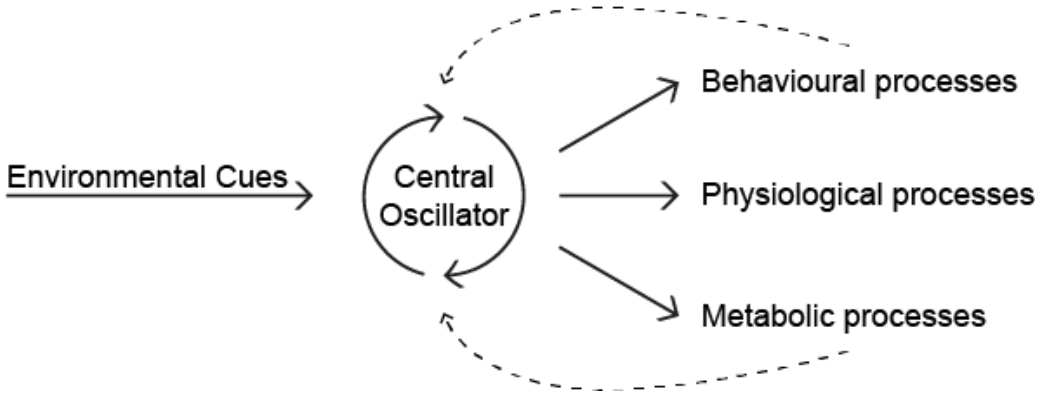


Figure 1.1: Schematic model of the circadian clock input and output signals. Dashed lines indicate feedback loop from the output signals to the central clock.

- entrainable, since circadian clocks can be synchronized by external cues
- self-sustained, i.e. oscillations can persist even in the absence of environmental cues
- temperature compensated, meaning that moderate variations in ambient temperature do not affect the period of circadian oscillation.

Circadian clocks represent an evolutionarily conserved adaptation to the environment that can be traced back to the earliest life forms. In animals they can be analyzed as an integrated system that begins with genes and concludes with behavioural outputs. This system is fundamental to achieve temporal homeostasis with the environment at the molecular level, by organizing physiology and behaviour in a proactive rather than a responsive manner. It is accomplished by regulating gene expression to create a peak of protein expression once every 24 h to control when a particular physiological process is most active throughout the day; among others, some of the functions that are under circadian regulation are food intake and metabolism, sleep-wake cycles, hormone levels, drug and xenobiotic metabolism, cell-cycle progression and body temperature, which all have a direct impact on fitness and survival (Bass and Takahashi, 2010; Mohawk et al., 2012; Partch et al., 2014).

Therefore, the key biological function of circadian clock is to maintain the physiological state of an organism and its alignment with the external environment, by acting as a driver of metabolic physiology, partitioning behavioural and metabolic processes according to time of day within each peripheral tissue, as shown in Figure 1.2.

In Figure 1.2 the different metabolic processes taking place during the wake or sleeping phase are detailed for a number of organs. 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. Thus, rhythmic expression helps to temporally separate incompatible biochemical processes, thereby preventing futile cycles (e.g., the simultaneous biosynthesis and degradation of a given molecule); this results in the fact that rhythmic gene expression is more energy efficient than

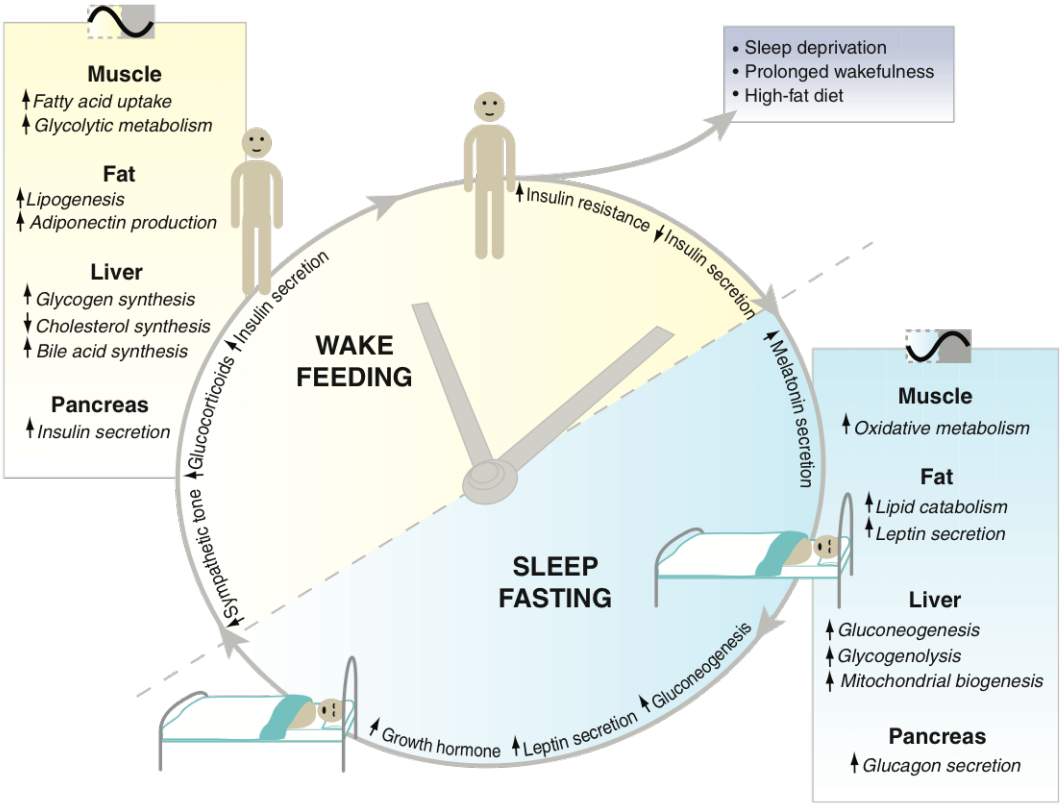


Figure 1.2: Metabolic processes alternation during the day and night cycle in peripheral tissues (Bass and Takahashi, 2010).

maintaining constant level of expression (Panda, 2016).

From Figure 1.2 it is possible to understand that proper functioning and alignment of peripheral clocks with the external environment keeps metabolic processes, which are mostly regulated by the clock, in synchrony with the environment. This is a crucial step for maintaining the health of the organism, since stimuli and phase of the clock are aligned. Circadian clock misalignment can be due to environmental factors (such as shiftwork and sleep deprivation), aging and diet (Bass and Takahashi, 2010), and it has been proven to cause several diseases, such as cardiovascular problems, metabolic syndrome and cancer (Panda, 2016; Sahar and Sassone-Corsi, 2009).

1.1.2 Hierarchical organization of the clock and entrainment

If on the one hand it is important that peripheral clocks are aligned to the external environment, on the other hand they must also be synchronized among each other since each tissue has its own circadian clock. The synchronization of peripheral clocks is ensured by the hierarchical organization of the circadian clock in the body (Figure 1.3).

In mammals, the master or central clock is located in the hypothalamic suprachiasmatic nucleus (SCN), a small brain region containing 10,000–15,000 neurons. The SCN clock can function autonomously, without any external input, thus generating approximately 24 h rhythms in the absence of

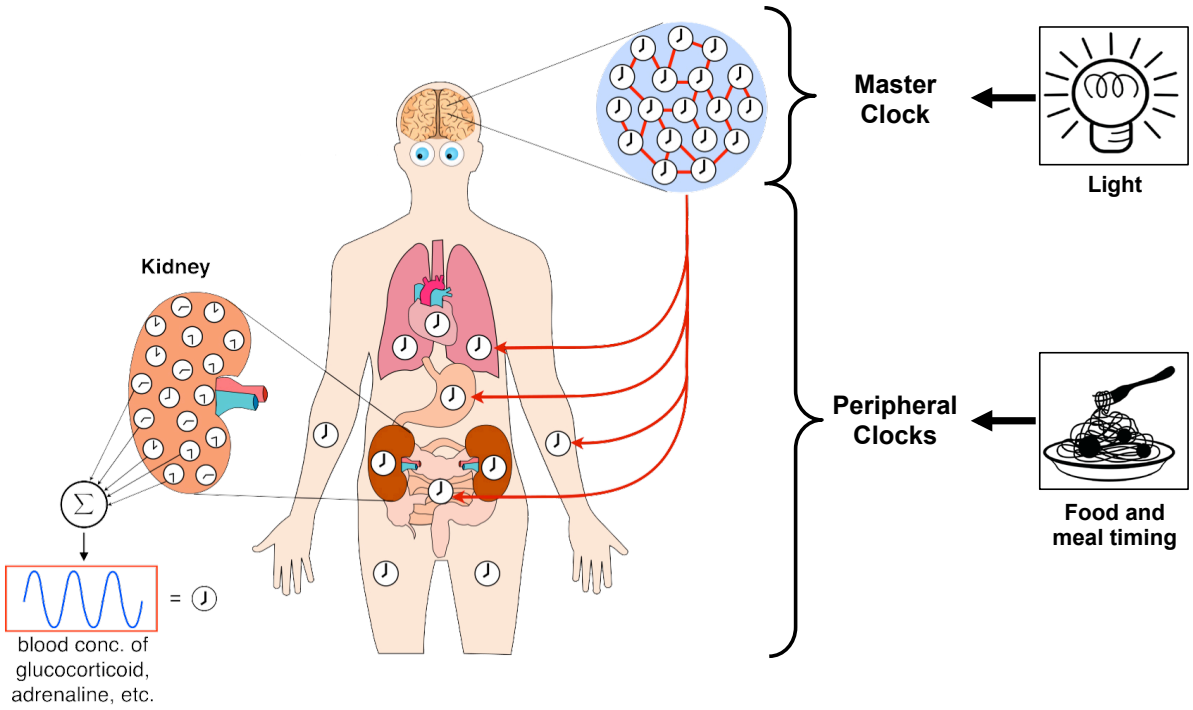


Figure 1.3: The hierarchical organization of the mammalian circadian clock and the entrainment cues. Adapted from Gupta et al. (2016).

environmental cues. Ablation of the mammalian SCN eliminates circadian patterns of behavioural activity, endocrine output, and many biochemical processes throughout the organism. It is also known as the circadian pacemaker (Bell-Pedersen et al., 2005; Sahar and Sassone-Corsi, 2009).

Clocks outside the SCN are referred to as peripheral clocks and are thought to be synchronized by the master clock to ensure temporally coordinated physiology. The synchronization mechanisms implicate various humoral signals, including circulating entraining factors such as hormones, sympathetic and parasympathetic systems, core body temperature and rest-activity rhythm (Lévi et al., 2010; Partch et al., 2014; Sahar and Sassone-Corsi, 2009). Peripheral clocks are present in almost all mammalian tissues, such as liver, heart, lungs, kidneys and skeletal muscles, where they maintain circadian rhythms and modulate transcription factors in a paracrine fashion to regulate tissue-specific gene expression (Bell-Pedersen et al., 2005; Partch et al., 2014; Sahar and Sassone-Corsi, 2009).

Entrainment of circadian clock to the external environment takes place through two main environmental cues: light-input pathway and feeding time. Light and dark cycles directly entrain the master-clock, which receives photic input through the retino-hypothalamic tract, while peripheral clocks are synchronized by the availability of metabolites and feeding time. In order to keep alignment of internal clocks, which is a feature of a healthy organism, it is necessary to have coherent stimuli from both environmental cues, meaning that, for diurnal animals, light stimuli must be matched with food availability, and dark phase must correspond to fasting.

Moreover, in Figure 1.3 another level in the circadian clock architecture is shown, which is the cell-

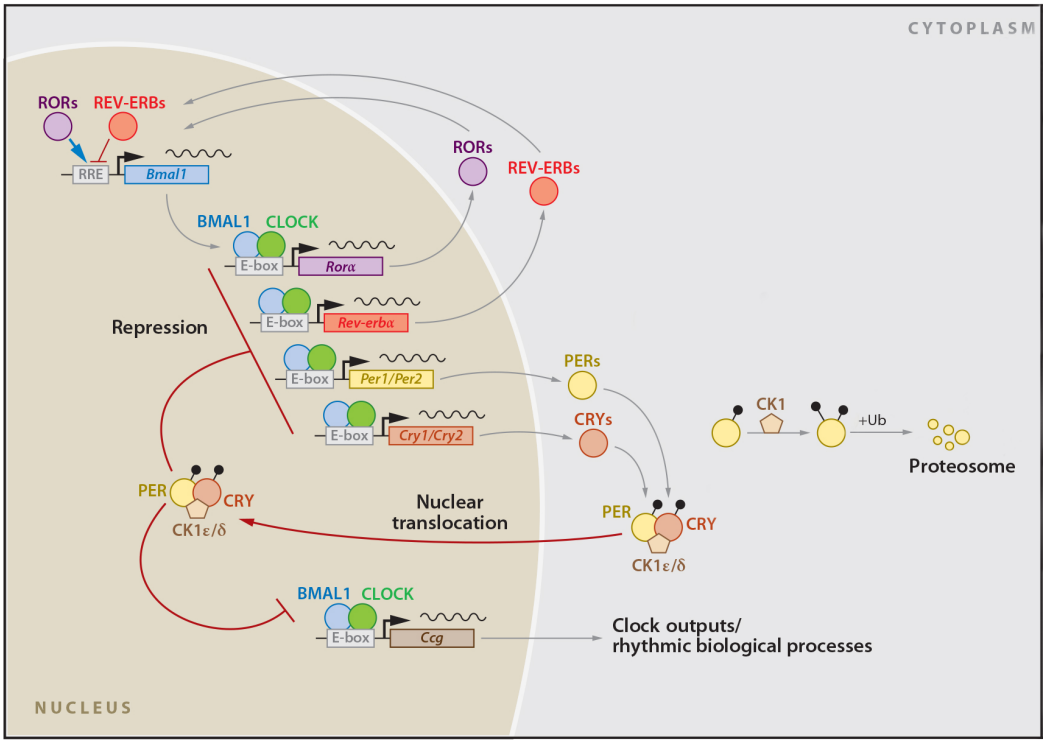


Figure 1.4: The molecular mechanism of the circadian clock in mammals. Adapted from Mohawk et al. (2012).

autonomous clock: each single cell in each tissue has its own molecular clock, and the clock of each tissue corresponds to the average of the clocks of all cells, which are kept synchronized by intercellular coupling mechanisms. Molecular clocks located in each single cell (neurons of the SCN and cells in peripheral tissues) share the same molecular architecture and capacity to generate sustained circadian rhythms, although one key difference between master and peripheral clocks lies in the degree of their intercellular coupling, being coupling among neurons of the SCN very strong compared to intercellular coupling in peripheral tissues (Gupta et al., 2016; Partch et al., 2014).

1.1.3 The molecular circadian clock

Even though each tissue has its own circadian clock, the molecular mechanism governing the cell-autonomous clock is conserved among tissues and mammalian species.

The cell-autonomous molecular clock in mammals is generated by two interlocking transcription-translation feedback loops (TTFL) that function together to produce robust 24 h rhythms of gene expression. The circadian clock molecular network is shown in Figure 1.4.

The core TTFL is driven by four clock proteins: two activators (CLOCK and BMAL1) and two repressors (PER and CRY), as well as by kinases and phosphatases that regulate the phosphorylation and thereby localization and stability of the clock proteins. Heterodimeric complex CLOCK:BMAL1 activates transcription of the repressor Per and Cry genes (binding to cis-acting E boxes in the promoter

region of Per and Cry), as well as other clock-controlled output genes; from Per and Cry genes, PER and CRY proteins are translated and transported from the nucleus to the cytoplasm. In the cytoplasm they heterodimerize and translocate to the nucleus to interact with CLOCK:BMAL1, inhibiting further transcriptional activation. As PER and CRY proteins are degraded through ubiquitin-dependent pathways, repression on CLOCK: BMAL1 is relieved and the cycle begins again with 24 h periodicity.

If, following synthesis, the repressor proteins PER and CRY translocate directly to the nucleus to repress CLOCK andBMAL1, the whole cycle would take just a few hours rather than one day. To maintain the daily oscillations of clock proteins, a significant delay between the activation and repression of transcription is required; the delay between the activation and repression of transcription is ensured by regulation through post-translational modifications. Reversible phosphorylation regulates important processes such as nuclear entry, formation of protein complexes and protein degradation.

The casein kinase CKIδ and CKIε play an important role in determining the intrinsic period of the clock by controlling the rate at which the PER:CRY complexes are either degraded or enter the nucleus, and their activity is either counteracted or regulated by the phosphatases PP1 and PP5, respectively.

A second negative feedback loop is generated through transcriptional activation of BMAL1 by the retinoid-related orphan receptors (RORα, β, and γ) and repression by REV-ERBα/REV-ERBβ, both expressed under the control of CLOCK:BMAL1 transcription complex. TTFL drives rhythmic changes in Bmal1 transcription and introduces a delay in Cry1 mRNA expression that is crucial for proper circadian timing.

The presence of cooperative, interlocking feedback loops provides robustness against noise and environmental perturbations to help maintain accurate circadian timing and also helps to generate phase delays in circadian transcriptional output that optimally time gene expression for local physiology.

1.1.4 Technology to study circadian clock

In this Section conventional approaches for the analysis of circadian gene expression in cell populations will be reviewed. They are based on quantitative real-time Polymerase Chain Reaction (qrtPCR) and on bioluminescence assay.

1.1.4.1 Quantitative real-time PCR

qrtPCR, currently the most sensitive method for RNA quantification, is a molecular biology assay aimed at precise quantification of the levels of a target mRNA sequence.

The assay performs amplification by Polymerase Chain Reaction (PCR) of cDNA, which has been retrotranscribed from the RNA extracted from the cells (cDNA is the complementary DNA sequence of RNA sequence). The amplification process takes place by repeating a specific thermal profile for a certain

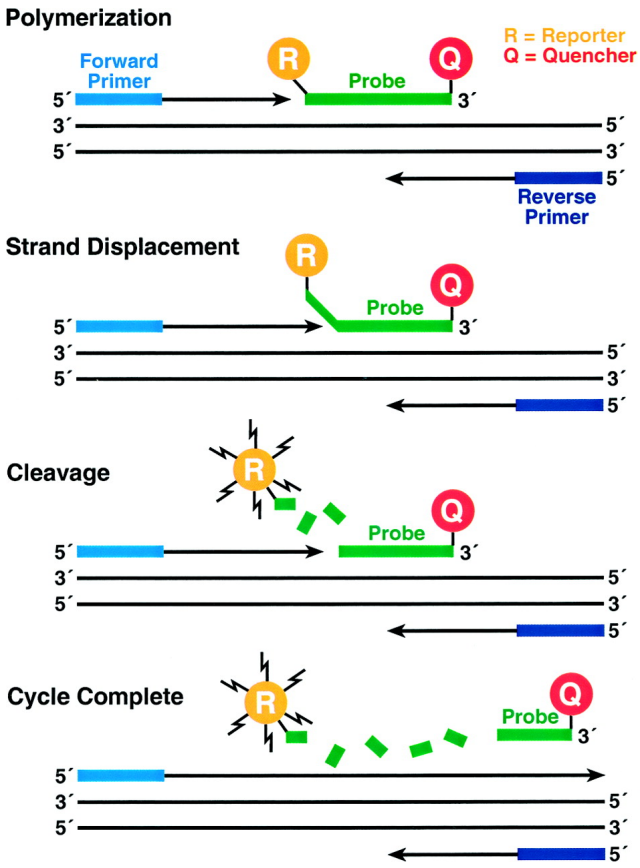


Figure 1.5: Schematic representation of qrtPCR performed with TaqMan probes

number of cycles[1]. The amplification process is measured over time by detection and quantification of a fluorescent probe, whose signal increases proportionally to the amount of sample amplified at each cycle. Thus, by measuring the fluorescence at each cycle, it is possible to monitor the exponential phase of the reaction, during which the increase of PCR product is linearly correlated to the initial amount of the sample.

A schematic representation of qrtPCR assay with TaqMan probes is shown in Figure 1.5. The fluorescence is generated by TaqMan probe, designed to hybridize to an internal specific region of the PCR product. The probe contains a fluorochrome reporter at 5' end and a quencher at 3' end, at a distance that allows the quencher to turn off the fluorescence of the reporter, in accordance to Fluorescence Resonance Energy Transfer (FRET). During the amplification of the target sequence, the 5' end of the TaqMan probe is hydrolyzed by the polymerase, thus the quencher and the reporter are cleaved and the fluorescence can be detected. The signal increases at each cycle in a proportional manner with respect to the probe hydrolysis, and so the amplification. At each amplification cycle, the amount of double stranded amplicons grows exponentially doubling at every cycle, and the fluorescence intensity of the sample grows pro-

[1]The thermal profile is formed by: i) step at 95 °C for 15 s, for cDNA denaturation, in which double stranded DNA molecules are broken apart; ii) step at 60 °C for 1 min, for primer annealing, in which the primer sequence binds to the specific target cDNA sequence, and for elongation, in which the polymerase to synthetize the complementary strand of DNA beginning with the primer sequence. Step i) and step ii) are repeated 40 times. Prior to step i) only before the first cycle, there is an incubation step at 50 °C fro 2 min and an enzyme activation step at 95 °C for 10 min.

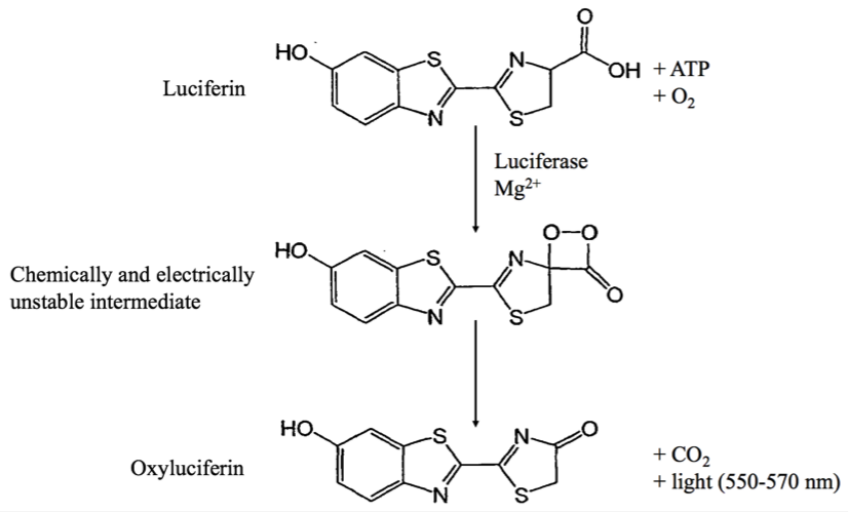


Figure 1.6: Luciferin oxidation into oxyluciferin catalysed by the enzyme Luciferase

portionally until the plateau of the reaction is reached. TaqMan assay ensures high sensitivity and high specificity in the quantification of the target mRNA, minimizing the probability of non-specific mRNA sequence detection and other possible technical artifacts.

1.1.4.2 Bioluminescence assay

The other conventional technique for gene expression quantification is the use of bioluminescence assay, by which it is possible to monitor over time the transcriptional activity of a specific gene of interest.

Bioluminescence genetic reporter systems are obtained by genetic engineering techniques (such as luciferase knock-in mice or stable plasmid transfection), by which, in the regulatory sequence of a specific gene to be studied, a reporter gene, encoding for the enzyme luciferase, is inserted, so that the expression of the reporter gene is regulated in the same way as the gene to be studied. Thus, by measuring the expression of the reporter gene, it is possible to relate it to the expression of the gene of interest.

In particular, the enzyme luciferase, which is encoded by the reporter gene, catalyses the energetically favoured chemical oxidation reaction in Figure 1.6, which shows the conversion of the substrate luciferin into its oxidised form (oxyluciferin); the reaction takes place with light emission, thus by measuring the emission of light it is possible to directly relate it to the amount of luciferase, which expression is equal to the expression of the gene of interest.

Light emission can be measured by highly sensitive detection instruments, such as Charge-Coupled Device (CCD) cameras, which are placed on dark acquisition chambers.

1.1.4.3 Comparison of conventional circadian gene expression techniques

The two conventional techniques described above are currently and widely used for gene quantification in *in vitro* circadian studies. By these techniques, broad and deep understanding of the circadian clock mechanisms has been achieved during the years.

By qrtPCR, it is possible to quantitatively measure gene expression of few genes of interest; however, in order to define the temporal pattern of expression of these genes, discrete sampling over time is required (normally a sample is taken every 4 h for 24 h), and a large number of cells is required at the beginning of the experiments.

On the other hand, luminescence assay allows continuous quantification of gene expression over time, since light emission is quantified over time in a non-invasive way, thus it does not require a large number of cells. However, the expression of only one gene can be assessed at a time.

In conclusion, conventional gene expression approaches allow the description of a restricted subset of genes, which is very important in the understanding of the clock mechanisms; however, they do not consider the temporal pattern of expression of other genes that are under circadian control, and it is known that circadian regulation governs a large number of biochemical processes.

1.1.5 Causes and consequences of clock misalignment

From the previous Sections it can be understood that the circadian clock plays a crucial role in keeping the homeostasis of the body, therefore the endogenous circadian system and 24-h environmental/behavioural cycles should be in phase. Cell-autonomous circadian clocks interact with daily light/dark (LD) and feeding/fasting cycles, thus their phase can be adjusted to be aligned to the environmental stimuli.

Since circadian clock has a complex architecture and it is entrained by two environmental cues, i.e. light/dark cycles and metabolites availability, circadian clock misalignment can take place at different levels: it can be referred to inappropriate timed sleep and wake in relation to the biological night, misalignment of sleep/wake with feeding rhythms, or misaligned central and peripheral clocks. These misalignments can be caused by:

- environmental factors, such as behavioural desynchronization, shiftwork, jet-lag, sleep restriction and social jet-lag
- genetic factors, such as clock gene alterations, uncoupling among SCN oscillators or among SCN and peripheral oscillators
- physiological, specially aging.

Disruption of circadian rhythms through any of these factors has been associated with physical disorders, such as cardiovascular disease, diabetes, obesity, cancer, as well as psychiatric problems as depression,

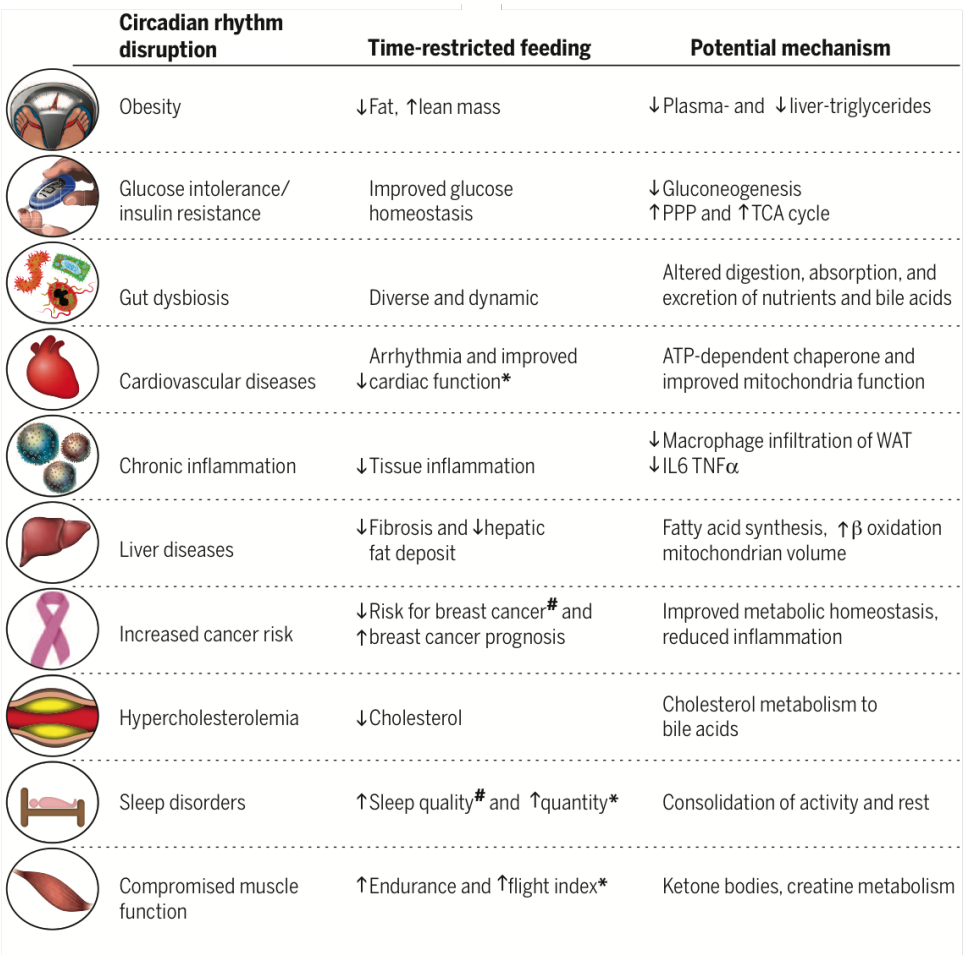


Figure 1.7: Consequences of circadian rhythm disruption by environmental factors, specially by altered feeding schedules, compared to benefical effects of time-restricted feedin. The potential mechanisms are largely based on rodent studies; few observations have been made in insects (*) and in humans (#). IL, interleukin; TNF, tumor necrosis factor. Adapted from Panda (2016).

bipolar, schizophrenia and attention deficit (Baron and Reid, 2014). Circadian mutant murine models have been found to have metabolic disorders, such as hyperlipidaemia, hepatic steatosis, hyperglycaemia and hypoinsulinaemia; in addition, they have been shown to be hypersensitive to insulin shock, suggesting a direct role for the molecular circadian clock in regulating glucose homeostasis (Takahashi et al., 2008).

Circadian misalignment has also been associated with an increase of inflammatory markers (Morris et al., 2016), as well as dysregulation of feeding behaviour, changes in appetite stimulating hormones, glucose metabolism and mood; other studies show that circadian disruption may actually drive pathogenesis of many neurodegenerative conditions, such as Alzheimer's disease, Parkinson's disease, and Huntington's disease, even though until few years ago it was thought that circadian misalignment was only among the earliest symptoms of these diseases (Musiek and Holtzman, 2016).

Therefore, since circadian clock regulates and control a wide number of downstream biological processes, the alignment of the endogenous circadian system and 24-h environmental/behavioural cycles is

of paramount importance to maintain organisms in a healthy status. Among all causes of circadian misalignment, the ones that are arising more attention are the ones related to environmental factors, since they are directly related to modern life style in western countries. More specifically, thanks to recent findings that revealed the strong synchronizing power of food timing, many studies are focusing on behavioural desynchronization related to food showing the negative effects on health of altered feeding schedules, as shown in Figure 1.7. This aspect will be deeply analysed in Section 1.2.

1.2 Circadian clock and metabolism: a feedback loop

In this Section the interconnection between circadian clock and metabolism will be presented, describing the circadian control of metabolism and the reciprocal regulation of metabolites on circadian clocks.

1.2.1 Circadian clock control of metabolism

The circadian clock is intimately connected to metabolism (Asher and Schibler, 2011; Bass, 2012; Eckel-Mahan and Sassone-Corsi, 2013; Green et al., 2008). Direct evidence is provided by the metabolic disorders observed in mouse models with targeted mutations of clock genes (Takahashi et al., 2008). Moreover, it has been discovered from whole genome profiling experiments that a significant fraction of the mouse transcriptome (20-30%) is under circadian regulation and that the majority of the pathways regulated by the clock are embedded in fundamental metabolic pathways (Takahashi et al., 2008). Genes involved in energy metabolism, redox state of the cell and cell signaling also showed circadian variation in their steady-state message levels. At a molecular level, circadian regulators drive rhythmic expression of thousands of target genes by binding cis-regulatory sites or through downstream transcriptional regulators, or interact with a number of coactivators, corepressors, and chromatin-associated factors that read, write, or erase chromatin histone modification marks to activate or repress transcription (Panda, 2016).

In the liver, which plays a central role in nutrient metabolism, basic cellular pathways such as glycolysis, fatty-acid metabolism, cholesterol biosynthesis, and xenobiotic and intermediate metabolism are under circadian regulation. Importantly, rate-limiting steps in these various pathways are key sites of circadian control, highlighting the fundamental role that circadian clocks have in cellular and organismal physiology, and, specially, in metabolic pathways (Takahashi et al., 2008).

Studies performed both in animals and humans, show that not only what and how much food is eaten but also timing of food consumption, independent of total caloric intake and macronutrient quality, determine different physiological responses: for example, postprandial glucose rise is lowest after breakfast and highest after dinner, among healthy adults eating identical meals. This is related to the fact that daily rhythms in the function of numerous genes prime the organism to assimilate nutrients, to mobilize these

nutrients for various functions, and to discard metabolic waste at specific times of the 24-hour day; this has been shown to exist even in absence of light cues (Panda, 2016).

Normal circadian physiology is maintained by consistent daily pattern of eating and fasting, which is the regular schedule that can be anticipated thanks to circadian rhythms. On the contrary, frequent disruptions in daily activity/rest and eating/fasting rhythms (as occurs in shiftwork) or genetic disruption of circadian clock in rodents predisposes to metabolic diseases. Certain diet regimens (e.g., the frequent eating of energy-dense food) and aging can dampen these daily oscillations and predispose to metabolic diseases (Panda, 2016).

1.2.2 Food as entrainment for clocks

Recent circadian and metabolic studies show that not only central nervous system, but also peripheral tissues modulate the response to the environment (Asher and Schibler, 2011; Bass, 2012; Eckel-Mahan and Sassone-Corsi, 2013; Takahashi et al., 2008). While originally thought to be limited to the brain, the occurrence of circadian rhythms has been noted throughout tissues of the body. Parallel to this experimental observation, it has been shown also that as a light pulse during the subjective night can phase advance or delay the SCN circadian clock, food can function as a potent zeitgeber for peripheral tissues, underscoring the important relationship between circadian and metabolic processes (Eckel-Mahan and Sassone-Corsi, 2013).

Therefore, in addition to the circadian control of metabolism (Section 1.2.1), it has become clear that metabolism itself can affect circadian clocks, and, in particular, the clocks of peripheral tissues, since metabolites are direct zeitgebers for peripheral tissues (Asher and Schibler, 2011; Eckel-Mahan and Sassone-Corsi, 2013; Green et al., 2008; Mohawk et al., 2012; Takahashi et al., 2008).

It has been shown that oscillations within some peripheral tissues, such as liver, appear to be dependent on communication with a rhythmic brain but principally on the feeding cycle. Feeding and circadian rhythms in gene expression are so tightly linked that, when food is restricted to a precise time of the day, the expression of a large number of hepatic genes is altered, with a new rhythm that follows that of the feeding cycle. Conversely, in a liver which lacks normal circadian rhythmicity, the majority of hepatic gene expression rhythms can be restored by exposure to a temporally restricted feeding schedule (Eckel-Mahan and Sassone-Corsi, 2013).

If there are experimental evidences that peripheral clocks are entrained by metabolites, the molecular mechanisms responsible for the interactions are still only partially understood. Comprehensive detailed reviews of the most studied pathways through which metabolites interacts with the clock are given by Bass (2012) and Panda (2016), where, for example, glucose metabolic pathway is analysed from the point of view of its interconnection with circadian clock, and the same is done for fasting signals as well

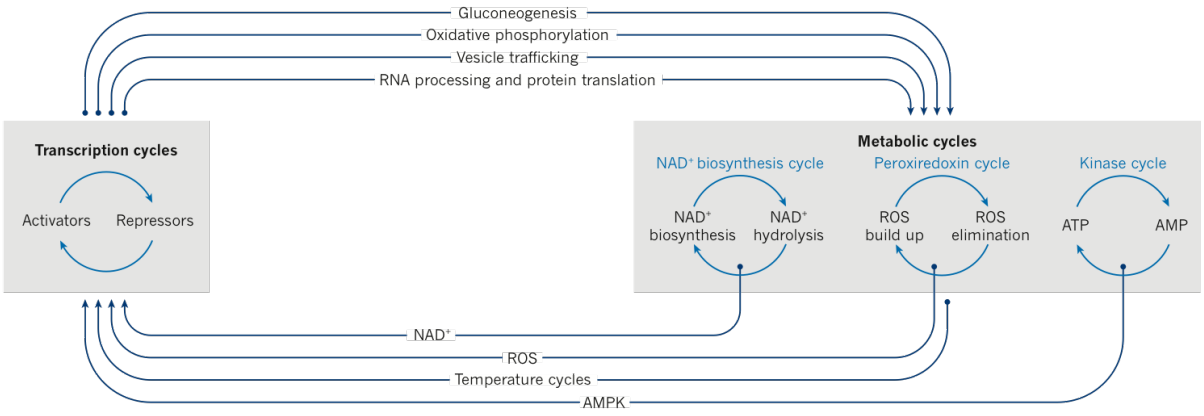


Figure 1.8: Cross-talk between clock transcription and metabolic systems at the molecular and physiological levels (Bass, 2012).

as for other relevant metabolic pathways.

If on the one hand it has been proven that food is the major entrainment cue for peripheral tissues, on the other hand whether peripheral clocks can affect, thus feed-back, the SCN, it is still debated.

It has been shown that the circadian rhythm of the SCN, which responds robustly to light, is largely unaffected by changes in feeding patterns, while circadian oscillators in the periphery and also a food-entrainable oscillator in the brain are strongly reset by restricted feeding. However, it has been observed that a free-running animal that is exposed to a restricted feeding schedule can show signs of SCN entrainment (as measured by locomotor activity, which is regulated by SCN) that may be caused directly or indirectly by food and the anticipatory activity involved in its administration (Eckel-Mahan and Sassone-Corsi, 2013; Panda, 2016).

Other studies showed that although peripheral signals provide feedback to the SCN, these signals are often insufficient to override the robust influence of light. Forcing nocturnal rodents to eat during the day changes the peak phase of expression for nearly all rhythmically expressed genes in the liver without affecting phases of gene expression in the SCN, which remains tied to the light/dark cycle. However, the phase of circadian gene expression in the pituitary, dorsomedial hypothalamus, and para-ventricular nucleus are affected by daytime feeding. Daytime access to a limited amount of food also elicits a survival strategy in nocturnal rodents by suppressing the natural circadian drive for daytime sleep and increasing food-seeking activity before the arrival of the daytime meal (Panda, 2016; Takahashi et al., 2008).

In summary, it is clear that circadian clocks and metabolism are intimately linked and reflect adaptive responses of living systems to the cyclic environment, being both part of an interconnected regulatory loop. Food serves not only as an output of the clock but also as a clock input mechanism, particularly for peripheral tissues, since physiological demands occur on a daily basis and metabolic processes vary with the physiological demands of the organism (Figure 1.8). Because peripheral tissues communicate back

to the brain via ghrelin, leptin, glucose, insulin, etc., circadian feeding contributes to an intertwining of the clock and metabolism that appears to be crucial for metabolic homeostasis. Disruption of the optimal tuning of rhythmic metabolic pathways could then have suboptimal consequences and important implications for metabolic disorders.

1.3 Motivation and aim

The circadian system orchestrates the temporal organization of behaviour, physiology and metabolism in synchrony with the 24-hour rotation of the Earth. The circadian system is a complex feedback network that involves interactions between the central pacemaker and peripheral tissues. Its biological function is to maintain the physiological state of an organism allowing to anticipate environmental changes and its alignment with the external environment; in fact the clock is self-sustained and it can also be entrained by environmental cues.

Until 10 years ago, it was thought that light and dark cycles, which are the entrainment cues of the masterclock, were the main synchronizing agent of the circadian clock since it was thought that SCN could drive the clock at a systemic level. However, recent findings noted the presence of circadian rhythms throughout almost all tissues of the body and more striking experimental evidences showed that food can function as a potent zeitgeber for peripheral tissues besides the phase of the SCN, suggesting that circadian regulation is intimately linked to metabolic homeostasis and that dysregulation of circadian rhythms can contribute to diseases.

Thus, a strong interconnected feedback loop exists between circadian clock and metabolism, being metabolic processes controlled in a circadian manner as well as an entrainment cue for circadian peripheral clocks. However, the mechanisms through which this tight regulation and control take place are still not deeply understood. Therefore, understanding the mechanisms by which food affect the daily sleep-wake cycle and metabolism could potentially reveal lifestyle and therapeutic interventions for preventing and treating metabolic diseases.

Currently, circadian clocks studies are performed *in vivo* and *in vitro*. In the former, animal models are exposed to physiological circadian stimuli, which are periodic and cyclic light/dark stimuli combined with different feeding schedules, and it has been shown that peripheral circadian clocks can be entrained by cyclic oscillations of endocrine factors and metabolic cues associated with dietary regimens of food intake and timing; however, in these models it is not possible to isolate the behaviour of peripheral clocks independently from the stimuli of the masterclock, i.e. systemic cues. In the latter, cell cultures, representing one peripheral clock independently from the whole system, are usually stimulated with a single-pulse stimulation, either of one metabolite or an endocrine factor. However, single-pulse stimuli do not mimic the oscillatory periodic intrinsic nature of environmental stimuli at which tissues

are exposed *in vivo*. Moreover, *in vitro* studies evaluate the effects of a single metabolite or a single endocrine factor, while very often their variations are tightly interconnected in biological organisms, as, for example, glucose and insulin.

Therefore, this research work is aimed at studying *in vitro* peripheral circadian clock entrainment through metabolic frequency-encoded stimulations, which mimic physiological oscillations at which peripheral tissues are exposed *in vivo*. In order to achieve this goal, specific technological tools are developed to perturb periodically an *in vitro* model and to continuously monitor its biological response. More specifically, microtechnologies will be developed since they ensure a high spatial control of the cell culture microenvironment, as well as a high precision in timing of the stimuli, considering that microtechnologies offer the possibility to be coupled with automation. This model will mimic more closely the pulsatile nature of metabolites and hormones and it will allow to simulate both physiological and pathological conditions.

Conventional techniques for molecular circadian clock studies are based on quantitative gene expression assays or on bioluminescence imaging. Even though the first technique allows to quantify more than one single gene at a time (which is what can be achieved by bioluminescence imaging), the number of the variables that can be quantified by these techniques is very limited. On the one hand, very precise information about the considered genes are obtained with these techniques, on the other hand they do not allow to obtain information regarding other unknown biological processes that may be involved in that phenomena, preventing the possibility of discovering new correlations between biological pathways. This is particularly important in the case of circadian studies, since a very large number of biological processes are under circadian control and researchers continuously identify new clock-controlled genes.

A more comprehensive approach for deeply understanding biological responses has started to be feasible thanks to recent technological development and cost reduction in RNA sequencing. By this approach, it is possible to identify genes that have a circadian pattern from genes that are constantly expressed as well as to group the genes under circadian regulations based on the phase of their expression, since this might indicate that the genes belong to the same regulatory network. Several algorithms and software are available to assess if a gene has a circadian expression (Fisher's G-test, Jonckheere-Terpstra-Kendall algorithm, COSOPT, ARSER, etc.); by the use of these algorithms it is possible to assess if a gene is significantly expressed in a circadian way, as well as to estimate the characteristics of the oscillation, i.e. amplitude, phase and period. However the algorithms do not allow to globally visualizing transcriptome and showing which genes have the same phase at whole gene expression level.

Therefore, during this research work data analysis tools have been developed to analyse whole transcriptome data, distinguishing between genes that are expressed in a circadian manner from the genes that are constantly expressed, to group genes based on the phase of their expression, to visualize globally

transcriptomic data and clearly identifying modifications at the transcriptome level from one biological condition to another one, or between organs or species. This tool will help in suggesting possible new gene interaction networks and to compare different biological conditions.

1.4 Dissertation structure

After giving the theoretical background regarding circadian clock and its tight interconnection with metabolism, and defining the motivation and the aim of the research work, in this Dissertation the study of circadian clock through frequency-encoded entrainment stimulations is proposed as follows:

- **Chapter 2** presents the findings of the innovative approach of frequency-encoded metabolic stimulations, which are used to perturb an *in vitro* model mimicking the daily oscillations at which tissues are exposed *in vivo*;
- **Chapter 3** presents the developed technology to automate frequency-encoded perturbations, in order to develop an *in vitro* model which can resemble physiological and pathological conditions at which peripheral tissues are exposed *in vivo*, achieving precise spatial as well as temporal control of cell microenvironment; the use of this technology represents a novelty in the field of mammalian circadian clock studies;
- **Chapter 4** presents the developed data analysis method to analyse whole gene expression of the cells, in order to identify and characterize circadian and non-circadian transcripts; this method allows to achieve a more comprehensive approach compared to conventional techniques and data analysis methods;
- **Chapter 5** presents the main achievements of this Ph.D. research work, along with the discussion of future investigations that may be carried out.

Chapter 2

Periodic oscillatory metabolic perturbations

The Chapter describes the state of the art of metabolic entrainment of peripheral clocks, it defines the motivation and the aim of performing this experimental investigation. After explaining the rationale of metabolic stimulations, it presents the obtained experimental results.

2.1 State of the art of metabolic entrainment of peripheral clocks

In Section 1.2 the tight interconnection between circadian clock and metabolism has been presented, describing how circadian clock controls metabolism and how metabolic signals feed-back to the circadian clock. This complex feedback mechanism is meant to keep metabolic homeostasis in the organisms and to synchronize peripheral tissues, and probably the whole clock, to the external environment.

However, the mechanisms underlying this regulation are poorly understood, thus a deeper understanding of the molecular pathways by which metabolic inputs affect the circadian clock is critically important for the prevention and treatments of diseases associated with circadian misalignment. This aspect is continuously becoming more important since it has been discovered that feeding/fasting cycles are one of the major entrainment cues in terms of synchronization of peripheral clocks, besides the phase of the SCN (Damiola et al., 2000; Hamaguchi et al., 2015; Vollmers et al., 2009). Specifically, it was shown that rhythmic feeding and caloric content are both necessary and sufficient to synchronize the circadian expression of peripheral clocks (Hara et al., 2001; Kohsaka et al., 2007; Kornmann et al., 2007). Accordingly, there is increasing evidence that peripheral circadian clocks can be entrained by cyclic oscillations of endocrine factors and metabolic cues associated with feeding and dietary regimens or food intake timing (Balsalobre, 2000).

All these studies are performed *in vivo*, where on the one hand it is possible to have information re-

garding the entire organism exposed to physiological or pathological periodic stimuli, while on the other hand it is not possible to dissect the contributions of the different entrainment stimuli (light/dark cycles, feeding/fasting schedules, temperature, ...) as well as to isolate the behaviour of the cell-autonomous clock of peripheral tissues independently from the entrainment stimuli received from SCN through hormones or metabolites.

Both these two limitations of *in vivo* approach could be overcome by *in vitro* experiments using cultured cells, since specific entrainment mechanisms can be isolated and analysed and non-systemic entrainment of peripheral clocks can be studied (Nagoshi et al., 2004; Welsh et al., 2004). For instance, cultured cell lines were used to investigate how endogenous factors, including growth factors, calcium, glucose, angiotensin II, retinoic acid, and nitric oxide, act on cell-autonomous circadian gene expression (Balsalobre, 2000; McNamara et al., 2001; Nonaka et al., 2001).

However, only few studies reported circadian clock entrainment by metabolic cues:

1. Hirota et al. (2002) showed that glucose down-regulates *Per1* and *Per2* transcription without causing phase resetting;

2. Yamajuku et al. (2012) reported the action of insulin as a clock synchronizer with phase shift of maximum 6 h in cultured hepatic cells;

3. Chaves et al. (2014) showed that glucose and insulin may lead to alteration of clock gene expression in the liver mediated by FOXO3 via signaling of PI3K axis;

4. Sato et al. (2014) reported the phase-dependent alteration of peripheral circadian clock with insulin as well as that insulin-induced phase shift was dependent on tissue type and involved PI3K- and MAPK-mediated signaling pathways;

5. Dang et al. (2016) demonstrated that insulin regulates BMAL1 transcriptional activity by reducing its nuclear accumulation via Akt-mediated Ser42 phosphorylation in the liver under physiological conditions.

All these works, in which large phase shifts caused either by glucose or insulin are not observed, are characterized by a single pulse metabolic stimulation, although metabolic entrainment stimuli, to which peripheral tissues are exposed *in vivo*, are periodic and oscillatory. Therefore, conventional *in vitro* models do not resemble physiological metabolic oscillations, which are a key feature of circadian entrainment stimuli.

Moreover, conventional *in vitro* model (Petri dish cell culture) offers only limited capability in investigating the oscillatory nature of environmental stimulations. Among different technological solutions, microfluidics is the most suitable technology for performing dynamic stimulation experiments under strictly controlled microenvironment conditions, including periodic cyclic stimulations (Luni et al.,

2014; Zambon et al., 2015). The microfluidic approach has been reported in a few works for studying the molecular mechanisms of circadian gene oscillations in algae and fungi, allowing long-term measurements with high spatio-temporal resolution (Lee et al., 2013; Teng et al., 2013; Danino et al., 2010), thus overcoming the major technical limitations of conventional systems. No experimental studies of the mammalian circadian clock using microfluidic systems has yet been reported.

2.2 Motivation and aim

Metabolic stimuli are a potent synchronizer for peripheral clocks; however, conventional *in vitro* approach, the one allowing for the investigation of metabolic entrainment on peripheral clocks, is based on single pulse metabolic stimulations, whereas metabolic stimulations at which peripheral tissues are normally exposed are periodic and oscillatory. Thus, the relationship between the feeding/fasting sequences and the cell-autonomous circadian clock should be more fully investigated under periodic stimulations.

The specific aim is to understand how the oscillatory nature of metabolic cycles, resembling feeding/fasting alternations, entrain the cell-autonomous circadian clock in peripheral tissues, and, eventually, if it is possible to reset the phase of peripheral clocks only by frequency-encoded stimulations. Moreover, the experimental investigation is aimed at dissecting the individual contribution of feeding and fasting in circadian phase reset.

To achieve this goal, microfluidic technology has been developed for performing accurate cyclic dynamic stimulations *in vitro* while continuously monitoring circadian oscillations. Cell culture volume has been reduced to microliters in order to achieve an accurate and homogenous spatial control of the metabolic perturbation.

2.3 Rationale of metabolic stimulations

The published works regarding circadian clock entrainment by metabolic cues presented in Section 2.1, investigated how glucose and insulin induce entrainment separately; however, glucose and insulin have a synergistic role, as insulin acts in response to an increase of glucose concentration in the blood for maintaining energy homeostasis. Thus, since the goal is to develop an *in vitro* model of a peripheral clock as close as possible to a physiological one, it is assumed that simultaneous variation of glucose and insulin mimic more closely physiological conditions.

Metabolic oscillations, which are formed by feeding and fasting phases, are resembled using two different culture media:

- high Glucose (25 mM) and high Insulin (100 nM) medium, namely H, for the feeding phase;

- low Glucose (2 mM) and low Insulin (10 nM) medium, namely L, for the fasting phase.

After stimulating the *in vitro* model by cyclic metabolic perturbations, the effect of the imposed entrainment is observed in the free-running period (FR), which correspond to the state of a not-perturbed system.

Cyclic metabolic perturbations have been applied to a murine fibroblast model (primary fibroblasts derived from PER2::LUC knock-in mouse from Takahashi's Lab), which has frequently been used to study the effect of glucose and insulin in peripheral clock resetting mechanisms.

Since the time, i.e. the phase of the oscillation, at which the stimuli are given determines a different response, all stimulations protocol have been started at the trough of *Per2* expression, which was shown to be the most sensitive time of the phase for clock resetting in hepatic cells (Yamajuku et al., 2012). In unit of Circadian Time (CT), the trough of *Per2* expression corresponds to CT6, in the fasting phase in a murine peripheral oscillator model[1].

2.4 *In vitro* metabolic stimulations

In this Section some of the results obtained and discussed in Gagliano (2015) are reported, since they are the basis for the subsequent research work.

2.4.1 Cyclic metabolic perturbations affect the circadian clock

First, it has been experimentally assessed if *in vitro* cyclic perturbations of glucose and insulin together could affect circadian clock even in fibroblasts, by stimulating fibroblasts cultured on Petri dishes every 12 h with the following media combinations: HH12:12, LL12:12, HL12:12 and LH12:12. PER2::LUC fibroblasts have been stimulated for 3 days, after a pre-treatment in L condition for 3 days (Figure 2.1a).

Figure 2.1b shows that *Per2* temporal pattern is significantly different across the four metabolic perturbations. The period of HH stimulation is significantly longer than all other conditions (Figure 2.1c) whereas the circular representation in unit of circadian time (Figure 2.1d) at day 4.5-5.5 of FR-L provides evidence of the peak of expression of *Per2* for all perturbations. The largest difference has been detected between HL and LH, the two anti-phase perturbations, with a phase shift of about 5 h maintained from day 5 to day 7 (Figure 2.1e).

In addition, the influence of the metabolic state during FR or of a pre-treatment (either in L or H medium) of 3 days before starting the cyclic metabolic perturbations has been considered; Figure 2.1f

[1] Under standard light/dark cycles, the time of lights on usually defines circadian time zero (CT0) for diurnal organisms, and the time of lights off defines circadian time twelve (CT12) for nocturnal animals. The unit of measure is the circadian hour (h). Mouse are nocturnal animals, thus their fasting phase ranges from CT0 to CT12 when they are exposed to light, while their feeding phase ranges from CT12 to CT0, while they are kept in darkness.

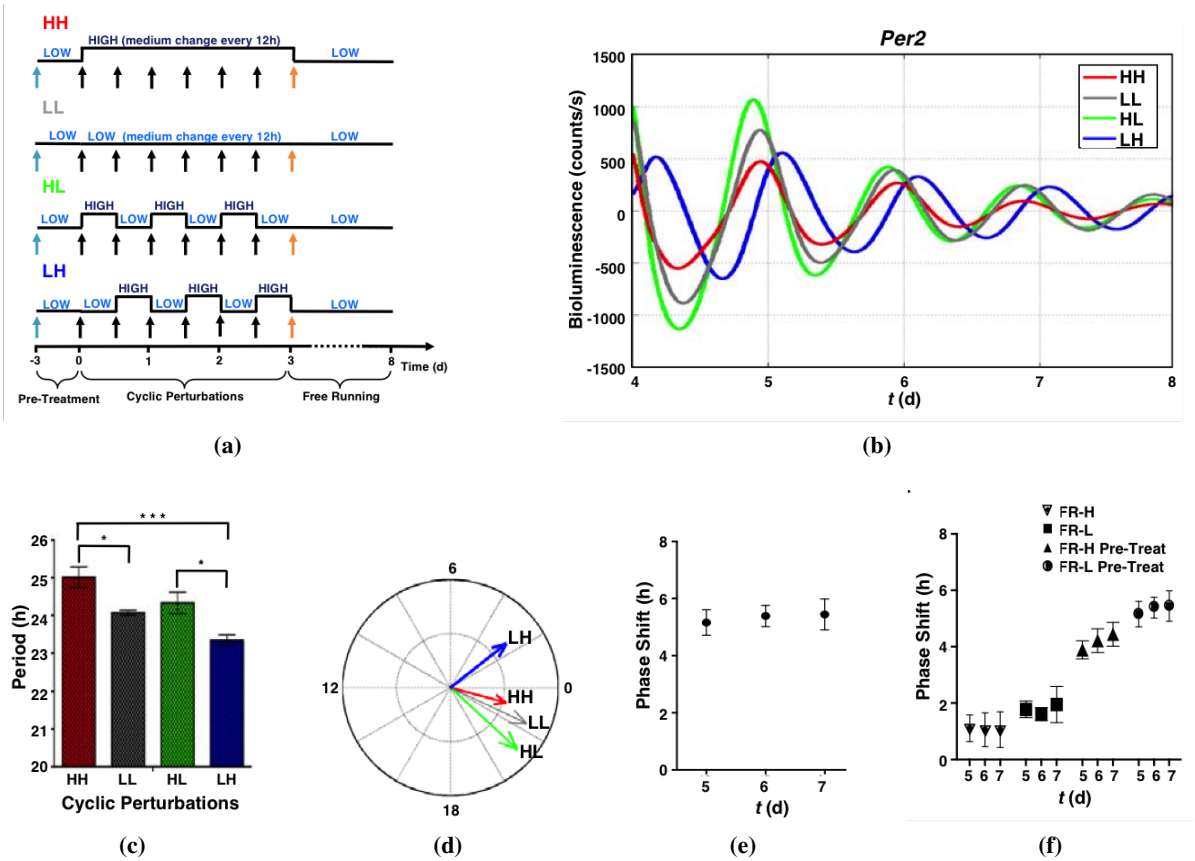


Figure 2.1: Metabolic behaviour of day and night in conventional in vitro model.
(a) Schematic representation of the 4 cyclic metabolic perturbations protocols, by alternating cycles of high and low glucose and insulin levels (H= 25 mM - 100 nM and L= 2 mM - 10 nM, respectively). The four 24 h cyclic perturbations implemented before the Free-Running period (FR), in L, are: HH (12 h of H and 12 h of H), LL (12 h of L and 12 h of L), HL (12 h of H and 12h of L) and LH (12 h of L and 12h of H); all perturbation cycles present a starvation pre-treatment, in L, for 3 days; the recording is performed in L (FR-L) **(b)** PER2::LUC bioluminescence patterns acquired after 3 days of stimulations. Data are represented as mean of bioluminescence signal, N=3 for each conditions. **(c)** Comparison of the circadian period for different metabolic perturbations during FR. Data are represented as mean ± s.d., N=3 for each conditions. *P<0.05, **P<0.01 one-way ANOVA with Tukey's multiple comparisons test. **(d)** Phase in the 4 different conditions; the arrow direction represents the phase (expressed in hours between the day 4.5-5.5 during the FR-L), its length the amplitude, and line thickness the standard deviation. **(e)** Comparison of the phase shift between two anti-phase conditions, HL and LH at the day 5, 6 and 7 of FR-L. Data are represented as mean ± s.d., N=3 for each condition. **(f)** Comparison of the phase shift between LH-HL at different conditions of FR (H and L) and pre-treatment, at the days 5, 6 and 7. Data are represented as mean ± s.d., N=3 for each conditions.

shows that a greater phase shift is observed with pre-treatment with fasting-like conditions (L) highlighting the importance of the metabolic state of the cell culture prior to metabolic perturbations.

In this investigation, it has been observed that 12 h cyclic metabolic perturbations alternating between H and L showed appreciable phase shift after few days; HL anticipates *Per2* expression while LH delays.

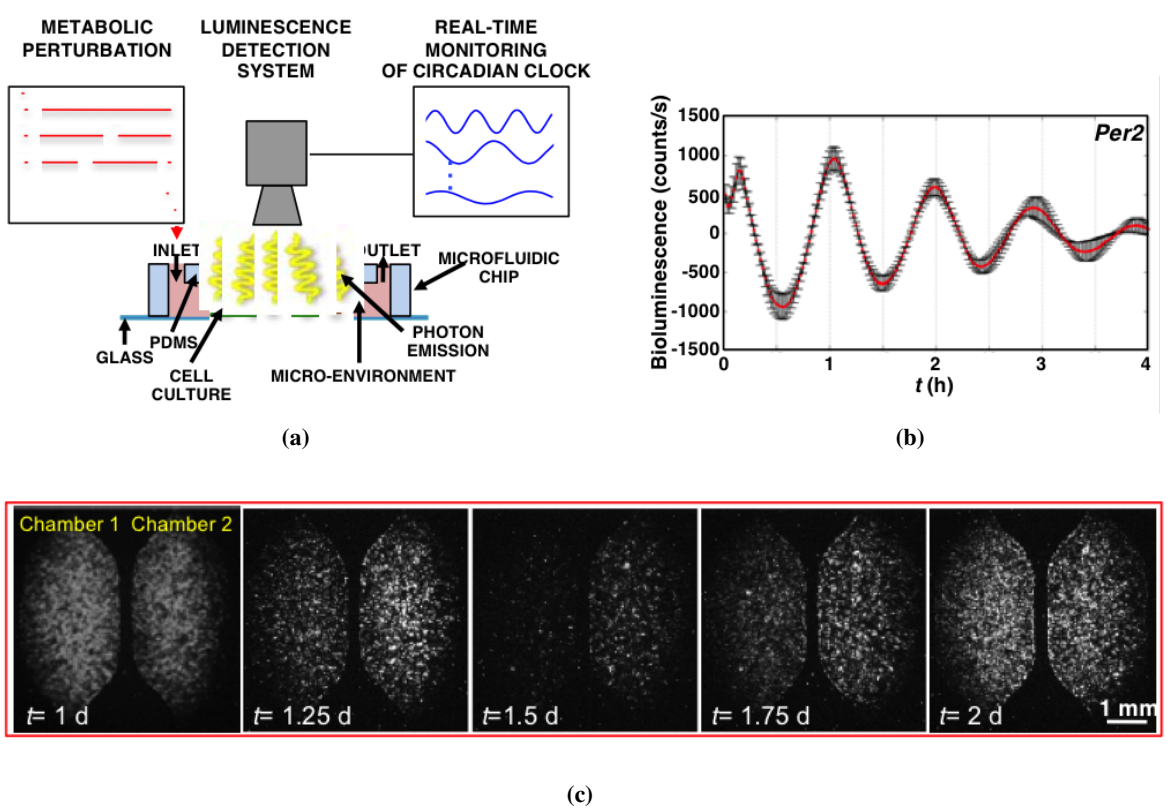


Figure 2.2: Luminescence recording for real-time measurement of clock gene dynamics.
(a) Schematic representation of experimental set-up for acquisition of bioluminescence signal from cell integrated microfluidic chip; microfluidic set-up is coupled with microscope. **(b)** PER2::LUC bioluminescence patterns acquired for 4 days from microfluidic chamber after dexamethasone shock. Data are represented as mean ± s.d., N=9 for each group. **(c)** Time lapse of bioluminescence images, captured every 6 hours, of mouse PER2::LUC fibroblasts, showing circadian rhythms of luminescence for 2 days.

2.4.2 Microfluidic technology for circadian studies

If on the one hand it has been shown that cyclic metabolic perturbations can entrain fibroblast, on the other hand the experimental result also clearly highlights the limit of using conventional Petri dish cell cultures to perform dynamic oscillatory stimulation while recording biological signals, since it was not possible to perform perturbation and recording at the same time. For these reasons, specific microtechnologies have been designed to ensure accurate control of cell culture conditions, while recording circadian gene expression by luminescence detectors (Figure 2.2a).

Since the use of microfluidic technology in the circadian field has been limited to algae and fungi and specially to single pulse stimulations, first it has been investigated if mammalian cell culture exhibits cell-autonomous circadian behaviour in microfluidic. This question is essential to further investigate circadian behaviour while perturbing the system with cyclic metabolic stimulations.

First, it has been experimentally shown that circadian synchronization of mammalian cell culture could be achieved in a microfluidic environment, i.e., in confined volume (microliter range). As proof of concept, Figure 2.2c shows changes of luminescence signal from PER2::LUC fibroblasts in a 1.5 µL

microfluidic chamber at single-cell resolution captured every 6 h. The microfluidic experimental set-up permits recording of highly reproducible and robust oscillatory behaviour after dexamethasone shock (Figure 2.2c).

Thus, this experimental result shows that the luminescence acquisition within a microfluidic platform allows the acquisition of highly time-resolved data (Yamazaki and Takahashi, 2005).

2.5 Frequency-encoded metabolic perturbations

In this Section the results obtained applying frequency-encoded metabolic perturbations are presented: first, the frequency that determines sustained oscillations is determined, then peripheral circadian clock resetting by metabolic perturbations is shown and finally the individual role of feeding and fasting is investigated.

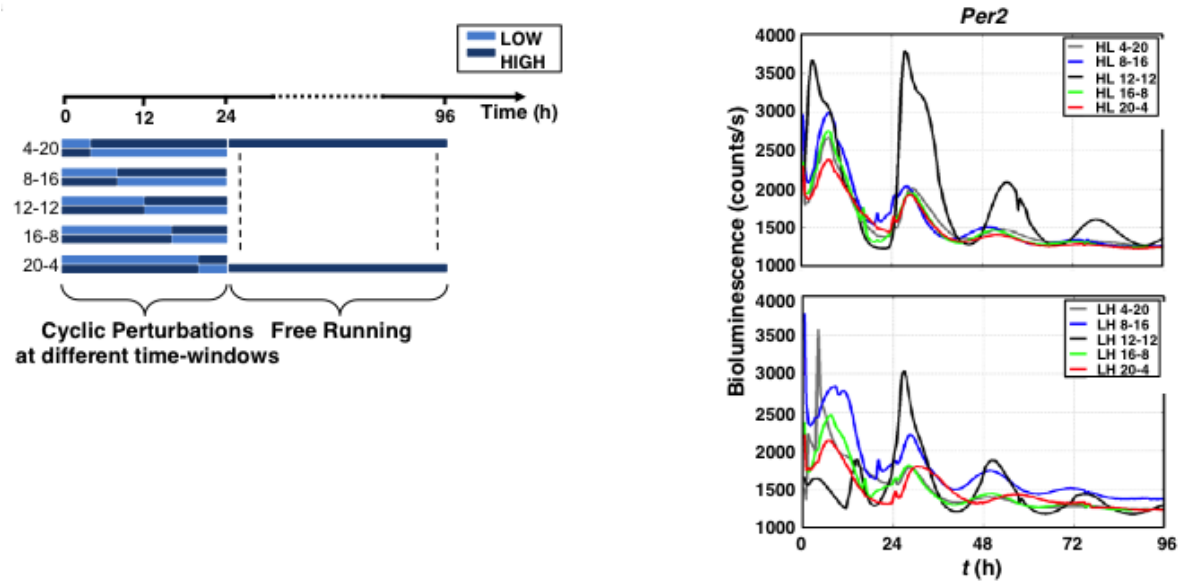
2.5.1 Sustained oscillations with frequency-encoded stimulations

Ater assessing that cyclic metabolic perturbations affect the circadian clock of cultured fibroblasts, and that mammalian cell culture exhibits cell-autonomous circadian behaviour, the effect of different frequencies in cyclic metabolic perturbations has been evaluated.

A series of experiments has been designed in which H and L media have been switched within a 24 h cycle, modulating the period of H and L in HL and LH as follows: 4:20, 8:16, 12:12, 16:8 and 20:4, as shown in Figure 2.3a. Figure 2.3b shows PER2::LUC expression during cyclic metabolic perturbation and free-running period for HL and LH, respectively, recorded from 5.4 μL volume microfluidic chamber. Interestingly, HL12:12 seems to provide the most sustained and robust circadian oscillations compared to all other conditions (both HL and LH). In general, LH shows unsustained oscillatory behaviour for all frequencies despite LH12:12.

On the one hand, the fact that 12:12 frequency is able to maintain sustained oscillations (both for HL and LH) is probably due to the match with the frequency of the intrinsic biological processes controlling well-defined diurnal and night actions (Qian and Scheer, 2016). On the other hand, the sustained oscillations observed for all HL stimulations, opposed to the unsustained oscillations of LH, is likely to be related to the match of the period of the perturbation frequency with the intrinsic cell-autonomous clock. For all conditions, it is important to notice that the first stimulation has been given at the trough of *Per2*, which is aligned with HL stimulation and is in 12 h anti-phase with LH.

According to these results, 12 h frequency media changes has been used for all subsequent HL or LH cyclic experiments.



(a) (b)

Figure 2.3: Circadian behaviour under different time-dependent cyclic perturbations.
(a) Schematic representation of daily perturbations under HL cycles imposed at different time-windows (4:20, 8:16, 12:12, 16:8 and 20:4), followed by a period of FR-H. **(b)** PER2::LUC bioluminescence rhythms of HL and LH imposed at different time-windows; bioluminescence was acquired from t=0 h to t=96 h at intervals of 30 min; data are represented as mean of bioluminescence signal, N=3 for each group.

2.5.2 Peripheral circadian clock resetting

Next, the number of complete cyclic metabolic perturbations required to reset the cell-autonomous clock has been evaluated.

An increasing number of cyclic perturbations for both HL and LH stimulations has been used to stimulate the peripheral clock model (Figure 2.4a). Figure 2.4c shows PER2::LUC luminescence patterns of fibroblasts in HL and LH conditions for 1, 2 and 3 cycles, up to 6 days. HL cyclic stimulations (which are matching with *Per2* phase) develop sustained and robust oscillations with remarkable consistency between all conditions during FR-H (Figure 2.4c, HL-onset). On the other hand, LH cyclic stimulations (which are in anti-phase with *Per2* expression) exhibits disrupted oscillatory behaviour during stimulation and higher variability in the FR-H observation (Figure 2.4c, LH-onset). In order to visualize the relative phase shift between HL and LH for different stimulation cycles, the circular diagram in Figure 2.4c has been used. It clearly shows that the phase difference between of HL and LH, at day 2 of FR-H, increases as long as the number of perturbation cycles increases, with a maximum of 8 h (Figure 2.4d). This result confirms the tight metabolic control by feeding/fasting cycle on the circadian clock.

Finally, it has been verified if 12 h metabolic oscillatory stimulations are able to reset the cell-autonomous clock, realigning it to environmental clock mimicked by feeding/fasting cycles; PER2::LUC fibroblasts at minimum of *Per2* expression have been stimulated with HL, LH, and HH cycles for 3 days.

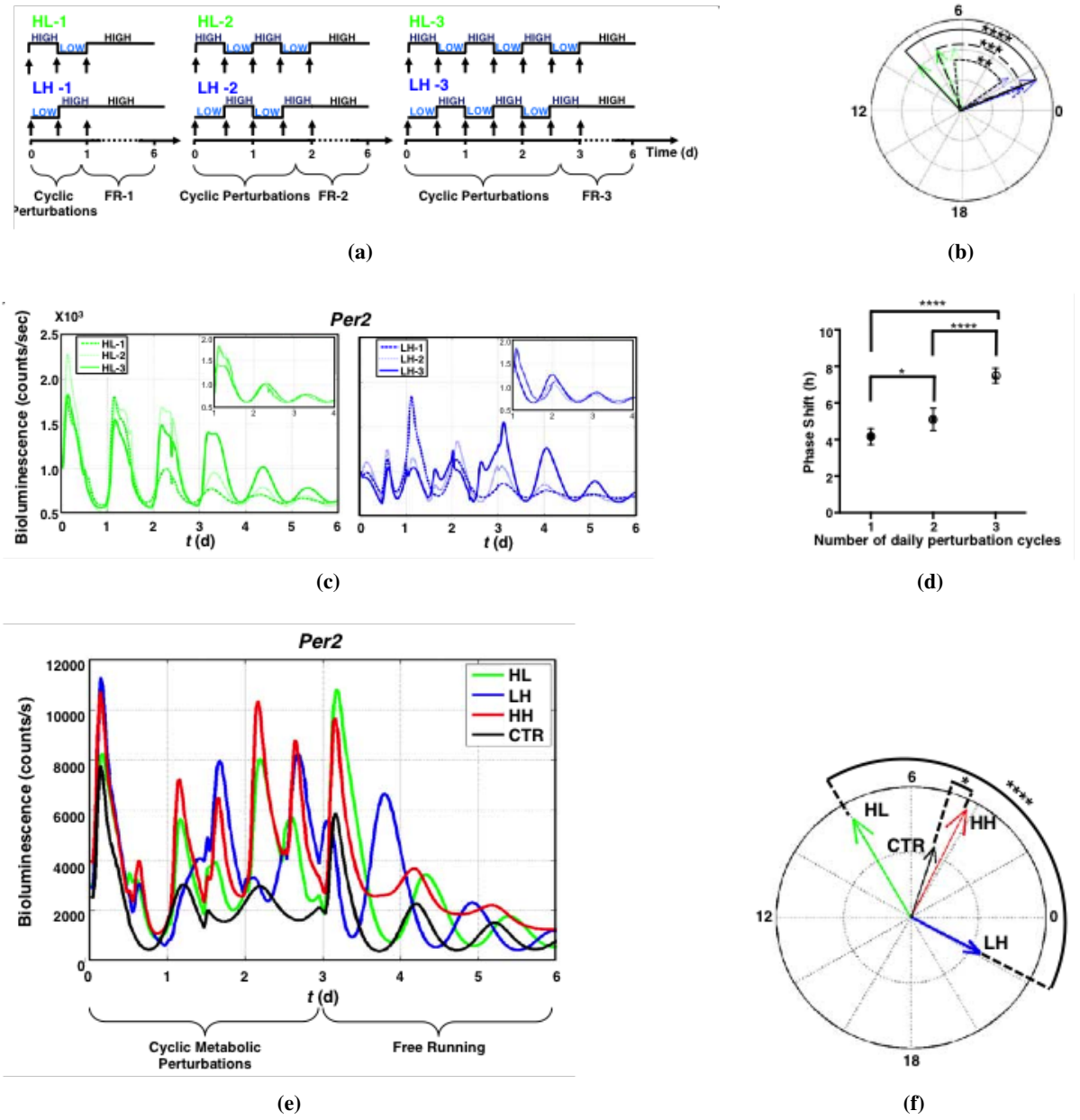


Figure 2.4: Metabolic behaviour of cyclic day and night by microfluidic technology.
(a) Schematic representation of feeding/fasting (HL) and fasting/feeding (LH) behaviour imposed for 1, 2 or 3 days and followed by a FR-H condition. **(b)** PER2::LUC bioluminescence patterns of PER2::LUC mouse fibroblasts in the two anti-phase conditions, HL (1-2-3) and LH (1-2-3), for 1, 2 or 3 days, acquired at interval of 30 min. The detail shows the data of PER2::LUC bioluminescence patterns for day 1 to day 4 of FR-H of each condition. Data are represented as mean, N=3 for each group. **(c)** Phase shift between the two anti-phase conditions, HL and LH; the arrow direction represents the phase (expressed in hours and calculated at day 2 of FR-H for each condition), its length the amplitude, and line thickness the standard deviation; phase shift between the anti-phase conditions, LH-HL (1-2-3), at day 2 of FR-H, N=6 for each group. **P<0.01, ***P<0.001, ****P<0.0001 Student's t-test. **(d)** Comparison of the phase shift between HL and LH conditions imposed for 1, 2 or 3 days, at day 2 of FR-H. Data are represented as mean ± s.d., N=6 for each group. *P<0.05, ****P<0.0001 one-way ANOVA with Tukey's multiple comparisons test. **(e)** PER2::LUC bioluminescence patterns acquired during the four cyclic metabolic stimulations (HL-LH-HH-CTR) at interval of 30 min; data are represented as mean, N=9 for each group. **(f)** Comparison of the phase among all conditions. The arrow direction represents the phase (expressed in hours and calculated at day 2 of FR-H), its length the amplitude, and line thickness the standard deviation. N=9 for each group. *P<0.05, ****P<0.0001 Student's t-test.

Figure 2.4e shows that a disrupted oscillatory behaviour is observed during the 3 days stimulation period with multiple peaks corresponding to 12 h media changes. On the other hand, the FR-H showed clear and defined *Per2* circadian patterns. Surprisingly, a 12 h divergence has been achieved between HL and LH at day 2 of FR. Interestingly, HH condition together with CTR does not show substantial phase shift, even if 12 h peaks are present during the stimulations. These results, summarized in Figure 2.4f, suggest that a cyclic metabolic stimulation, which includes a L phase (mimicking food restriction conditions), is required to reset the cell-autonomous clock. Notably, HL and LH match the phase of light/dark cycle with a proper alignment of high metabolic phase with maximum of *Per2* expression. Interestingly, HL and LH show a phase shift of 12 h after 3 days of cyclic entrainment, which means that the metabolic entrainment has a very strong impact on circadian rhythms.

This experimental result clearly demonstrates that prolonged fasting/feeding cycles are crucial in the cell-autonomous clock resetting. Moreover, it suggests that starvation may play a crucial role in peripheral clock resetting.

2.5.3 Dissection of feeding and fasting contributions for phase resetting

The 24 h cyclic perturbations described previously consist of 12 h periods of caloric delivery and caloric restriction. However, it is not possible to assess the individual role of feeding as opposed to fasting in driving phase reset. To dissect the individual contribution of feeding- and fasting-associated signaling pathways (Figure 2.5a), fibroblasts have been perturbed with a number of small molecules that target specific pathways involved in circadian regulation. Specific inhibitors of insulin signaling pathway (TNF-α, mTOR, PI3K and GSK3) for investigating pathways associated with feeding have been used, whereas resveratrol, a potent activator of Sirt1, has been used for its association with caloric restriction (Mansur et al., 2017; Oike and Kobori, 2008).

Figure 2.5b illustrates the phase shift relative to the control at day 2 after stimulation. On the one hand, the data clearly suggest that all inhibitors acting on metabolic pathways activated during the feeding period provide a phase advance of at most 4 h for PI-103 and CHIR (inhibitors of PI3K and GSK3, respectively). These results are consistent with data published recently (Chaves et al., 2014; Sato et al., 2014). On the other hand, resveratrol, which is a starvation enhancer, leads to large phase delay of 7.4 $\pm$ 0.8 h. These results are likely to suggest that resveratrol associated pathways deeply affect circadian clock. Resveratrol is a strong activator of Sirt1, which is NAD^+-dependent protein deacetylase (Lagouge et al., 2006; Nakahata et al., 2008). During fasting, levels of NAD+ are high, and the activity of SIRT1 is high, while, when energy is in excess, NAD^+ is converted to NADH, and SIRT1 activity decreases.

Consistently with this evidence, Figure 2.5c shows the effect of resveratrol on *Per2* expression in two different metabolic states, FR-H and -L, compared with the control conditions. As shown in Figure 2.5d,

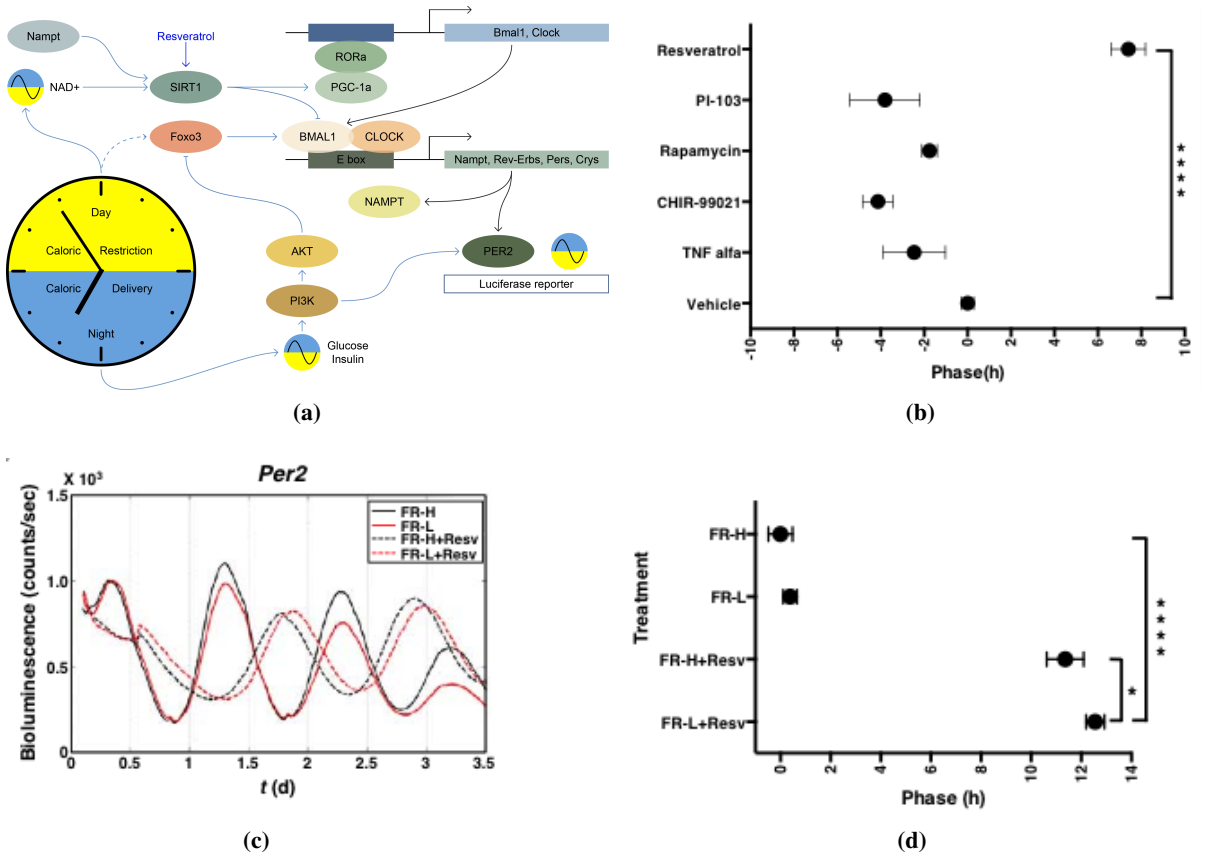


Figure 2.5: The role of fasting/feeding in inducing circadian entrainment.
(a) Regulatory network of the clock core genes; the clock receives feedback from day and night pathways. **(b)** Phase of the conditions treated with resveratrol and inhibitors; the phase is calculated at the second peak observed after addition of inhibitors. Inhibitors were added to medium with 25 mM glucose concentration. Data are represented as mean ± s.d., N=9 for each group. ****P<0.0001 one-way ANOVA with Tukey's multiple comparisons test. **(c)** PER2::LUC bioluminescence rhythms acquired during FR and FR with resveratrol in H and L conditions. Data are represented as mean ± s.d., N=4 for each group. **(d)** Phase of all conditions; the phase is calculated at the second peak observed after addition of inhibitors, using FR-H as control. Data are represented as mean ± s.d., N=4 for each group. *P<0.05, ****P<0.0001 one-way ANOVA with Tukey's multiple comparisons test.

resveratrol provides a phase delay up to 11 and 12 h in H and L, respectively. The effect of resveratrol is more significant in L condition, likely because in L condition Sirt1 is more active.

According to the hypothesis of Sirt1 activation, whether metabolic restriction alone is capable of perturbing the circadian clock by accumulation of NAD^+ during prolonged starvation has been verified. Different levels of glucose for dissecting the role of starvation have been imposed as schematically represented in Figure 2.6a. After a period of one day of pre-treatment in L for resetting the system, the cells have been stimulated with increasing levels of glucose (0, 2 and 5 mM) for one day. FR has been performed in 2, 5 and 25 mM of glucose. Figure 2.6b shows the phase of the fourth peak for all conditions and confirms the relevant role of starvation; enhanced starvation condition (0 and 2 mM of

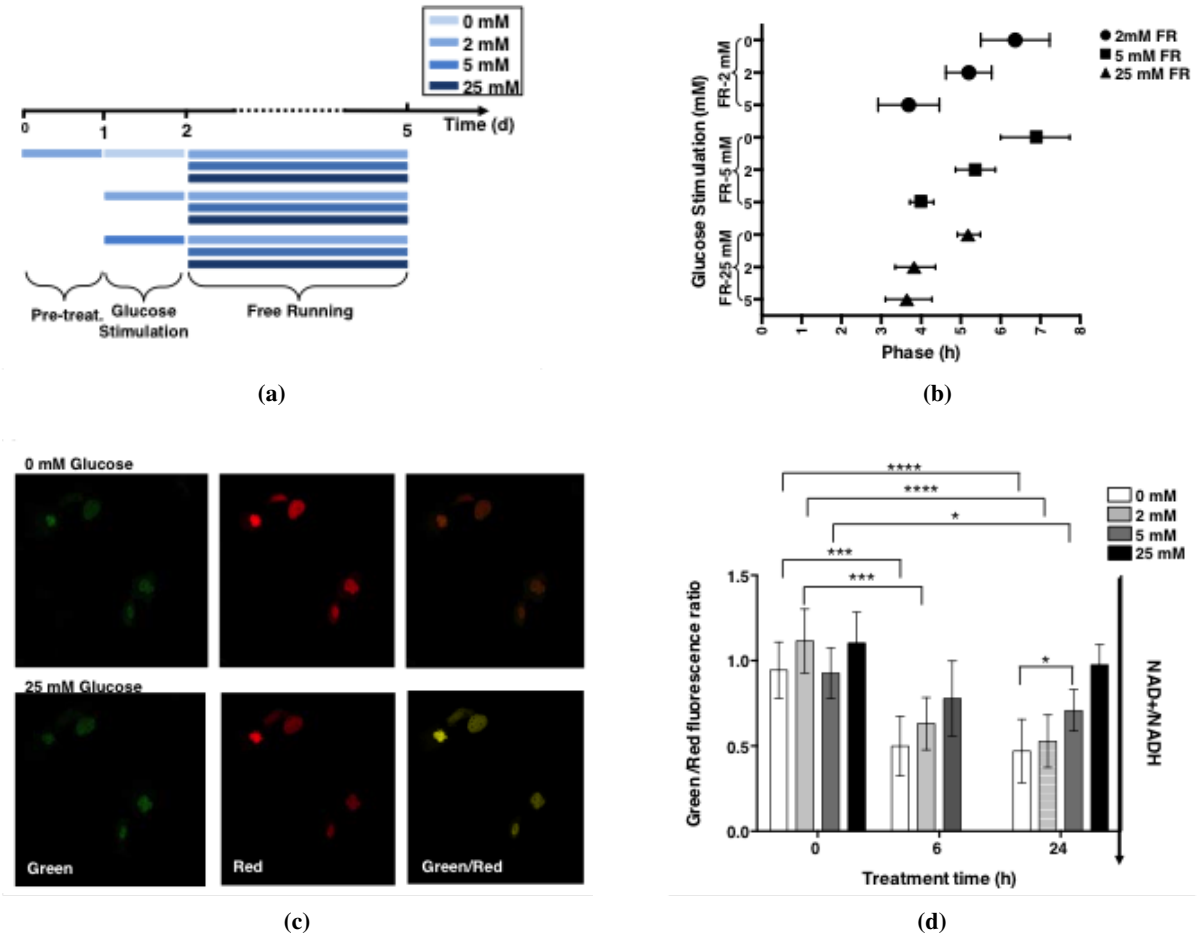


Figure 2.6: Fasting-mediated circadian phase shift of *Per2* expression.
(a) Schematic representation of starvation imposed at different levels of glucose. **(b)** Comparison of phase of PER2::LUC expression among all starvations, calculated at the second peak of FR. Data are represented as mean ± s.d., N=3 for each group. **(c)** Confocal green and red fluorescence and green-red ratio images of PER2::LUC fibroblasts expressing Peredox after 0 mM glucose (top) and 25 mM glucose (bottom) treatments. **(d)** Green/Red fluorescence responses of PER2::LUC fibroblasts measured at different time points (t=0, 6 and 24 h) for the 4 glucose levels (0, 2, 5 and 25 mM). Data are represented as mean ± s.d., N=8 for each group. . *P<0.05, ***P<0.001, ****P<0.0001 one-way ANOVA with Tukey's multiple comparisons test.

glucose) leads to a more relevant phase delay, at most 5.9 h, that could derive from NAD^+ accumulation.

These results have been supported by the measurement of NADH-NAD^+ redox state of PER2::LUC fibroblasts within the microfluidic platform through the engineered fluorescent biosensor Peredox (Hung and Yellen, 2014), as shown in Figure 2.6c. The cells have been transfected with the probe and maintained in 25 mM glucose for 16 h before the measurement; at time t=0 the cells have been treated with three different levels of glucose (0, 2 and 5 mM), with a control at 25 mM. NAD^+/NADH redox state becomes significantly more reduced as long as glucose decreases over time; this is even more evident at 0 and 2 mM glucose concentration (Figure 2.6d). It is also interesting to consider that there is a significant difference in NAD^+/NADH accumulation at 24 h between 0 and 5 mM glucose, which is consistent with

the crucial role of starvation in resetting the circadian rhythm.

Notably, circadian clock controls the NAD salvage pathway via the enzyme NAMPT, catalyzing a key step in the synthesis of NAD. This regulation governs the NAD^+/NADH ratio. Conversely, SIRT1 regulates directly the expression of clock and clock-controlled genes via deacetylation of clock proteins and histones, creating a feedback loop between redox homeostasis and circadian rhythmicity (Eckel-Mahan et al., 2013; Imai and Guarente, 2014; Ramsey et al., 2009).

These experimental findings show the relevance of starvation in driving peripheral clock phase resetting, and suggest that the molecular mechanism may be regulated by Sirt1.

2.6 Conclusion

All studies reported in the literature investigating the impact of metabolic and endocrine signals on peripheral clocks have been focused mainly on single pulse stimulation, even though the intrinsic nature of physiological metabolic and endocrine stimuli is periodic and oscillatory.

Therefore, an innovative approach able to overcome the substantial limitations of conventional experimental macro-scale approaches, i.e. Petri dish cell culture, has been developed by means of microfluidic technology, as well as the experimental strategy to mimic cyclic oscillatory metabolic cues on an *in vitro* model.

The experimental results show that fasting/feeding cycles are sufficient for driving the cell-autonomous clock resetting, confirming that metabolic entrainment has a very strong impact on circadian rhythms. By cyclic metabolic perturbations it is possible to obtain larger phase shifts than the ones obtained by a single pulse stimulation. Moreover, in the dissection between feeding and fasting individual contributions, the crucial role of fasting in driving phase resetting has been identified, and it has been suggested that the molecular mechanism might be regulated by SIRT1, which expression is under circadian control and conversely it regulates the expression of clock genes, highlighting the strong interconnection between circadian clock and metabolism.

However, even though this experimental approach overcomes limitations of conventional technology, achieving a high spatial control of cell culture microenvironment and the possibility to perform frequency-encoded stimulations, it is still not feasible to perturb the system with more complex combinations of stimulations frequencies. Moreover, the timing of stimuli might not be precise when stimuli are given by an operator. Therefore, a more robust *in vitro* model can be established, by exploiting all advantages of microfluidics, including the possibility to automate the system, which allows to mimic both physiological and pathological conditions.

Chapter 3

Automation of frequency-encoded perturbations

This Chapter describes the design, the fabrication and the validation of a multilayer microfluidic platform that enables automated frequency-encoded stimulations at a high spatio-temporal resolution for circadian studies.

3.1 Microfluidic large scale integration

Microfluidic large-scale integration (mLSI) refers to the development of microfluidic chips with thousands of integrated micromechanical valves and control components, which enable hundreds of assays to be performed in parallel with multiple reagents in an automated manner. It has been used in applications such as high-throughput screening and single cell analysis (Melin and Quake, 2007; Thorsen et al., 2002). mLSI aim is to automate biology, being a candidate for replacing today's conventional automation paradigm, which consists of fluid-handling robots (Melin and Quake, 2007).

mLSI devices are fabricated with the well-established multilayer soft lithography technique by which multilayer microfluidic platforms can be produced. Multilayer microfluidic platforms are formed by two or three Polydimethylsiloxane (PDMS) layers bound one on top of the other, and then sealed on top of a rigid substrate (normally a glass slide). In this Dissertation only two layers microfluidic platform will be considered. These platforms are formed by a control layer and a flow layer (Figure 3.1a). Control layer harbors all channels required to actuate the valves, situated at the bottom of the flow layer, which contains the network of channels being controlled; its thickness is around 50 μm. All biological assays and fluid manipulations are performed on the flow layer (Thorsen et al., 2002). A valve is created where a control channel crosses orthogonally a flow channel. The resulting thin membrane in the junction between the two channels can be deflected by pneumatic actuation: when pressure is applied to the fluid flowing into

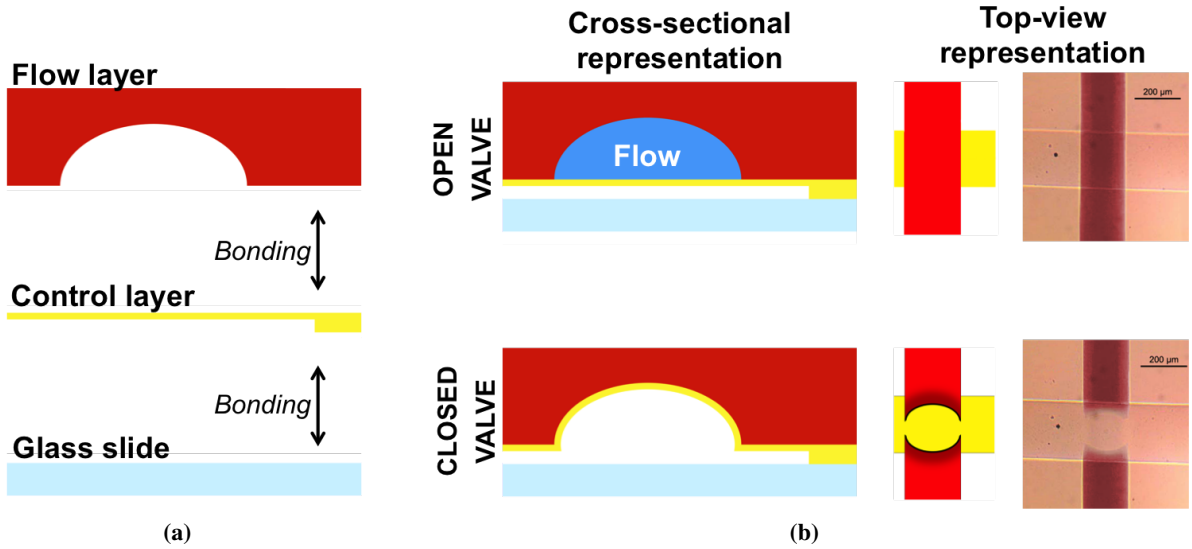


Figure 3.1: Schematic representation of the layers of a multilayer microfluidic platform and of the monolithic microfluidic valve.
(a) Schematic representation of the structure of a multilayer microfluidic platform. PDMS flow layer (thickness of 3 mm) is bound on top of PDMS control layer (thickness of 50 μm). These layers are bound together first, and then on a rigid substrate such as a glass slide. **(b)** Schematic cross-sectional and top-view representation and image of a monolithic microvalve, when the control channel is not pressurized (valve open) and when the control channel is pressurized (valve closed).

the control channels, the membrane bends, causing sealing of the flow channel, thus preventing fluid flow inside it (Figure 3.1b). Introduction of fluid into these devices is accomplished through steel pins inserted into holes punched through PDMS, since silicone is soft and forms a tight seal around input pins.

In order to achieve automation, mLSI technology has two main requirements: an actuator to control fluids, and a method for addressing and controlling these valves (Thorsen et al., 2002). Monolithic micromechanical valves are the actuators to control fluid, being the building block of mLSI and the basic unit for fluid-handling functionality; they are leakproof, scalable and resistant. The control is achieved by computer-controlled external solenoid valves that allow actuation of microvalves.

Monolithic micromechanical valves are the most simple unit used by mLSI; by combining in series at least 3 valves, micropumps are formed and they can be used to control precisely the flow rate of fluid flow. Also micromixers and separators can be included in the design, in order to achieve homogeneous composition (overcoming the fluid segregation typical of the laminar flow regime) or to separate particles based on their size, respectively. Moreover, on-chip motion is more uniform compared to the one obtained by lab-scale syringes or peristaltic pumps, thus a higher control of conditions in the chip is achieved.

By increasing the number of valves, microfluidic platforms can become very versatile and suitable for many applications; however, they parallel become highly complex integrated systems, with a high

number of control lines and complex flow architectures, thus slowing down their applicability and their actual use specially in the industrial laboratories and in routine processes (Ouellette, 2003).

Therefore, by mLSI technology is possible to combine the advantages of microfluidics, which are precise and accurate spatial control of fluid composition, concentration of secreted factors and reduction of reagents usage, to the advantages of automation, which are precise timing of the stimuli and feasibility of all stimuli combinations and timing.

3.2 Motivation and aim

In the obtainment of efficient human *in vitro* models for study diseases and biological processes, the ability to study cells under simulated *ex vivo* physiological microenvironments is of crucial importance. To achieve this goal, it is necessary to define a method to stimulate the cell culture in a physiological way and to enable quantitative real time control of the cellular response.

The miniaturization of the systems can offer many advantages, such as precise and accurate spatial control of fluid composition, physiological concentration of secreted factors and reduction of reagents usage; but in order to manage tiny volume with high spatio-temporal distribution, microfluidic systems are useless without accurate liquid handling, distribution and control.

Circadian rhythms are processes that take place over time and that are entrained by a series of frequency-encoded environmental stimuli, such as the periodic alternation of day and night and meal timing. Therefore, an *in vitro* model aimed at studying circadian clock entrainment due to different stimuli, mimicking *in vivo* conditions, should allow to perform frequency-encoded stimuli.

The *in vitro* model used for the biological investigation reported in Chapter 2, allowed to achieve a high homogeneity in the cell culture environment as well as high control of the chamber composition, to investigate the effects of frequency-encoded stimulations. However, in a manually operated model, timing of the stimuli may not be highly precise and frequency-encoded stimulations are not always feasible.

Therefore, the specific aim is to develop an *in vitro* model to perform automated frequency-encoded stimulations at highly precise timing in order to resemble physiological and pathological conditions at which peripheral tissues are exposed *in vivo*. The goal has been achieved by mLSI technology, which allowed to design and fabricate a versatile multilayer microfluidic platform to perform circadian studies with frequency-encoded stimulations, increasing the experimental throughput without requiring assistance for extended period of time. The design of the platform has been a trade-off between flexibility and complexity of the platform, in order to keep it robust and as user-friendly as possible.

3.3 Multilayer microfluidic platform for circadian applications

In this Section the rationale for the design of the multilayer microfluidic platform will be presented, followed by the procedure for fabricating and automating the platform.

3.3.1 Rationale of platform design

The design of a microfluidic platform is driven by the fulfillment of technical and biological requirements. The general idea of the design of this platform is to create the simplest design that allows to perform frequency-encoded stimulations, varying frequency, sequence and total number of stimuli independently between biological conditions performed within the same platform. Moreover, the design should minimize the number of control lines, in order to make as simple as possible the handling of the platform. The platform must be suitable for cell culture, it has to permit to switch between two different stimuli and it must include four independent conditions within the same platform, each one with four independent biological replicates.

The maximum surface of the platform has to be 60x40 mm, since PDMS layers will be bound on a 75x50 mm glass slide (a factor of 0.8 is used for maximum dimensions to ensure high yield in fabrication procedure).

It is decided to use the pressure as the driving force for fluid motion, coherently with the idea to keep the design as simple as possible (rather than including a peristaltic micropump in the design, which would have allowed to manipulate the flow rate of the flow layer).

The design of the platform is shown in Figure 3.2.

It can be observed that it has two inlets, which connect to all 16 culture chamber inlets; culture chambers are grouped in four independent groups made of four chambers each. All chamber outlets are connected to the same outlet channel. In order to avoid heterogeneous composition inside culture chambers (due to diffusion from inlet or outlet channels or from the other chambers) each culture chamber is isolated by monolithic micromechanical valves, which are closed during normal operations and selectively opened when stimuli are given. On-chip valves are also present in correspondence to the two inlets as well as close to the outlet. Valve area is 170x200 μm, wider than minimum required surface, to ensure lower actuation pressure, thus reducing the mechanical stress of the valves. Each culture chamber is 2 mm wide, 10 mm long and it has a height of 200 μm.

3.3.2 Fabrication

In this Section the technique to fabricate molds, i.e. multilayer soft lithography, will be described, as well as the procedure to fabricate the multilayer microfluidic platform.

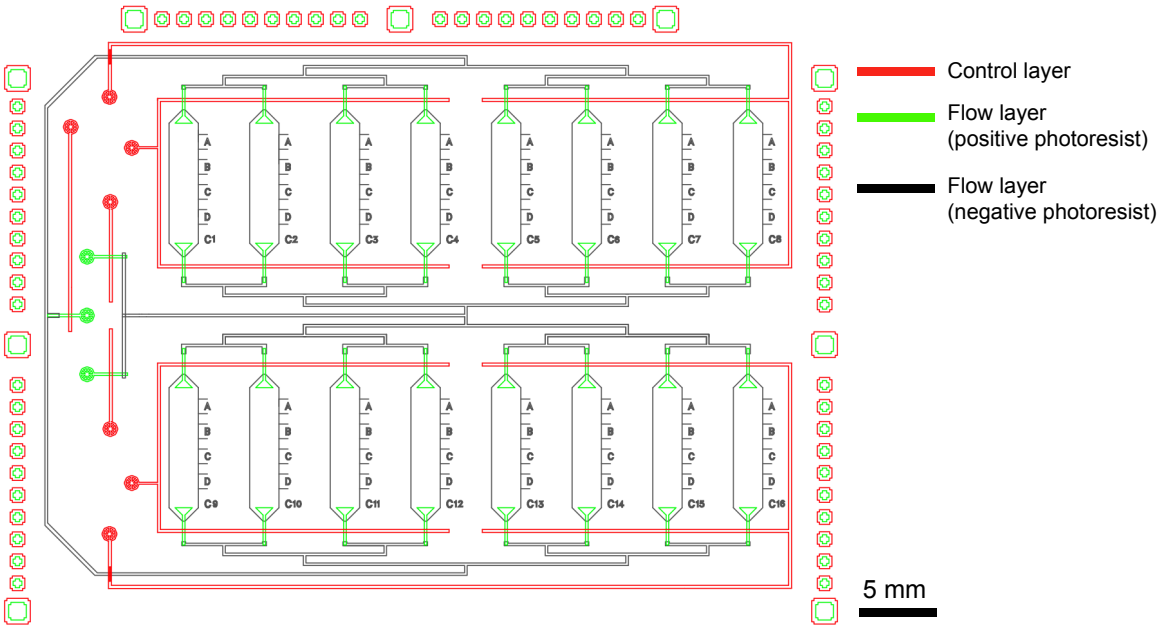


Figure 3.2: Design of the multilayer microfluidic platform: the control layer is reported in red, the flow layer done with positive photoresist is reported in green, the flow layer done with negative photoresist is reported in black.

3.3.2.1 Multilayer Soft Lithography

Multilayer soft lithography is the technique used to produce molds for the multilayer microfluidic platform; it combines soft lithography with the capability to bond multiple patterned layers of elastomer (Unger, 2000). As reported in Section 3.1, multilayer platforms are formed by binding two (or three) PDMS layers, each of which is separately cast from a micromachined mold. The two molds required for this platform are defined as the control mold and the flow mold, which will be used to pattern the control and the flow layer, respectively. Both molds are multilayer photoresist-based molds. The detailed protocol and materials for multilayer soft lithography are given in Appendix A on page 93.

Control mold is made by two layers of negative photoresist that allows to form channels with a rectangular cross-section; the first layer, which is deposited on the silicon wafer, is defined as the blanket, and it provides a surface with uniform wetting properties for PDMS casting. The second layer contains the pattern of the channels of the control mold.

Flow mold is made by one layer of positive photoresist and one layer of negative photoresist; positive photoresist is used in the portions of the channels where a valve will be formed, since to enable the valve to seal completely during actuation, the flow-channel must have a rounded profile and positive photoresist allows to shape channels with rounded section. The other portions of the channels and culture chambers are done by negative photoresist, thus they will have a rectangular cross-section.

The photomasks used for patterning control and the flow molds are given in Appendix A on page 93.

3.3.2.2 Replica Molding

Once molds are produced by soft lithography, PDMS layers are cast on molds. PDMS consists of two components: a base and a curing agent. They are mixed in a 10:1 weight ratio, to obtain a material with good mechanical properties (with a 10:1 ratio a proper balance between stiffness and elasticity is obtained for multilayer platforms). PDMS is poured on the flow mold and thermally cured; the thin PDMS layer of the control layer is obtained by spinning PDMS on the control mold at a spinning speed that ensures to obtain the desired thickness, and then it is thermally cured.

After punching inlets and outlet in the flow layer, it is bound to the control layer by air plasma treatment (which creates strong covalent bonds between PDMS layers). Then the two layers are punched together to created the inlets for control layer and finally they are bound to the glass slide, which has been previously covered by a very thin layer of PDMS (5-10 μm), to ensure better adhesion between control layer and the rigid substrate. The detailed replica molding protocol is provided in Section A.5 on page 102.

Figure 3.3 shows the fabricated microfluidic multilayer platform.

In Figure 3.3 in red it is possible to see the flow layer: it is composed by chambers and ramifications of channels and on the left-hand side of the platform the two inlets and the outlet are clearly visible. In yellow it is possible to see the control layer: it is composed only by dead-end channels which intersect the flow ones (forming 35 on-chip valves) and on the left-hand side of the platform there are its seven inlets.

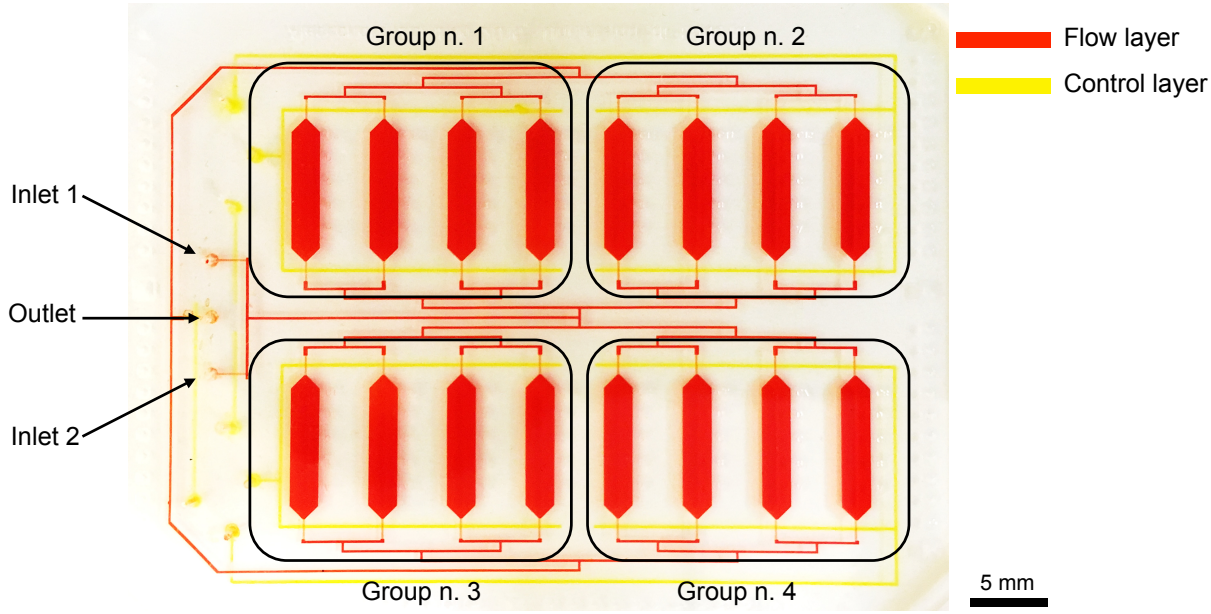


Figure 3.3: The image shows the multilayer platform, with the flow layer depicted in red and the control layer depicted in yellow.

3.3.3 Automation

The automation is achieved through valve control and a custom software, developed using LabVIEW (Laboratory Virtual Instrument Engineering Workbench, National Instruments). LabVIEW is needed to create the interface for controlling the instrumentation from a computer; a proper LabVIEW code has to be prepared depending on the specific application.

All the on-chip valves are driven by pneumatic solenoid valves which are in turn controlled by an electronic unit for the generation of digital signal connected to the USB port of a computer. Each solenoid valve can switch on-chip valve between atmospheric pressure (on-chip valve open) and 250 kPa (on-chip valve closed) through a pressurized liquid; liquid pressure is controlled through a pressure regulator. The details about automation instrumentation are given in Section A.6 on page 103.

The developed interface is user-friendly, it allows to switch from a manual mode, in which the operator can open and close valves one by one as it is needed (for example during coating and cell seeding), to an automated mode, where the duration of medium change, the time interval between medium changes, which one of the two media should be be used in a specific group of chambers and the total number of media changes can be set at the beginning of the experiment, without requiring any further assistance of the operator for all the duration of the experiment.

3.4 Microfluidic platform validation

The designed microfluidic platform has to be characterized from the mechanical and fluid dynamic point of view and than validated for biological applications prior to be used for relevant biological experiments. This Section reports the the mechanical and fluid dynamic characterization and the biological validation.

3.4.1 Platform characterization

This section describes the results concerning the mechanical and fluid dynamic characterization of the multilayer microfluidic platform.

3.4.1.1 Determination of volumetric flow rate vs pressure

As stated in the rationale of the design of the platform (Section 3.3.1), the fluid flow in the microfluidic platform is pressure-driven. Therefore, it is very important to estimate the relation between volumetric flow rate and pressure to quantify the minimum duration of media changes to ensure a complete change of the composition inside all culture chambers. The variation of volumetric flow rate as a function of pressure is shown in Figure 3.4.

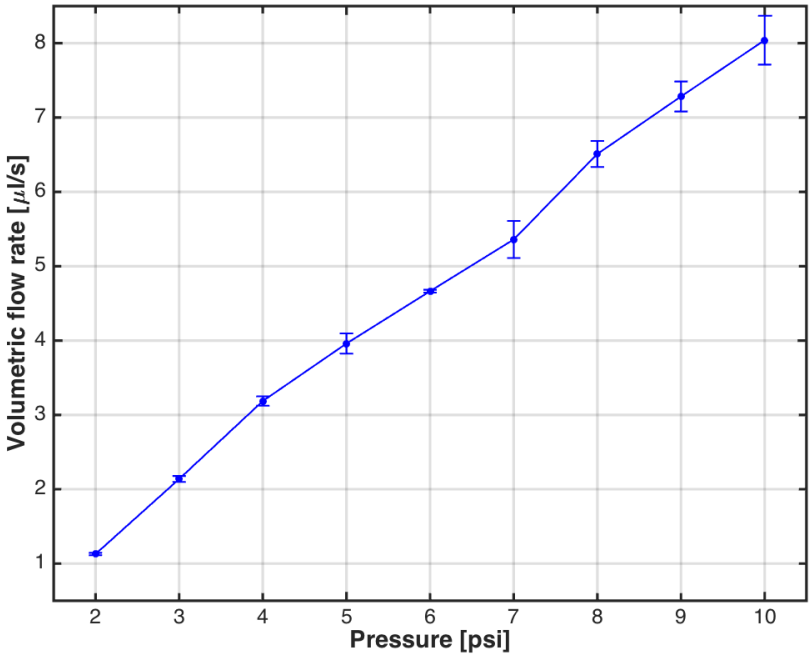


Figure 3.4: Volumetric flow rate [μl/s] vs pressure [psi] in multilayer platform (N=3).

In the operative pressure range, i.e. from 2 psi to 10 psi, the flow rate increases from 1.1 μl/s to 8 μl/s, respectively, with a linear trend. A linear trend was expected since the flow regime in the microfluidic chamber is laminar, thus the relation between fluid velocity and pressure drop is linear; if the flow regime had been turbulent, the relation between pressure and volumetric flow rate would have been quadratic.

3.4.1.2 Mechanical resistance of the platform

Since multilayer microfluidic platforms are pressurized during the entire duration of experiments, it is important to test their mechanical resistance, to ensure that they are properly and reliably working until the end of experiments. Mechanical resistance of the platform means that it does not delaminate (i.e. PDMS layers do not detach one from the other one or from the glass slide) as well as all valves do not collapse or leak for all the duration of the experiment. Circadian experiments normally last for 5 days, therefore the platforms have been tested for a longer time, i.e. 7 days.

Table 3.1 reports the results of the mechanical resistance tests.

From Table 3.1 it is possible to see that 80% of the platforms properly worked for at least 5 days, and 60% of them worked for 7 days. Note that Experiment #2 could have lasted longer, since it had been stopped for a technical problem in the compressed air supply line.

The pressure of the flow has been set equal to 5 psi in order to ensure a proper volumetric flow rate (thus duration of media changes) without exposing the cells to a too high shear stress. ΔP between the two layers has to be equal to the minimum pressure that closes all valves, which is usually around 10 psi

Table 3.1: Results of mechanical resistance experiments and their relative operating conditions (✓ indicates that the platform successfully worked and × indicates that the platform did not work).

Experiment	Flow pressure [psi]	Control pressure [psi]	Testing Time [day]	Result
#1	5	15	7	✓
#2	5	15	5	✓
#3	5	15	7	✓
#4	5	18	7	✓
#5	5	/	/	×

(the actual value depends on the thickness of the membrane); by adding ΔP to the pressure of the flow, the pressure of the control is determined. The higher the pressure of the flow, the higher the pressure of the control, thus the stronger the mechanical stress on the platform.

Among the tested ones, only one platform did not work due to a fabrication problem (the membrane had a thickness of 90 μm, thus it required a very high actuation pressure, higher than 30 psi).

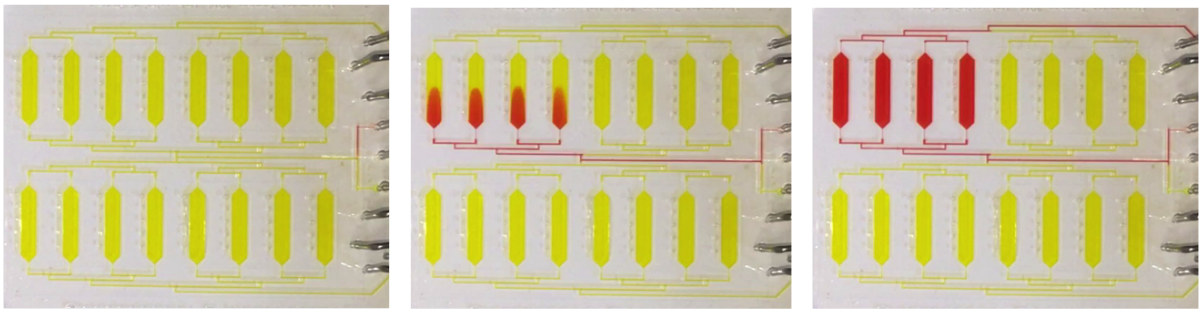
This experimental result is very important in the assessment of the suitability of the designed multilayer microfluidic platform for circadian experiments: if the platform is correctly fabricated, it works for all the duration of the experiment.

3.4.1.3 Fluid dynamic characterization

Since different biological conditions are performed in the four groups of chambers in the multilayer microfluidic platform, it is important to prove that the four groups have the same fluid dynamic behaviour, in order to avoid biased biological responses. Moreover, it is also important to verify that the four independent chambers belonging to the same group have the same fluid dynamic behaviour, in order to be sure to relate variability in cellular responses only to biological variability.

To quantitatively characterize fluid dynamic behaviour, media changes have been performed with food colorants at different combinations of flow and control pressure, and the quantification has been performed by evaluating the average pixel intensity over specific Regions of Interest (ROIs), placed close to the inlet and to the outlet of culture chambers. A generic example of media change with food colorant from yellow to red in one of the group of chambers can be seen in Figure 3.5.

To evaluate homogeneity of fluid dynamic behaviour between the four groups of chambers, the average signal of the 4 ROIs placed close to the inlet of the culture chambers of the same group has been compared to the average signal of the other 3 groups of chambers, and the same has been done with the



(a) **(b)** **(c)**

Figure 3.5: Time course of a medium change in one of the four group of chambers.
(a) Initial situation (t = 0 s). **(b)** Intermediate situation (t = 3 s). **(c)** Final situation (t = 7 s).

average signal measured in correspondence to the outlet of the chambers.

Figure 3.6, Figure 3.7 and Figure 3.8 show the homogeneity of fluid dynamic behaviour of the four groups of chambers. The profiles have been measured keeping the pressure of the control layer at 25 psi and setting the pressure of the flow layer at 3 psi, 5 psi and 10 psi, respectively[1]. The value of the intensity has been normalized between 0 and 1, where 0 corresponds to the intensity of the initial food colorant (in this case yellow) and 1 corresponds to the intensity of the final situation (in this case red colorant).

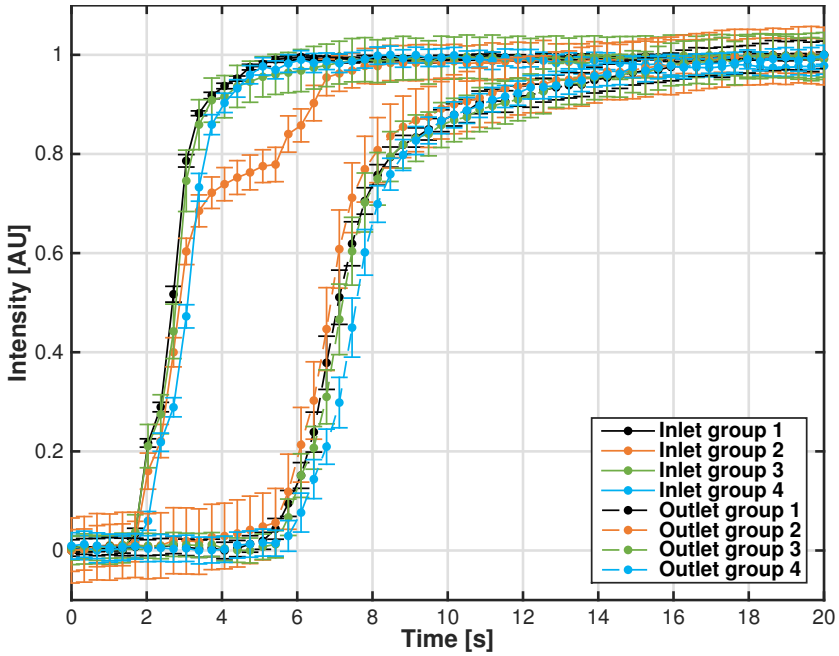


Figure 3.6: Average intensity [AU] vs time [s] of ROIs placed at the inlets and outlets of different groups of chambers for a flow pressure of 3 psi and a control pressure of 25 psi (N=3).

[1] 3 psi is the minimum pressure to have fluid flow in the microfluidic platform, due to pressure drops; 5 psi is the normal operating condition, which ensures acceptable duration of media changes and tolerable shear stress on the cells; 10 psi is considered the highest pressure that could be used since previous experimental results showed that cells can tolerate the shear stress caused by fluid flow driven by a pressure of 15 psi at maximum; moreover, with this flow pressure, the pressure of the control must be 25 psi, thus the platform will be exposed to a high mechanical stress that may cause delamination

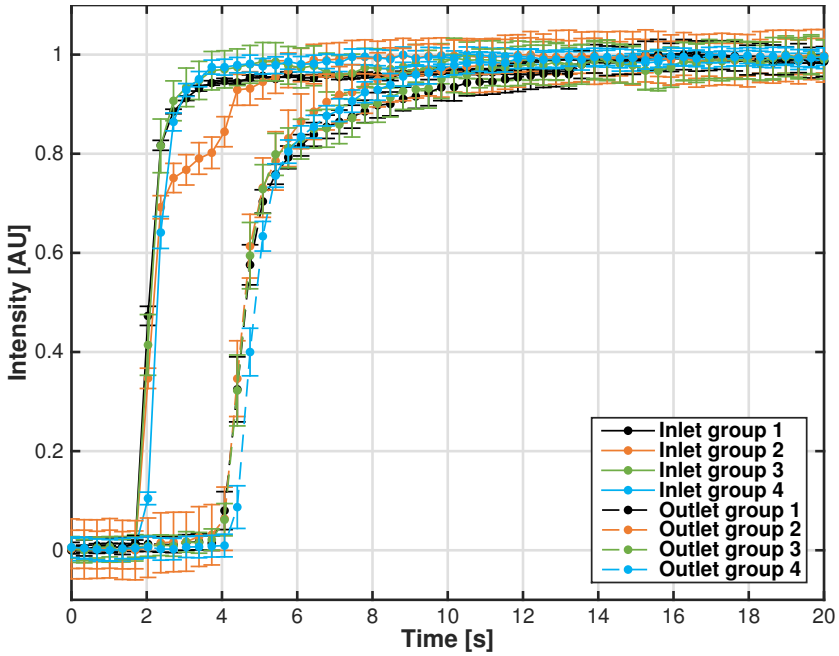


Figure 3.7: Average intensity [AU] vs time [s] of ROIs placed at the inlets and outlets of different groups of chambers for a flow pressure of 5 psi and a control pressure of 25 psi (N=3).

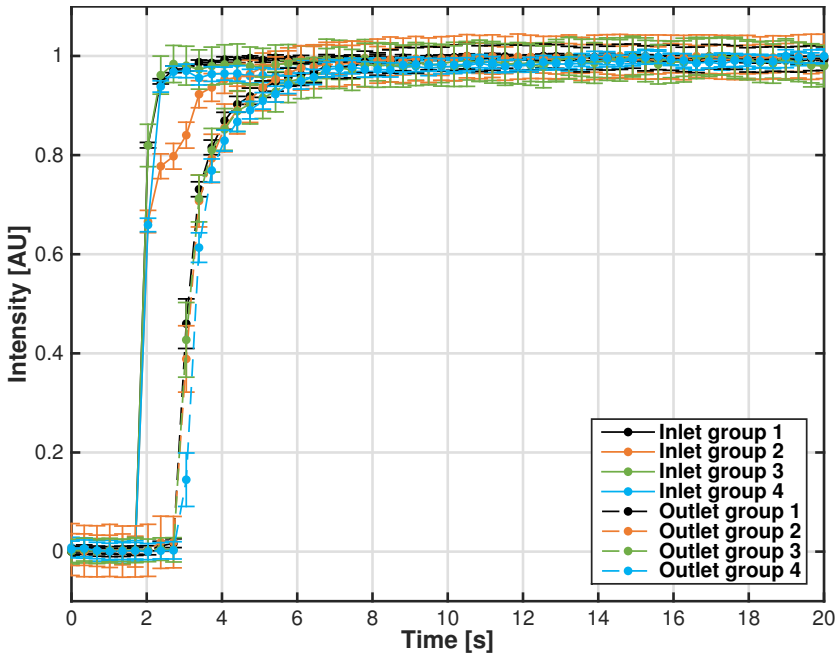


Figure 3.8: Average intensity [AU] vs time [s] of ROIs placed at the inlets and outlets of different groups of chambers for a flow pressure of 10 psi and a control pressure of 25 psi (N=3).

From Figure 3.6, Figure 3.7 and Figure 3.8 it is possible to observe that all groups behave in a comparable way from the fluid dynamic point of view; only group 2 is slightly different (only in correspondence to the inlet) but it reaches comparable steady state conditions with the other 3 groups within the same time. Increasing the pressure of the flow causes more homogeneous flow behaviour between the four groups. The lag phase of around 2 seconds that is observed for all inlets profiles is due to solenoid and

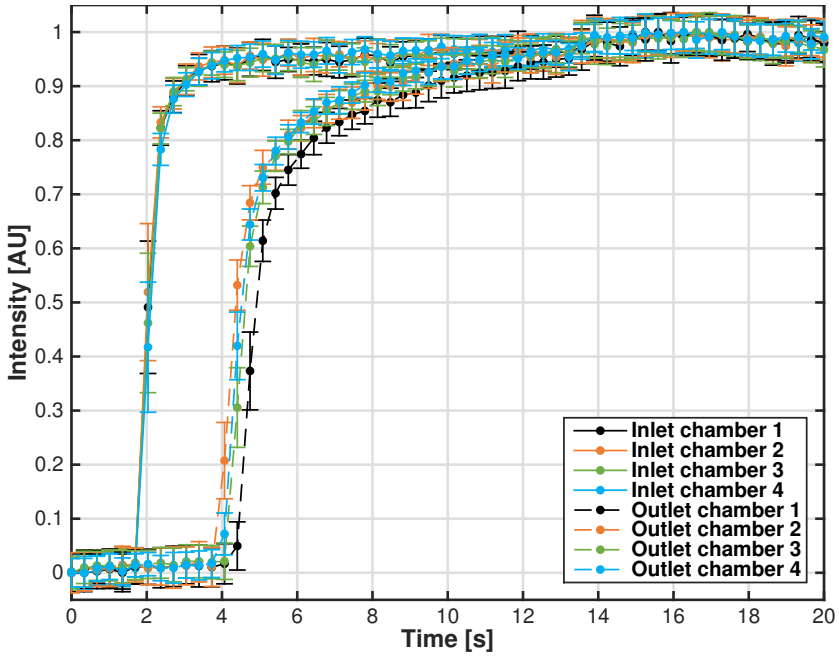


Figure 3.9: Intensity [AU] vs time [s] of ROIs placed at the inlets and outlets of group of chambers 1 for a flow pressure of 5 psi and a control pressure of 25 psi (N=3).

on-chip valves dynamics and for the time the fluid takes to reach culture chamber inlets from the inlet of the platform (time 0 corresponds to the time when the valve of one of the two inlets is opened). The time required for outlets of the chambers to reach the plateau decreases with the increase of pressure; by looking at outlet profiles it is possible to quantify the time required for a medium change, which are 16 s, 12 s and 8 s, for a flow pressure of 3 psi, 5 psi and 10 psi, respectively.

The effects of control pressure on the fluid dynamic of the platform has also been evaluated (keeping flow pressure constant), but no significant differences were observed on the fluid dynamic behaviour.

In order to verify that the four chambers belonging to the same group behave in a similar way, the signal has been quantified at the inlet and at the outlet of each chamber. Here it is reported only the quantification of the signal of chambers belonging to the first group of chambers, which has been measured keeping the pressure of the control at 25 psi and fixing the pressure of the flow at 3 psi, 5 psi and 10 psi. The profiles relative to flow pressure of 5 psi and control pressure of 25 psi are shown in Figure 3.9.

Fluid behaviour of different chambers is similar for all the different flow pressures, since the curves have a similar profiles; the only exception is the outlet of chamber 4, which is slightly slower than the others to reach plateau condition. The same considerations done for the average intensity of the four groups can be done.

Fluid dynamic characterization of multilayer microfluidic platform has shown that the fluid behaviour of the chambers belonging to the same group and of different groups of chambers is similar, thus, the same fluid dynamic is ensured for all biological replicates performed during an experiment with this

device.

3.4.2 Biological validation

This Section describes the biological validation of the multilayer microfluidic platform, showing cell seeding, growth and quantification of cell viability, as well as bioluminescence acquisition.

3.4.2.1 Cell viability

All biological validation experiments have been performed with murine fibroblasts PER2::LUC since these cells are used for circadian experiments. Cells are seeded using the automation interface developed in LabVIEW, after coating channels and culture chambers with Fibronectin at 25 μg/ml. Cells are seeded at seeding density of 700 cells/mm^2, to make sure cells are confluent.

The homogeneity of cell seeding within each chamber and in all chambers is checked right after seeding, and cell growth and attachment have been monitored; the results are shown in Figure 3.10.

From Figure 3.10 it is possible to see that cells are seeded uniformly at Day 0, close to the inlet of the culture chamber and in an area in the middle of the chamber close to the chamber wall (cells are present also very close to the border of the chamber); they are attached and confluent at Day 1, displaying a typical fibroblast morphology. Comparing Day 1 and Day 2, it is possible to see that cell morphology and density are comparable, while dead cells (which are visible being the brightest points), are present at Day 1 and not at Day 2, since they have been removed by fluid flow during media changes.

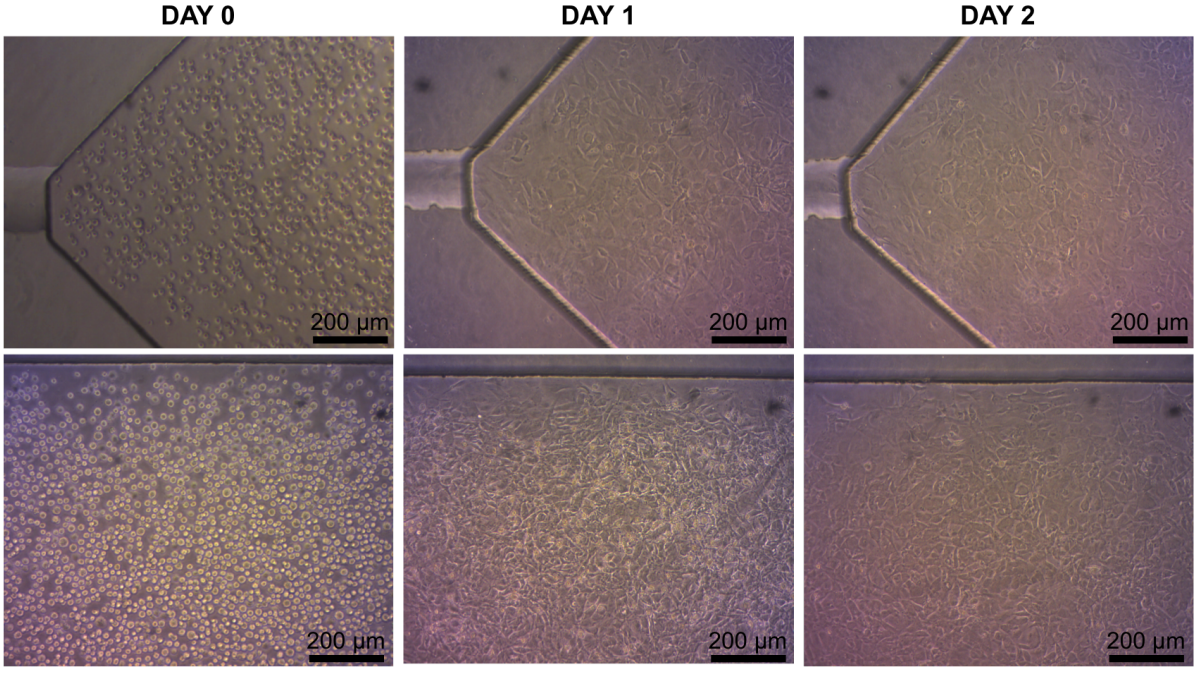


Figure 3.10: Murine fibroblasts PER2::LUC growth inside different culture chambers at Day 0 (seeding), at Day 1 (24 h after seeding) and at Day 2 (48 h after seeding). Cells are seeded at seeding density of 700 cells/mm^2.

Figure 3.11: Image of one entire culture chamber of multilayer platform at Day 7 (i.e. 7 days after seeding at a cell density of 700 cells/mm^2).

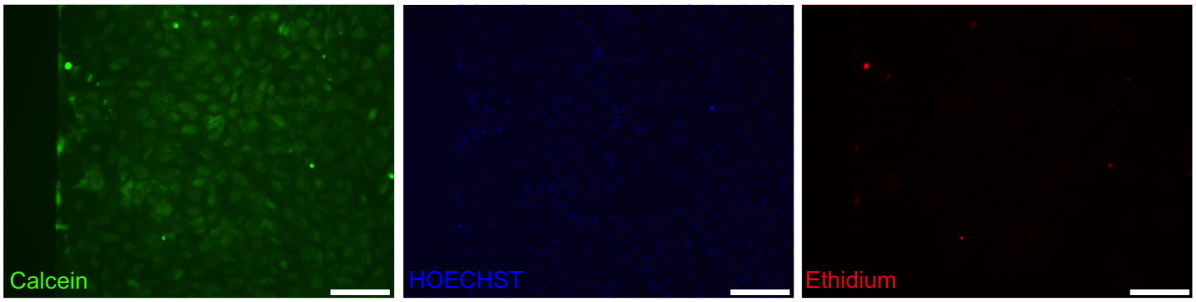


Figure 3.12: LIVE/DEAD at Day 7 (i.e. 7 days after seeding). Calcein (in green) marks live cell cytoplasm, HOECHST (in blue) marks all cell nuclei and Ethidium Bromide (in red) marks cells in the final stage of apoptosis. Scalebar corresponds to 200 μm.

After 7 days from seeding, cells have completely covered the entire surface of the culture chambers; cell homogeneity is good, also close to the walls of the culture chambers. An image showing one entire culture chamber is shown in Figure 3.11.

After 7 days from seeding, LIVE/DEAD assay is performed, in order to quantitatively assess cell viability in the culture chambers. Cell staining of the three used markers is shown in Figure 3.12. Cell viability, defined as the ratio of live cells over the total number of cells, has been quantified to 98.6 $\pm$ 0.14 %.

Thus, it is possible to state that multilayer microfluidic platform is suitable for cell growth and viability regarding this specific cell line.

3.4.2.2 Bioluminescence acquisition

Murine fibroblasts PER2::LUC, which are a luciferase reporter cell line for *Per2* gene, have been used to test the applicability of the platform for bioluminescence acquisition. From Figure 3.13 it is possible to observe the average bioluminescence signal of the four chambers belonging to the same group, with the characteristic periodic dampened oscillation profile.

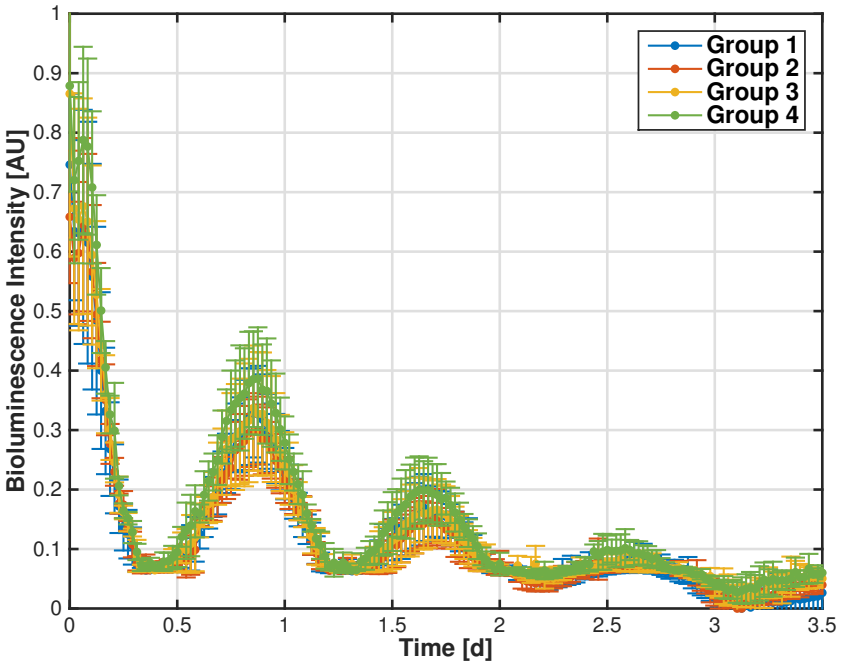


Figure 3.13: Average bioluminescence intensity [counts/min] vs time [d] of chambers belonging to the same group (N=4).

The acquired bioluminescence signal is stable for three days and its dampening, due to cell death and to cell desynchronization, is typical of *in vivo* cell cultures. From the Figure, it is possible to see that all chambers are in phase during all acquisition: this is due to the fact that they have been stimulated at the same time with the same media, thus the same expression of *Per2* is expected.

Thus, the platform is suitable for circadian applications with this cell line.

3.5 Conclusion

In this Chapter, mLSI technology has been used to design a multilayer microfluidic platform for circadian applications. As described in Section 2.4.2, monolayer microfluidic technology has been limitedly used in the circadian field, and up to now only for circadian clock studies of algae and bacteria; in the literature microfluidic monolayer platforms have never been used for circadian clock studies in mammals.

As far as multilayer microfluidic platforms are concerned, this is the first example of microfluidic multilayer platform designed for circadian clock studies, therefore it represents a completely new tool for frequency-encoded circadian clock studies, which have been shown to give new insights into the interconnection between circadian clock and CCGs involved in metabolic pathways.

The automated platform has been designed, fabricated, automated, characterized from the fluid dynamic point of view and biologically validated. It allows to perform four independent biological conditions in four replicates within the same platform, which has been proved to work successfully in normal operating conditions for 7 days. It is possible to operate the platform without requiring assistance of

operators for all the duration of the experiment thanks to the implemented automation, in which duration of medium change, time interval between medium changes, which one of the two media should be be used in a specific group of chambers and the total number of media changes can be decided. It has also been shown that the platform ensures a very high cell viability for 7 days.

Therefore, the designed multilayer microfluidic platform is suitable for *in vitro* circadian studies and it enables frequency-encoded stimulations, which are particularly relevant for circadian studies, allowing to resemble *ex vivo* physiological conditions. It combines the advantages of microfluidic technology, such as high homogeneity in the cell culture environment, high control of the chamber composition and reduction of reagent usage, with the advantages of an automated system, which are precise timing of the stimuli, operator-independent operations and feasibility of highly complex experimental designs.

Even though the experimental throughput could be increased by the use of the microfluidic platform, enabling to investigate many stimuli combinations, the main limitation consists in the fact that the expression of only one gene is monitored at a time with a bioluminescence reporter cell line, while many biological processes are under circadian control and regulation. Therefore, a more comprehensive approach is required in order to deepen circadian regulatory mechanisms and discover new clock-controlled genes.

Chapter 4

Whole transcriptome data analysis

This Chapter describes the state of the art of technology for transcriptome studies and algorithms related to analysis of temporal RNA sequencing data. It presents a new data analysis method for transcriptomic data measured over time based on Principal Component Analysis: first the method is characterized and then its possible applications are shown.

4.1 State of the art of circadian transcriptome studies

This Section describes the relevance of circadian transcriptomic studies, the available techniques to perform them and the most common algorithms for identification and characterization of circadian transcripts from whole transcriptome data.

4.1.1 The relevance of circadian transcriptome studies

Circadian regulation of rhythmic biological processes is achieved through core clock proteins, that drive the rhythmic expression of clock-controlled genes (CCGs), which are the ones mediating circadian clock's biological influence. CCGs do no participate directly in the molecular mechanism of the circadian timekeeper, but they impose rhythmicity on downstream cellular and physiological functions. As it can be understood, the number and the type of CCGs are highly tissue-specific, since specific functions of each tissue require the expression of different genes even though the core clock machinery is largely conserved among tissues (Hughes et al., 2009; Li et al., 2015; Zhang et al., 2014). Generally, the total number of cycling transcripts ranges from a few hundred to several thousand transcripts (Li et al., 2015; Zhang et al., 2014).

Thus it is really important to identify CCGs to study the molecular mechanism of circadian clock output across tissues and species and to discover pathways and processes that are clock-controlled. Moreover, grouping transcripts by phase may suggest a common underlying regulatory network under

circadian regulation (Li et al., 2015). This understanding will have a direct application in chronotherapeutics, which studies the interactions between the circadian network and pharmaceuticals, since it has been recently shown that efficacy and specially side effects of drugs are dependent on the time of the day they are administered, jet drug metabolism is under circadian control (Ballesta et al., 2017; Hughes et al., 2009; Lévi et al., 2010).

Thus, the final aim of circadian transcriptomic studies is to develop an accurate systems-level understanding of the dynamic interacting regulatory networks that underlie molecular, behavioural, and physiological circadian control, across and between tissues and species.

4.1.2 Techniques

Two techniques are available for quantitatively measuring genome-wide expression: microarrays and RNA sequencing. Over the past decade, the first circadian gene expression studies were performed using microarrays, due to their low cost and well-established analysis methods; however, in the last years, researchers have begun to use RNA sequencing for circadian experiments due to recent technological improvements and reduction in sequencing costs, as well as for intrinsic advantages of sequencing over microarrays for circadian studies (Bumgarner, 2013).

4.1.2.1 Microarray

Microarrays are microscope slides that are printed with thousands of tiny spots in defined positions, with each spot containing a known DNA sequence or gene. The DNA molecules attached to the slide act as specific probes to detect gene expression.

The schematic workflow of a microarray analysis is shown in Figure 4.1. To perform a microarray analysis, RNA is extracted from the cells and converted to a labeled cDNA or cRNA, which is allowed to bind to the microarray (hybridization); each cDNA or cRNA molecule can theoretically bind to one specific probe, located in a known position, in the microarray slide. Then, the signal is detected by measuring fluorescence at each microarray spot. The intensity of the signal on each spot is taken as a measure of the expression level of the corresponding gene (Bumgarner, 2013).

Even though microarray technology is well-established, it has a number of limitations: first, microarrays provide an indirect measure of relative concentration since the signal measured at a given position on a microarray is usually assumed to be proportional to the concentration of a single species in solution that can bind to that spot. However, kinetics of hybridization cause a non-linear dependence between signal and concentration both at high and at low concentration, thus reliable concentration estimations are possible only over a limited concentration range. Second, probes on the microarray slides are not sufficiently selective so that many genes can bind to the same probe. Finally, a microarray can only de-

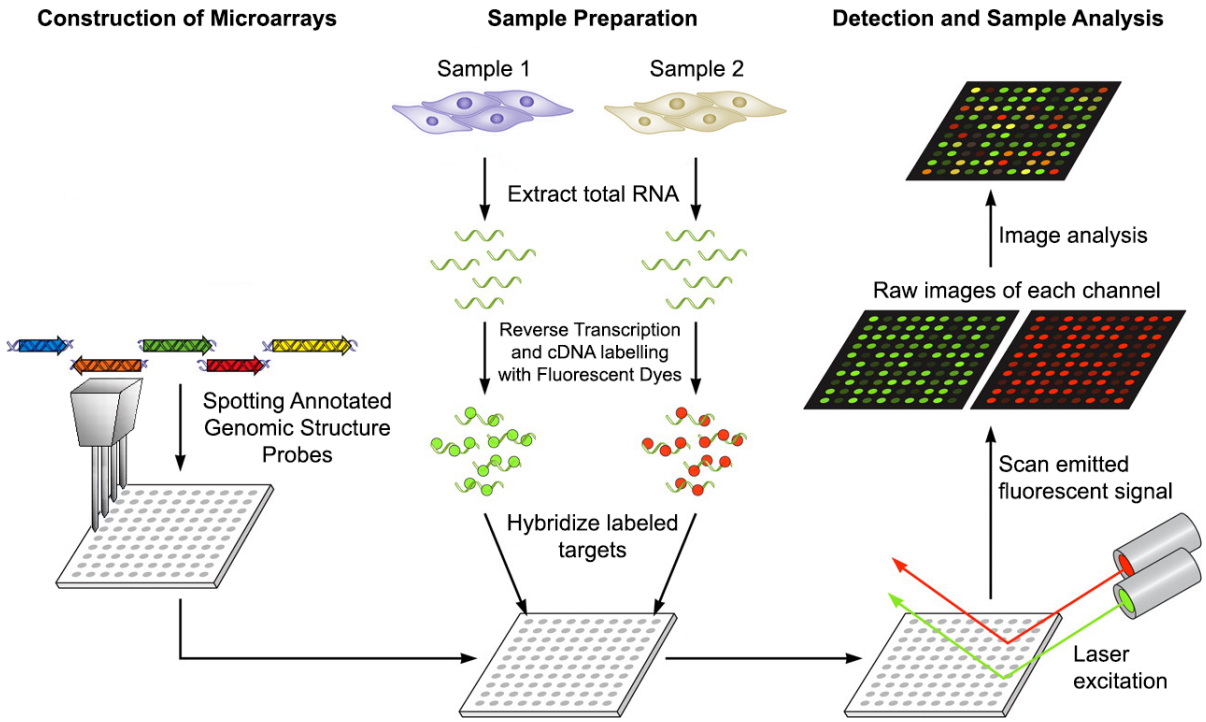


Figure 4.1: Workflow of microarrays analysis. It includes microarrays construction, sample preparation and detection and sample analysis.

tect the genes that the array was designed to detect, thus unknown genes are not detected by microarrays (Bumgarner, 2013).

4.1.2.2 RNA sequencing

RNA sequencing is a high-throughput technology to sequence every RNA molecule and profile the expression of a particular gene by counting the number of times its transcripts have been sequenced; thus, RNA sequencing is a direct measurement of which nucleic acids are present in solution.

A schematic workflow of RNA sequencing analysis is shown in Figure 4.2. RNA is extracted from cells or tissues, then it is retro-transcribed to cDNA and fragmented. The sequence of bases that form each fragments is then identified, and fragments are mapped, i.e. realigned, to the genome, in order to identify which genes are expressed and how many times. In this way it is possible to quantitatively assess the abundance of each gene. This technology can be applied to cell populations, thus the analysis is performed on the averaged gene expression of a multitude of cells; it can also be applied to gene expression of single-cells, allowing to identify of cell-to-cell variability as well as discover cell subpopulations (Duncombe et al., 2015; Halldorsson et al., 2015). Technology for performing single-cell RNA-seq has been partially established (Appendix B on page 105).

Since the bases forming each fragment are read by sequencing, the technology does not require prior knowledge of which nucleic acids may be present and it can also independently and precisely detect

closely related gene sequences, novel splice forms or RNA editing that may be missed due to cross hybridization on microarrays (Bumgarner, 2013).

Considering the intrinsic technological advantages of RNA sequencing over microarrays, RNA sequencing is becoming in the last years the mostly used technique for circadian transcriptome analysis.

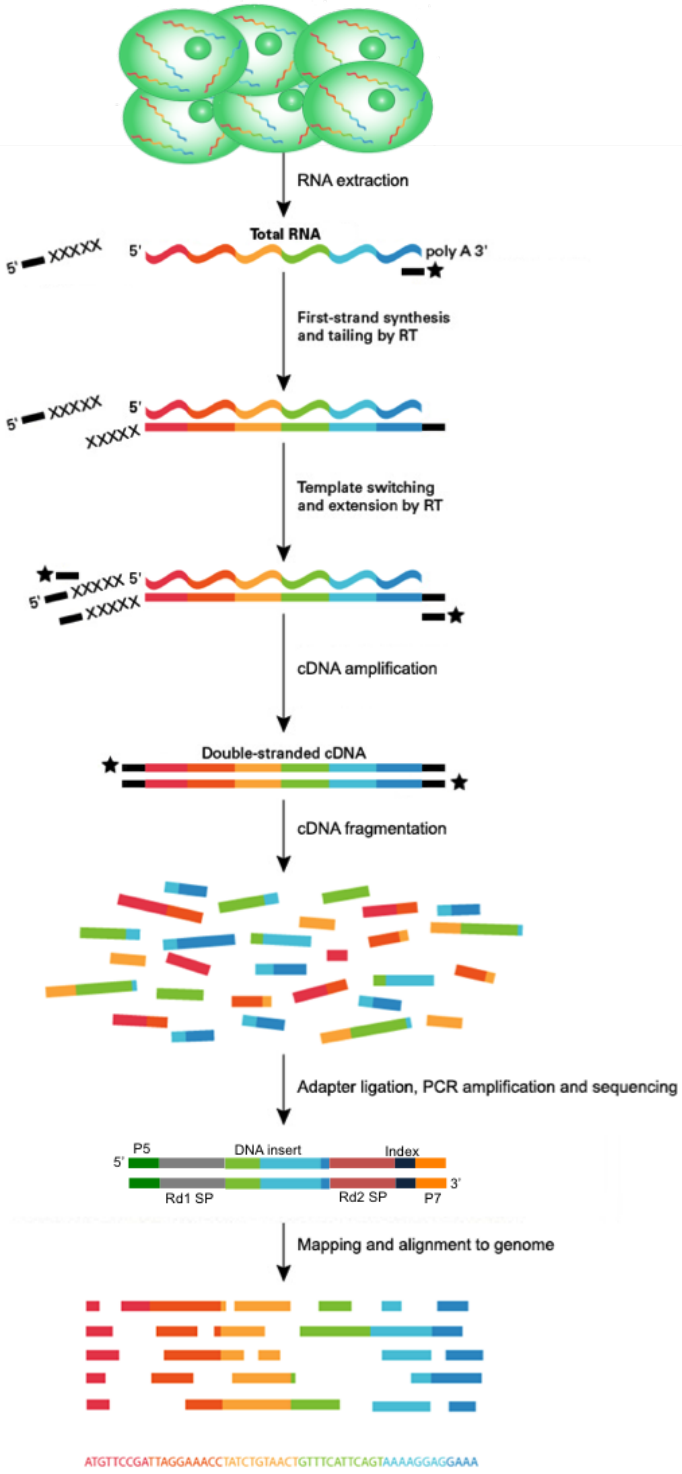


Figure 4.2: Workflow of RNA sequencing analysis. It includes RNA extraction from cells, retro-transcription (RT) into cDNA, cDNA amplification and fragmentation, fragments amplification, sequencing and mapping of fragments to the reference genome.

4.1.3 Algorithms for detection of oscillatory patterns

Besides technical issues in collecting time series data, the challenging matter to be addressed is which genes are rhythmically expressed; this can be classified as a periodicity identification task. Several computational problems arise when dealing with rhythmicity identification in gene expression time series since usually data are short time series with low sampling frequency (i.e. sparse determination) at high levels of noise. Typically, data are collected every 4 h over a course of 24 or 48 h, generating expression profiles with 6 or 12 time points (Appendix C on page 111). Budget constraints and dampening of the circadian rhythm in *in vitro* studies (or desynchronization in *in vivo* studies) are the two factors limiting the number of feasibly obtainable samples. Moreover, budget constraints limit replicate number. It should also be considered that these algorithms are applied to large dataset (transcriptomic studies include 20,000 - 30,000 genes) thus relatively fast automated algorithms are required as well as appropriate statistical tools for addressing multiple comparison problems (Straume, 2004; Wichert et al., 2003; Yang and Su, 2010).

After detection of cycling components in datasets, it is also important to determine crucial attributes of their rhythms including period, phase, and amplitude.

Many statistical tests have been developed to detect cycling components in datasets, including but not limited to COSOPT, Fisher's G-test, F24, Haystack, ARSER, JTK_CYCLE, Lomb Scargle, and Delichtenberg (Li et al., 2015). The methods fall into two major categories: time-domain and frequency-domain analyses. Typical time-domain methods rely on cosine-based pattern, while frequency-domain methods are based on spectral analysis methods (Yang and Su, 2010). In this Section COSOPT, Fisher's G-test, ARSER and JTK_CYCLE will be described since they are the most used ones in published papers of whole transcriptome studies through RNA sequencing technology (Appendix C on page 111).

4.1.3.1 COSOPT

Among time-domain methods, COSOPT, which was published in 2004 by Straume, is a well-known algorithm frequently used to analyze circadian microarray data in *Arabidopsis*, *Drosophila* and mammalian systems (Straume, 2004; Yang and Su, 2010). COSOPT measures the goodness-of-fit between experimental data and a series of cosine curves of varying phases and period lengths. It is highly intuitive and accurately measures period, phase and oscillatory amplitude of rhythmic transcripts in a most conservative manner compared to other methods, minimizing false positives at the expense of false negatives (Straume, 2004). However, because it is permutation-based, it is statistically under powered, it has a large computational cost, it is noise-sensitive and model-dependent, thus not effective at finding periodic signals that are not perfectly sinusoidal (Hughes et al., 2010; Yang and Su, 2010).

COSOPT algorithm comprises the following steps:

1. it calculates the mean expression intensity and its corresponding standard deviation;

2. it calculates linear-regression-detrended data $y_{lr}(t)$ by performing an arithmetic linear-regression detrend of the original time series to obtain stationary oscillations;

3. it calculates mean and standard deviation of the detrended time series;

4. it requires user-specified number and range of periods (test periods are spaced uniformly in period-space);

5. it generates, for each test period, 101 unit-amplitude test cosine basis functions $y_b(t)$, varying from minus one-half the period to plus one-half the period (i.e., such that phase is considered in increments of 1% of each test period);

6. it calculates, for each $y_b(t)$, the least-squares optimized linear correspondence between $y_{lr}(t)$ and $y_b(t)$ as a function of t such that the approximation of the detrended time series data is optimized by $y_b(t)$ for all t values in terms of two parameters, α and β, where $y_{lr}(t) \approx \alpha + \beta y_b(t)$; the quality of the approximation is assessed by the sum of squared residuals (SSR) between $y_{lr}(t)$ and its approximation given by $\alpha + \beta y_b(t)$. SSR values are used to identify the phase at which the optimal correspondence between $y_{lr}(t)$ and $y_b(t)$ is obtained for each test period. Thus, for each test period, the values of α, β and SSR are obtained at the optimal phase. β is the most significant parameters since it represents a parametrized measure of the magnitude of the oscillatory amplitude expressed by $y_{lr}(t)$ as modeled by a cosine curve of corresponding period and optimal phase;

7. it performs empirical resamplings by 1000 Monte Carlo cycles to generate both randomly shuffled temporal sequences and sequences with pseudo-Gaussian-distributed noise in proportion to the corresponding values of experimental uncertainty (estimated by experimental replicates), to take into account the influence of temporal patterning and experimental uncertainty, respectively. Empirical resampling is used to assess statistically the probability that a significant rhythm is present in $y_{lr}(t)$ as modeled by $y_b(t)$ at the optimal phase;

8. it calculates parameters α, β and SSR as in step 6 for all resampled time series;

9. based on a normality assumption, it calculates a one-sided significance probability for each test period and optimal phase based on mean and standard deviation of the resampled β values;

10. it assesses if oscillations are statistically significative.

4.1.3.2 Fisher's G-test

Among frequency-domain methods, Fisher's G-test was proposed by Wichert et al. (2003) to detect periodic gene expression profiles and it has been used to analyze circadian microarray data of *Arabidopsis* and mammalian systems. Fisher's G-test searches periodicity by computing the periodogram of experimental data and tests the significance of the dominant frequency using Fisher's G-statistic; however, it is limited by low frequency resolution for short time-series generated by circadian microarray experiments, which means it is often not adequate to resolve the periodicity of interest (Yang and Su, 2010). It is computationally efficient and more powerful than COSOPT since it is noise-insensitive and model-independent but it is not suitable for characterizing the properties of identified cycling transcripts, as amplitude and phase (Hughes et al., 2010).

Fisher's G-test makes use of the periodogram:

$$I(\omega) = \frac{1}{N}\left|\sum_{t=1}^{N} Y_t \exp(-i\omega t)\right|^2, \qquad \omega \in [0,\pi] \tag{4.1}$$

where N is the sample size, Y_t is the generic component of a time series Y_1, ... Y_N, ω is the frequency of the time series and t is time. Periodogram can be graphically represented as a function of ω (treated as a continuous or discrete variable assuming values equal to $2\pi k/N$, with $k = 0, 1, \ldots, N/2$, i.e. from 0 to π, known as Fourier frequencies). If a time series has a significant sinusoidal component with a frequency ω_0, periodogram will exhibit a peak at that frequency with a high probability. Conversely, the plot of the periodogram is a straight line if the time series is a merely random process.

When G time series data of N elements are available, the average periodogram can be defined as

$$AI(\omega) = \frac{1}{G}\sum_{i=1}^{G} I_i(\omega) \tag{4.2}$$

where $I_i(\omega)$ is the periodogram of the i-th time series. By plotting $AI(\omega_k)$ vs. $\omega_k = 2\pi k/N$, where $k = 0$, 1, ..., $N/2$ it should be possible to understand if the data contains time series with a strong periodicity because the periodogram of these time series will have the largest contribution in Equation (4.2), while non-periodic time series will have negligible contribution to the average periodogram. However, average periodogram yields accurate results only if some time series show identical periodic behaviour.

After assessing that the periodogram of a single time series contains a peak, it should be stated whether this peak is significant or not by the use of g-statistic for the maximum periodogram coordinate:

$$g = \frac{\max_k I(\omega_k)}{\sum_{k=1}^{N/2} I(\omega_k)} \tag{4.3}$$

Large values of g lead to the rejection of the null hypothesis (i.e. purely random process). The exact distribution of g is given by

$$P(g > x) = n(1-x)^{n-1} - \frac{n(n-1)}{2}(1-2x)^{n-1} + \ldots + (-1)^p \frac{n!}{p!(n-p)!}(1-px)^{n-1} \tag{4.4}$$

where $n = N/2$ and p is the largest integer less than $1/x$. Equation (4.4) yields a p-value that allows to test whether a gene behaves like a purely random process or whether it displays some periodic expression pattern.

When large data set are considered, it is important to address the problem of multiple testing, which in Fisher's G-test is done with the method of False Discovery Rate (FDR) as multiple comparison procedure (Wichert et al., 2003).

4.1.3.3 ARSER

ARSER algorithm was published by Yang and Su (2010) and it combines time domain and frequency domain analysis for extracting and characterizing rhythmic expression profiles from temporal microarray data. It estimates the period of rhythmic transcripts from the frequency spectrum by autoregressive (AR) spectral analysis and then it models the rhythmic patterns by using a harmonic regression model to fit the time-series in the time domain, estimating mean, amplitude and phase and statistically validate them (Yang and Su, 2010). ARSER is robust to noise, it quickly and exactly estimates periodicity and gives comprehensive and statistically significant results in analysis of short time-series (Yang and Su, 2010).

The methodology followed by ARSER to detect circadian rhythms in gene expression profiles is as follows:

1. it performs data detrending by ordinary least squares to obtain stationary time series;

2. it smoothes detrended time series to remove pseudo-peaks caused by noise;

3. it determines the period of detrended smoothed time series within the range of circadian period length (20-28 h) by AR spectral analysis, which is a high-resolution spectral analysis technique compared to Fourier transform periodogram. AR spectral estimation applies AR model of order p to fit the time series using

$$x_t = \sum_{i=1}^{p} \alpha_i x_{t-i} + \varepsilon_t \tag{4.5}$$

where x_t is a uniformly sampled time-series at Δ sampling interval, α_i are model parameters, ε_t is white noise and $p = 24/\Delta$. After AR spectral modeling, AR spectral analysis estimates the

spectrum with model parameters using

$$p_x(\omega) = \frac{\sigma_\varepsilon^2}{\left|1 + \sum_{k=1}^{p} \alpha_k \exp(-i\omega k)\right|^2}, \qquad 0 \leq \omega < \pi \tag{4.6}$$

where σ_ε^2 is the variance of white noise. AR spectral density curve calculated with Equation (4.5) will show peaks at each frequency associated with circadian period length in the time series;

4. with the estimated periods, it uses harmonic regression to model cyclic components in the detrended time series estimating its amplitude, phase and mean. This step involves fitting a sinusoidal model to a time series by

$$x_t = \mu + \sum_{i=1}^{n} \beta_i \cos(2\pi f_i t + \phi_i) + e_t \tag{4.7}$$

where μ is the mean value of the time series, β_i is the amplitude of the waveform, f_i are the dominant frequencies derived from Equation (4.6) (note that $f_i = \tau_i/\Delta$, where τ_i is the period estimated in step 3), t is the sampling time, ϕ_i is the phase and e_t are the residuals. The unknown parameters β_i, ϕ_i and μ are estimated with ordinary least squares method;

5. to distinguish between real rhythms and random oscillations, ARSER uses F-test to assess the significance of the regression parameters;

6. FDR is used for addressing the problem of multiple comparisons in large dataset.

4.1.3.4 JTK_CYCLE

JTK_CYCLE is a non-parametric statistical algorithm designed to identify and characterize cycling variables in large datasets which was published by Hughes et al. (2010). It distinguishes between rhythmic and non-rhythmic transcripts more reliably and efficiently than COSOPT and Fisher's G-test since it displays increased resistance to outliers (Hughes et al., 2010). It also accurately measures period, phase, and amplitude of cycling transcripts. Finally, it is several orders of magnitude faster than COSOPT, making it ideal for large scale datasets (Hughes et al., 2010). JTK_CYCLE's advantages over COSOPT and Fisher's G-test are more evident in the analysis of synchronized cell lines than tissue samples from animals, since transcriptional rhythms in cell lines are not reinforced by systemic cues, resulting in generally low amplitude rhythms that are more affected by experimental and biological noise (Hughes et al., 2010).

JTK_CYCLE applies Jonckheere-Terpstra-Kendall (JTK) algorithm to alternative hypothesized group orderings derived from cosine curves corresponding to a range of user-defined period lengths and phases. The algorithm finds the optimal combination of period and phase that minimizes the exact p-value of

Kendall's tau correlation between an experimental time series and each tested cyclical ordering. Each minimal p-value is Bonferroni-adjusted for multiple testing and consequently, the adjusted minimal p-values reported by JTK-CYCLE are uniformly conservative.

More in detail, JTK applies Jonckheere-Terpstra (JT) test to a group of alternative hypothesized group orderings. JT test is a non-parametric test similar to Kruskal-Wallis test and it has more statistical power in monotonically ordered alternatives; it is used to accept or reject the null hypothesis (H_0: time series have the same period and phase) with respect to the alternative hypothesis (H_A: time series are monotonically ordered) based on the distribution of test statistic Z calculated as

$$Z = \frac{U_{xy} - E(U)}{\sqrt{Var(U)}} \tag{4.8}$$

with $E(U)$ and $Var(U)$ defined as

$$E(U) = \frac{N^2 - \sum_{i=1}^{N} n_i^2}{4}, \qquad Var(U) = \frac{N^2(2N+3) - \sum_{i=1}^{N} \left[n_i^2(2n_i+3)\right]}{72} \tag{4.9}$$

where U_{xy} is the total number of observations in group y that are larger than each observation in group x, N is the total sample size and n_i is the sample number in each group.

Kendall's tau is a measure of rank correlation that is used to assess the ordinal association between two measured quantities, and it is defined as

$$\tau = \frac{C - D}{\frac{n(n-1)}{2}} \tag{4.10}$$

where C is the number of concordant pairs, D is the number of discordant pairs and n is the number of observations per group. Any pair of observations (x_i, y_i) and (x_j, y_j), where $i \neq j$, are concordant if both $x_j > x_i$ and $y_j > y_i$ (or $x_j < x_i$ and $y_j < y_i$) while they are discordant if the observations do not have rank correlation.

JTK_CYCLE estimates also the amplitude of each optimal cyclical pattern by calculating the one-cycle median sign-adjusted deviation from the median, where each sign-adjusted deviation equals the product of the deviation and the associated sign of the optimal cosine pattern.

4.1.3.5 Comparison of methods

Table 4.1 summarizes the characteristics of the algorithms for rhythmic transcript identification described in Section 4.1.3.

Table 4.1: Comparison of algorithms for rhythmic and non-rhythmic time series identification and characterization

Algorithm	General description	Domain	Distinction of circadian and non-circadian	Estimation		
				Period	Phase	Amplitude
COSOPT	Cosine fitting	Time	Based on permutations resampling	✓	✓	✓
Fisher's G-test	Fourier Analysis	Frequency	Based on G-statistics	✓	×	×
ARSER	AR spectral estimation and harmonic regression	Frequency and Time	Based on F-statistics	✓	✓	✓
JTK_CYCLE	Non-parametric statistical test	Time	JT algorithm	✓	✓	✓

The algorithms are based on different mathematical assumptions thus they differ in the number of identified rhythmic transcripts from the same dataset, as well as in parameters estimation. Their performance has been compared by Deckard et al. (2013) and Hughes et al. (2010). Overall, it is possible to state that JTK_CYCLE can identify more cycling transcripts than COSOPT and Fisher's G-test, and it can estimate period, phase and amplitude more accurately than COSOPT requiring less computational time (Hughes et al., 2010). Methods relying on comparing experimental data with reference curves, such as COSOPT, ARSER and JTK_CYCLE, perform most accurately when the data match the reference curves; moreover, as noise increases or sampling rate decreases, the ability of these algorithms to classify non-standard curve shapes degrades much more rapidly than for true cosine shapes (Deckard et al., 2013).

Besides the strengths and weaknesses of each method, they are all concerned with identification of the number of rhythmic transcripts over the total number of expressed transcripts in a given tissue or cell line, or with grouping rhythmic transcripts based on their oscillation period. However, they are not able to represent both whole transcriptome and circadian transcriptome in a global way, in order to clearly visualize all transcriptome data at a glance, as well as to clearly visualize changes at the transcriptome level between different biological conditions.

4.2 Motivation and aim

A large number of biological processes are under circadian control, and the number of newly identified CCGs is constantly increasing. The identification of CCGs has a crucial biological importance since it allows to understand which processes are rhythmic and how their expression varies over time. This knowledge, for example, could be used for chronotherapy studies, since it is known that drug intake, metabolism, action and detoxification are regulated in a circadian manner; thus the discovery of new CCGs will have direct implications for improving the quality and the efficacy of medical treatments in a relatively simple way, because it will require only to modify the time of administration of drugs.

However, conventional techniques, i.e. qrtPCR and bioluminescence reporter, are limited in the analysis of only one of few genes at a time, giving very detailed descriptions of specific mechanisms or biological phenomena but loosing the description of changes at a more comprehensive level. This approach also limits the discovery of new regulatory pathways or unknown interactions between genes, since the experimental investigation with these techniques is restricted to specific targets, which are often identified based on *a priori* knowledge of the gene interaction network. Therefore, technologies such as RNA sequencing are very helpful in unveiling gene interactions and regulations at the omics level, specially in the case of rhythmic gene expressions.

Parallel to technological improvements, also proper data analysis tools are required to treat and visualize transcriptomic datasets at a global level, besides identifying and characterizing the properties of rhythmic transcripts. Currently available methods are powerful tools in discriminating between rhythmic and non-rhythmic transcripts and in characterizing the properties of their oscillation; however, they do not allow to represent whole transcripts in a clear and comprehensive graphical representation.

Thus the aim of this part of the research activity is to define a method for globally visualizing RNA sequencing data measured over time, for distinguishing between rhythmic and non-rhythmic transcripts and for discovering new regulatory network or new interactions between genes, by looking at the phase of the expression and at the differences at the transcriptional level between different biological conditions.

4.3 Latent Variable Models Approach

This Section provides a general overview of Latent Variable Models, a description of Principal Component Analysis, focusing both on the algorithms and on the interpretation of the results. It describes also two important aspects connected to Principal Component Analysis, which are data pre-treatment and selection of principal components.

4.3.1 Latent Variable Models

Latent Variable Models (LVMs) are statistical models that have been developed to analyze large amounts of usually correlated data. The underlying idea of LVMs is that real data can be expressed as a linear combination of factors, namely Latent Variables (LVs), that describe the dominant trend of the data and that can be interpreted based on the knowledge of the physical and chemical phenomena involved in the system. LVs represent the driving forces acting on the system and responsible for the variability of the data. Thus, LVMs are used for data compression as well as for data interpretation, assuming that the information can be extracted by analysing how one LV changes with respect to the others.

The objective of LVMs is to explain the correlation structure of the variables to understand the relations among them. Principal Component Analysis (PCA) is one of the most useful techniques to this purpose. PCA is the proper method to deal with both correlated and dynamic data (Camacho et al., 2010; Nomikos and MacGregor, 1994), which is the case of gene expression data measured over time. The theoretical background described in this Dissertation has been taken from Meneghetti (2016), Ottavian (2014) and Tomba (2013).

4.3.2 Principal Component Analysis

Principal component analysis is a multivariate statistical method that performs dimensionality reduction while maintaining the information embedded in a dataset $\mathbf{X}$ $[I \times N]$ of I samples and N (correlated) variables, by projecting the data through a linear transformation onto a new coordinate system of orthogonal, i.e. independent, principal components (PCs); PCs identify the direction of maximum variability of the original data (Jackson, 1991). The original dataset $\mathbf{X}$ can be represented by PCA as the sum of the R scores-loadings vectors outer products

$$\mathbf{X} = \sum_{a=1}^{R} \mathbf{t}_a \mathbf{p}_a^{\mathrm{T}} \tag{4.11}$$

where R is the rank of matrix $\mathbf{X}$, $\mathbf{t}$ are the score vectors and $\mathbf{p}$ are the loading vectors for each component a. Score vectors contain information on how samples are related, and loading vector contains information on how variables are correlated.

The determination of scores and loadings of the decomposition is done by solving the optimization problem (Burnham and Viveros, 1996)

$$\begin{aligned} \max_{\mathbf{p}_a} \quad & \left(\mathbf{p}_a^{\mathrm{T}} \mathbf{X}^{\mathrm{T}} \mathbf{X} \mathbf{p}_a\right) \\ \text{subject to} \quad & \mathbf{p}_a^{\mathrm{T}} \mathbf{p}_a = 1 \end{aligned} \tag{4.12}$$

since $\mathbf{p}_a$ represents the latent direction of maximum variance in the data on which the original data can

be projected, by obtaining the vector

$$\mathbf{t}_a = \mathbf{X}\mathbf{p}_a \tag{4.13}$$

that represents the coordinates of the data in the PC space along the a latent variable. A iterations are performed to identify model parameters for the A selected PCs.

The analytical solution of the maximization problem in Equation (4.12) is obtained by the following eigenvalue problem (López-Negrete de la Fuente et al., 2010)

$$\text{cov}(\mathbf{X})\mathbf{p}_a = \mathbf{X}^{\text{T}}\mathbf{X}\mathbf{p}_a = \lambda_a \mathbf{p}_a \tag{4.14}$$

where it is clear that $\mathbf{p}_a$ is the eigenvector corresponding to eigenvalue λ_a of the covariance matrix of $\mathbf{X}$. Thus, the maximization problem in Equation (4.12) can be explained geometrically as the maximization of the variance captured by λ, which represents the variance explained by the product $\mathbf{t}\mathbf{p}^{\text{T}}$. The eigenvector problem in Equation (4.14) can be used to determine A loadings of the PCA model, which are the A orthonormal eigenvectors of the covariance matrix of $\mathbf{X}$. As a consequence, the resulting score vectors are orthogonal and they have length equal to the eigenvalue λ associated to a-th PC. The computation of PCA model parameters, i.e. scores and loadings, can be done with the Singular Value Decomposition (SVD) of the correlation matrix $\left(\mathbf{X}^{\text{T}}\mathbf{X}\right)/(I-1)$ (where $\mathbf{X}$ and $\mathbf{X}^{\text{T}}$ are autoscaled, Section 4.4.1) or non-linear iterative partial least-squares algorithm (NIPALS) (Cordella, 2012).

As a result of the maximization problem in Equation (4.12), the PCs are ordered according to the variance of the original dataset that they capture, being the first PC the one describing the largest possible variance of the data (this is obtained by iterating the maximization problem in Equation (4.12), performed with dataset $\mathbf{X}$ at the first iteration and with the error matrix $\mathbf{E}$, obtained properly deflating $\mathbf{X}$, for all subsequent steps). Usually A PC are sufficient to adequately describe $\mathbf{X}$, where $A \ll N$, because correlated variables identify a common direction of variability that is captured by a single PC. Thus, assuming that only the first A PCs are used to represent $\mathbf{X}$, equation (4.11) can be written as

$$\mathbf{X} = \sum_{a=1}^{A} \mathbf{t}_a \mathbf{p}_a^{\text{T}} + \sum_{a=A+1}^{N} \mathbf{t}_a \mathbf{p}_a^{\text{T}} = \mathbf{T}\mathbf{P}^{\text{T}} + \mathbf{E} = \hat{\mathbf{X}} + \mathbf{E} \tag{4.15}$$

where $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_A]$ is the score matrix, $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_A]$ is the loading matrix and $\mathbf{E}$ is the error matrix $[I \times N]$ matrix of the residuals generated by $(N-A)$ discarded PCs of the PCA model when $\mathbf{X}$ is approximated by the first A PCs since $\hat{\mathbf{X}} = \mathbf{T}\mathbf{P}^{\text{T}}$. Graphically, the decomposition of the data matrix $\mathbf{X}$ into score, loading and error matrices is shown in Figure 4.3.

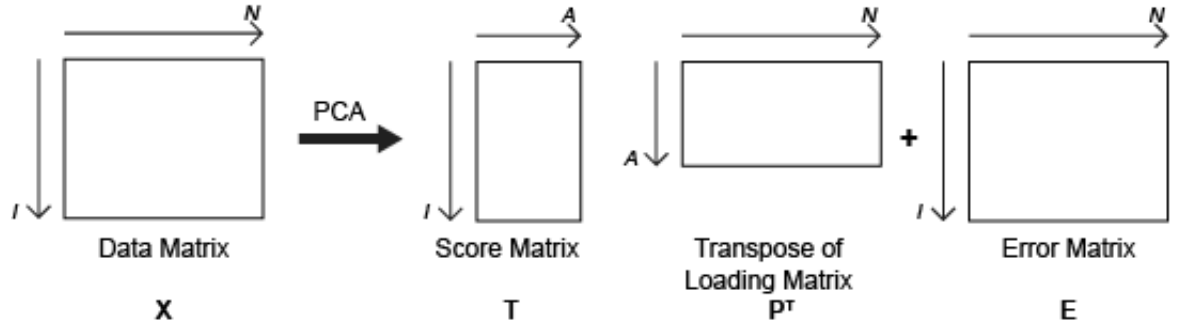


Figure 4.3: Graphical representation of the linear decomposition of the data matrix **X** by PCA into the score matrix **T**, the loading matrix **P** and the error matrix **E**.

In general, when models are used, data are separated into two parts: the systematic part explained by the model and the noise (Eriksson et al., 2006). In the case of a PCA model, if the correct number of PCs is selected, $\hat{\mathbf{X}}$ should include all systematic part of the data, whereas the noise and eventually the remaining unmodeled part of the data is included in **E**.

A graphical representation of geometrical interpretation of PCA model is given in Figure 4.4. A dataset of 9 samples and 2 variables is considered. By applying a PCA model, the direction of maximum variability of the data is identified by PC1, which represents the trend of the data in the space of the original variables. The original dataset can be projected to the new latent space of the PCs, which is determined by a lower number of variables compared to the original one.

The model loadings $p_{1,1}$ and $p_{1,2}$ represent the director cosines of $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively, on PC1, which are the cosines of the angles between the latent direction of the model and the axes of the original variable space. Each score $t_{1,n}$, with 9, represents the coordinate of the n-th sample of matrix **X** in the new model space, represented by PC1. The residual $e_{1,1}$ represents the information not captured by the model based only on PC1 (graphically it corresponds to the distance of sample 1 from PC1 direction). However, if also PC2 was considered, it would account for the orthogonal distance of each projection from PC1 direction; but PC2 would capture a very limited variability of the data compared to PC1. In this case, only PC1 is sufficient to adequately describe **X**.

Besides being used to model a given dataset **X**, PCA can be used to predict or compare external dataset using the information achieved by modeling the calibration set (**X**). The new dataset $\mathbf{X}^{\mathbf{v}}$ (namely validation set) can be projected into the score plane of the calibration set by

$$\hat{\mathbf{T}}^{\mathbf{v}} = \frac{\mathbf{X}^{\mathbf{v}}\mathbf{P}}{\mathbf{P}^{\mathrm{T}}\mathbf{P}} \tag{4.16}$$

where $\hat{\mathbf{T}}^{\mathbf{v}}$ is the $[I \times A]$ matrix of the scores of the validation set calculated from the model parameters of the calibration set, $\mathbf{X}^{\mathbf{v}}$ is the $[I \times N]$ validation dataset and **P** is the $[N \times A]$ loading matrix of the model.

Thus, PCA model can be used to:

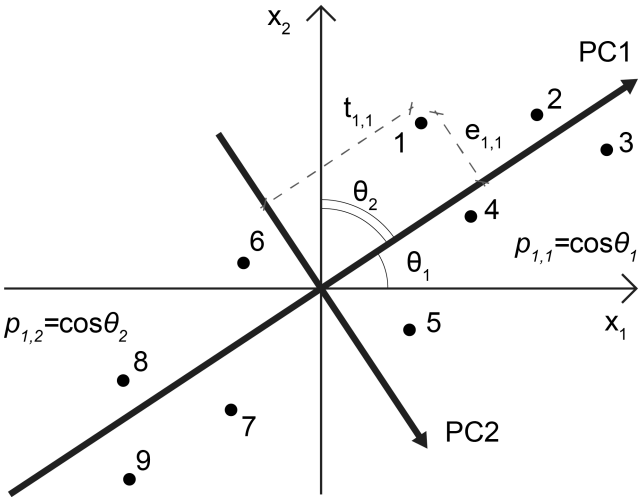


Figure 4.4: Geometrical interpretation of the PCA scores and loadings for a dataset **X** [9×2]

- represent a multivariate system with a reduced number of latent variables compared to the number of the original variables, i.e. data reduction and simplification, extracting the most important information from the dataset
- understand the relation between samples by interpretation of the score plot
- understand the correlation among variables by interpretation of the loading plot
- understand which are the variables that correlate some samples by the combined interpretation of the loading and score plots
- detect outliers
- perform prediction for new data $\mathbf{X}^{\mathbf{v}}$ after building a PCA model based on dataset **X**.

4.3.3 Data pre-treatment

Before modeling a dataset by PCA, dataset are usually pre-treated; the appropriate data pre-treatment depends on the characteristics of the data and on the objective of the analysis and it may include filtering, denoising, transformations, advanced scaling and data compression (Eriksson et al., 2006).

Dataset may collect many variables of different type and physical meaning and to correctly model them by PCA they should be weighted in a similar way. The most common data pre-treatment is autoscaling, i.e. mean-centering the data and scaling them to unit variance. Mean-centering, i.e. subtracting to each column of **X** its mean value, avoids to detect differences among mean values of different variables as significant directions of variability. Scaling to unit variance, i.e. dividing each column of **X** by its standard deviation, makes the analysis independent of the measurement units.

So, autoscaling is applied to reduce magnitude heterogeneity among variables; when the same pre-treatment is applied in order to solve heterogeneity among samples, the pre-treatment is called standard normal variate (SNV), i.e. subtracting to each row of $\mathbf{X}$ its mean value and dividing each row of $\mathbf{X}$ by its standard deviation.

4.3.4 Selection of the number of PCs

As stated before in Section 4.3.2, one of the application of PCA is to perform dimensionality reduction, which means representing the original data by a smaller number of independent latent variables. The selection of the dimensionality of the latent space of the model, namely the selection of the number of PCs to be retained, is a critical aspect in the development of a PCA model, since it may affect its capability in representing in a meaningful way the original dataset. Besides being used to model a given dataset $\mathbf{X}$, PCA can be used to predict or compare external dataset using the information achieved by modeling the $\mathbf{X}$ dataset (called calibration set). Therefore, the number of selected PCs may also limit the ability of prediction of the model, i.e. how accurately an external dataset can be predicted based on the model created based on $\mathbf{X}$ dataset.

In the selection procedure of the appropriate number of PCs, several aspects should be considered, such as the number of samples, the total explained variance, the relative size of the eigenvalues and the subject-matter interpretations of the PCs (Johnson and Wichern, 2007).

Among the available methods, the scree test and the eigenvalue-greater-than-one rule will be used and described (Jackson, 1991; Mardia et al., 1979).

The scree test is an empirical and graphical procedure based on the analysis of the profile of an index (explained variance) indicating the variability of the original data captured by the PCA model per PC as a function of PC number. The explained variance R^2 quantifies the variability of the original data captured by the model and is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{I}\sum_{n=1}^{N}(x_{i,n} - \hat{x}_{i,n})^2}{\sum_{i=1}^{I}\sum_{n=1}^{N}(x_{i,n})^2} \tag{4.17}$$

where $x_{i,n}$ and $\hat{x}_{i,n}$ represent respectively the element in the i-th row and n-th column of the original matrix $\mathbf{X}$ and of the modeled matrix $\hat{\mathbf{X}}$. The method is based on the idea that above a certain number of PCs, additional PCs will describe the variability associated to random errors, thus they are not needed to describe the variability of the systematic part of the data. This number can be graphically identified in the scree plot in correspondence of the elbow, or right before where the curve stabilizes.

The eigenvalue-greater-than-one rule is a simple rule for which all the PCs whose corresponding eigenvalues are lower than one are not considered in the model. The basic idea behind this method is that, if data are auto-scaled, the eigenvalue corresponding to a PC represents roughly the number of

original variables whose variability is captured by the PC itself. If so, a PC capturing less than one original variable should not be included in the model. However, when an eigenvalue is just slightly smaller than 1, it may be convenient to include it in the model since it may contribute significantly to explain the systematic variability of the dataset.

4.3.5 Data unfolding

Data unfolding methods have been developed to include in PCA modeling data variability among different replicates or over time; these two types of data are available from batch processes, in which the same process is repeated *k* times theoretically in the same way, and continuous processes, where the variables describe continuous trajectories over time, respectively. Different replicates or time measurements add a third dimensions to dataset, transforming it into a three-ways array. Unfolding methods transform three-ways arrays into large two-ways arrays, that are then pre-treated prior to perform PCA. Schematic representation of the three unfolding methods is shown in Figure 4.5.

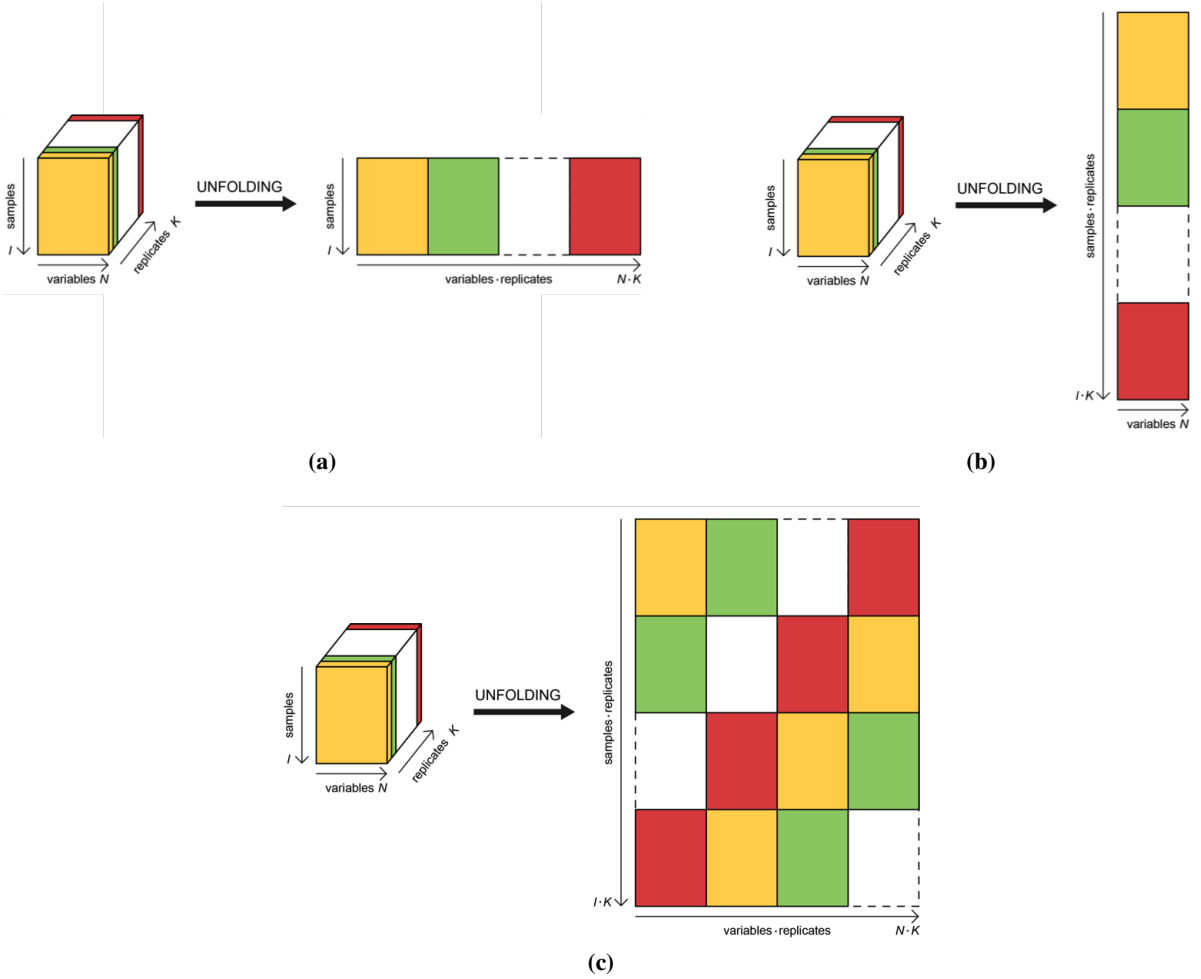


Figure 4.5: Schematic representations of the three unfolding methods.
(a) Batch-wise unfolding. **(b)** Continuous-wise unfolding. **(c)** Hybrid unfolding.

Two unfolding methods have been developed in the literature:

- batch-wise unfolding, by which the different replicates are concatenated along the horizontal direction, accounting for correlations within the same replicate and between replicates;
- continuous-wise unfolding, by which measurements over time are concatenated along the vertical direction and take into account only the "average" correlation of variables in time.

A third unfolding method, namely hybrid unfolding, has been recently published by Facco et al. (2017), and it combines both above mentioned unfolding methods. By this method, firstly a three-ways array matrix is batch-wise unfolded, allowing to account for the correlation of the variables within a single replicate and between different replicates. Then, another rearrangement is performed by permuting the order of the replicates along the horizontal direction, and concatenating them vertically. This second rearrangement accounts for variance and cross-covariance within replicates and the relationship between variables averaged on all replicates (Facco et al., 2017). In this Dissertation hybrid unfolding method has been used in order to include in PCA model all above mentioned correlations and relationships.

4.4 Liver transcriptome analysis

In this Section the results obtained by PCA of whole transcriptome data will be presented. In particular, the structure of the data, their pre-treatment and unfolding will be described, then the characterization of the method will be presented. Finally, applications of the method will be shown.

4.4.1 Structure of the data and data pre-treatment

PCA has been applied for the investigation of whole transcriptome data over time obtained with RNA sequencing. In particular, it has been decided to analyse whole gene expression of the liver, since it has been shown that circadian clock in the liver plays an important role in regulating metabolism and energy homeostasis; circadian control of hepatic metabolism is exerted by controlling expression and/or activity of key metabolic enzymes, transcription factors, signaling molecules, and transporters. Reciprocally, intracellular metabolites modulate the molecular clock activity in response to the energy status, thus defining an interconnected regulation between circadian clock and metabolism. Moreover, circadian clock dysregulation has been implicated in common chronic liver diseases (Tong and Yin, 2013). Thus, it is relevant from a biological point of view to investigate gene expression, and specially circadian gene expression, in the liver.

Murine hepatic gene expression data over time have been taken from Yang et al. (2016). Data are deposited in Gene Expression Omnibus (GEO), a publicly available data repository, under accession number GSE70497. Entrainment and sampling was performed as shown in Figure 4.6. Briefly, mice

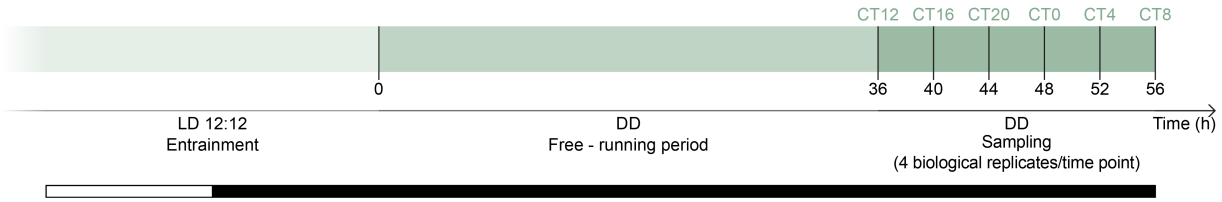


Figure 4.6: Timeline of entrainment, free-running period and sampling as described by Yang et al. (2016). CT indicates Circadian Time (h), CT12 corresponds to the onset of activity, CT0 corresponds to onset of resting phase (see note 1 on page 20) White/black bars indicate light/dark periods.

were maintained under LD12:12, then released into dark/dark (DD) condition and, starting 36 hour after release, sampling was started and performed in constant darkness. Liver samples were taken every 4 hour for 24 hour, i.e. 6 time points, taking four samples per time point.

Two murine genotypes are considered by Yang et al. (2016): wild type mouse (WT) and inducible knock-out mouse (iKO). The former is a healthy organism with a phenotype as it occurs in nature, the latter is a biological model in which the target gene of interest (in this case *Bmal1*) can be specifically inactivated at a given time point. Before induction, the gene of interest is functionally expressed. *Bmal1* knock-out (KO) mice have become a valuable resource for studying the effects of circadian disruption since among canonical clock genes, *Bmal1* is the only one whose deletion results in complete loss of circadian rhythms (it is the sole non-redundant gene in the core molecular clock) (Yang et al., 2016). For both genotypes, the same samples are taken, i.e. 6 samples every 4 hour for 24 hour with 4 biological replicates per time point. In the first part of the Dissertation, gene expression of WT will be analysed, while from Section 4.4.3.1 iKO data will be discussed.

WT gene expression data are arranged in four $[23733 \times 6]$ data matrix $\mathbf{X}_k$, where $k = 1, \ldots, 4$ are replicates, genes (samples) are indicated in the rows and time point (variables) are along the columns. Each data matrix $\mathbf{X}_k$ is formed by randomly selecting one of the four replicates measured at each time point. The analysis does not depend on the replicate combination in the data matrix $\mathbf{X}_k$.

In the paper by Yang et al. (2016), transcripts are divided into circadian and non-circadian ones by the use of JTK_CYCLE (described in Section 4.1.3.4); by this algorithm, 5457 genes are identified as circadian among all expressed genes in the mouse liver (23733 genes are identified in total).

As far as data pre-treatment is concerned, in the case of dataset containing values of gene expression at different time points, variables are homogeneous and they have very similar magnitudes, since the total and average number of transcripts expressed by a tissue at each time point are constant. However, different genes are not expressed with the same abundance, thus in gene expression dataset is more biologically relevant to weight in the same way the expression of the genes. Thus, SNV has been applied, i.e. the mean of the expression of each gene over time has been subtracted to the value of expression of

the same gene at each time point, and this difference has been divided by the standard deviation of the expression of that gene over time

$$x_{i,j}^{SNV} = \frac{x_{i,j} - \frac{\sum_{j=1}^{N} x_{i,j}}{N}}{\sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(x_{i,j} - \frac{\sum_{j=1}^{N} x_{i,j}}{N}\right)^2}}, \quad i = 1, \ldots, I \tag{4.18}$$

where $x_{i,j}$ is the gene expression value in the i-th row and j-th column in dataset $\mathbf{X}$ and $x_{i,j}^{SNV}$ is the SNV of $x_{i,j}$.

Another pre-treatment that has been used is normalization from -1 to 1 of the expression of each gene over time

$$x_{i,j}^{N} = 2\frac{x_{i,j} - \min_{j=1}^{N}(x_{i,j})}{\max_{j=1}^{N}(x_{i,j}) - \min_{j=1}^{N}(x_{i,j})} - 1, \quad i = 1, \ldots, I \tag{4.19}$$

where $x_{i,j}^{N}$ is the normalization of $x_{i,j}$.

By these two pre-treatments, differences in amplitude of expression of the different genes are not considered, while emphasis is given only to the shape of the temporal expression profile, i.e. period and phase.

PCA model has been performed both on $\mathbf{X}^{SNV}$ and $\mathbf{X}^{N}$, and rows containing only 0, i.e. either non-detected gene or non-expressed gene, have been removed prior to perform PCA. PCA has been performed in MATLAB and model parameters have been computed with SVD algorithm.

4.4.2 Characterization of the method

In this Section the characterization of transcriptome analysis by PCA modeling will be presented. Firstly, results about PCA modeling on a single replicate dataset will be presented, analysing separately circadian genes, non-circadian genes and whole transcriptome (according to the classification done by Yang et al. (2016)). Then, the results of PCA modeling of a multiple replicates dataset will be presented.

4.4.2.1 Circadian genes of a single replicate

Since PCA is able to describe a multivariate dataset by a reduced number of LVs that describe the direction of largest variability in the data, firstly PCA has been performed on circadian genes. The result of model parameters calculated on SNV dataset is shown in Figure 4.7.

In Figure 4.7a the loading plot of the two first PCs is reported. By the loadings it is possible to describe the correlation among variables; in this case it is possible to observe that the time points are

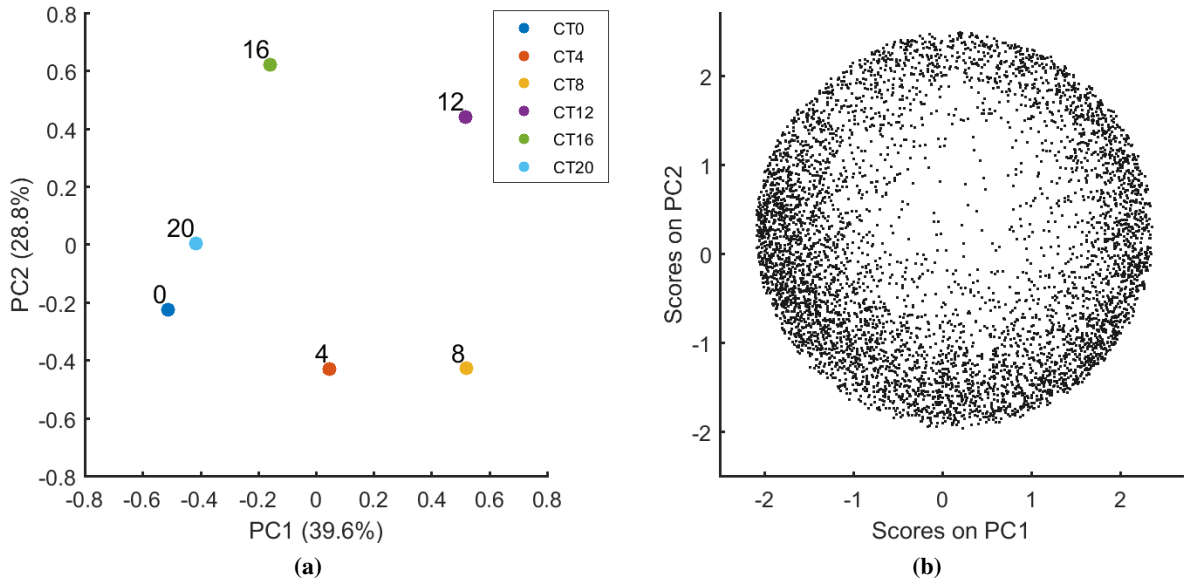
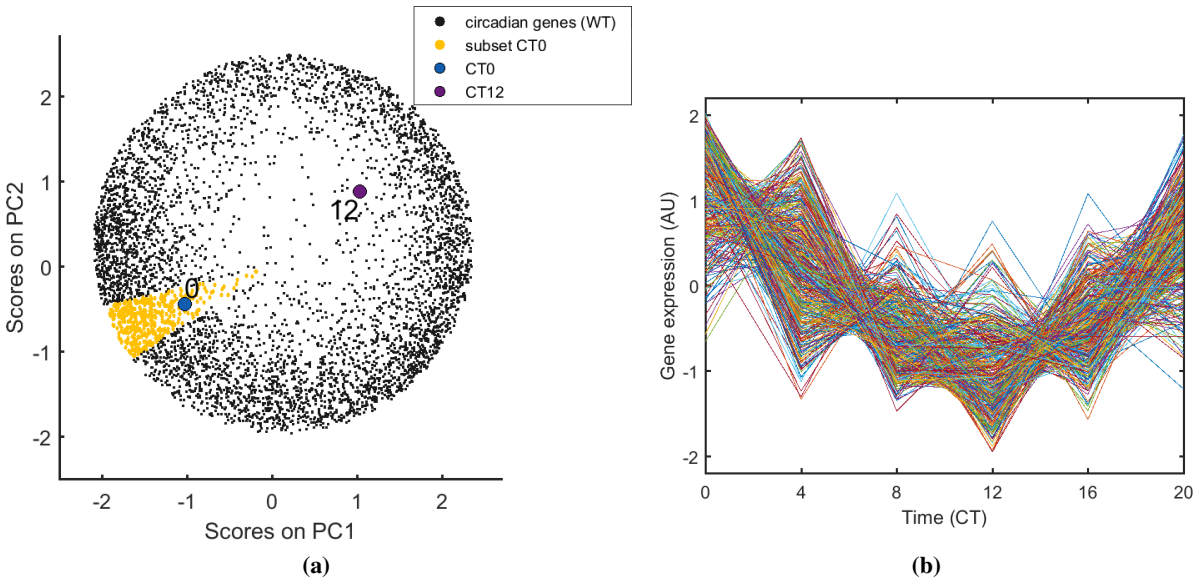


Figure 4.7: PCA model calculated on circadian genes of a single replicate dataset after SNV.
(a) Loading plot. **(b)** Score plot.

located in an approximately circular pattern, and their order along the pattern is in agreement with the temporal sequence of the variables. Moreover, time points that are separated by 12 h interval, i.e. the ones that correspond to anti-phase behaviour, are located in the opposite quadrant with respect to the center of the PCs plane, meaning that only by PC1 and PC2 it is possible to describe the 24 h temporal pattern of circadian genes. It can also be observed that each PC describes a relevant fraction of the variability of the dataset and that the first two PCs account for a comparable variability. The total explained variability by PC1 and PC2 is 68.4%. Based on the eigenvalue-greater-than-one rule, only the first two PCs are considered in the PCA model (scree plot in Figure D.1 on page 119).

In Figure 4.7b the score plot corresponding to PC1 and PC2 is reported; it shows the relation among the samples, i.e. genes. It is possible to observe that the points are mostly located in the border of the circle and that few genes are located in the center of the plot in correspondence to the origin of the axis. Genes that are closely located (i.e. they have similar score values) have a similar temporal profile, meaning that they share the same phase. This is shown in Figure 4.9 and in Figure 4.8.

In Figure 4.8a, a group of 418 circadian genes is highlighted: they are selected from the score plot, defining a sector of amplitude of 20° centered on the position of CT0 (which is derived from the loading plot in Figure 4.7a). Figure 4.8b shows the profiles of the above defined group of genes; it is clear that they all share the same temporal profile, thus they are in phase. Moreover, they all have maximum expression at CT0, and a trough at CT12: this is due to the position of these genes in the score plot compared to the loading plot, meaning that they have high expression values in correspondence to the loading that is located in the corresponding position in the loading plot, and minimum expression values

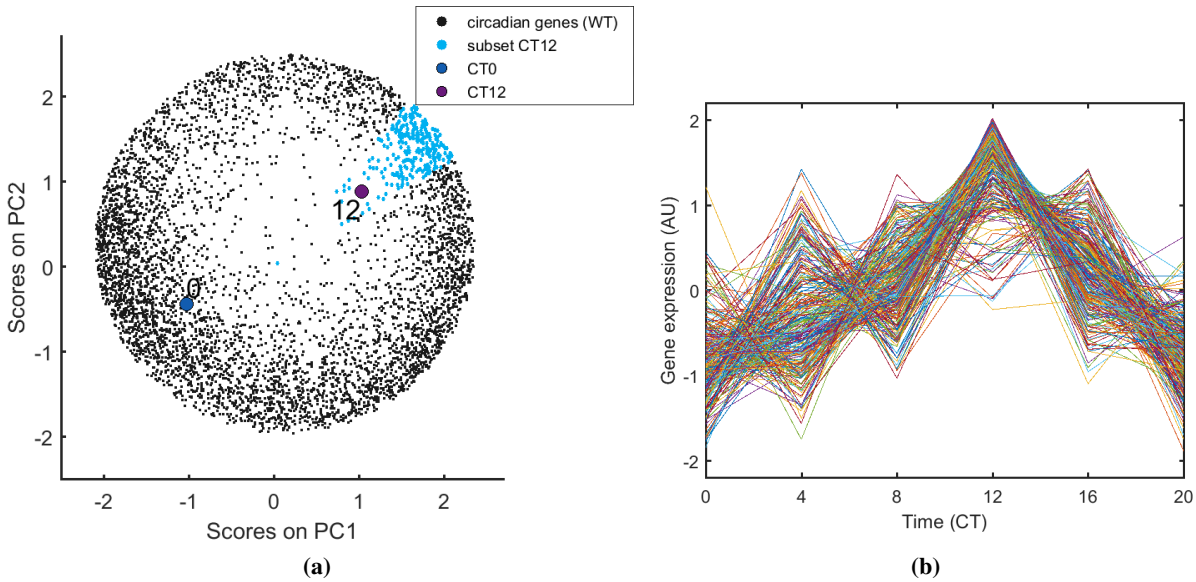


(a) (b)

Figure 4.8: Bi-plot and gene expression profile over time of genes having a peak at CT0.
(a) Bi-plot, with highlighted a group of circadian genes located in correspondence of CT0. **(b)** Gene expression over time of 418 circadian genes highlighted in Figure 4.8a.

in correspondence to the loading that is located in the opposite quadrant in the loading plot.

Similarly, in Figure 4.9, a group of 275 circadian genes is highlighted: they are selected from the score plot, defining a sector of amplitude of 20° centered on the position of CT12 (which is derived from the loading plot in Figure 4.7a). In this case in Figure 4.9b it is possible to observe the genes that are in anti-phase with the genes shown in Figure 4.8b, since these genes display a trough where the previous ones have a maximum.



(a) (b)

Figure 4.9: Bi-plot and gene expression profile over time of genes having a peak at CT12.
(a) Bi-plot, with highlighted a group of circadian genes located in correspondence of CT12. **(b)** Gene expression over time of 275 circadian genes highlighted in Figure 4.9a.

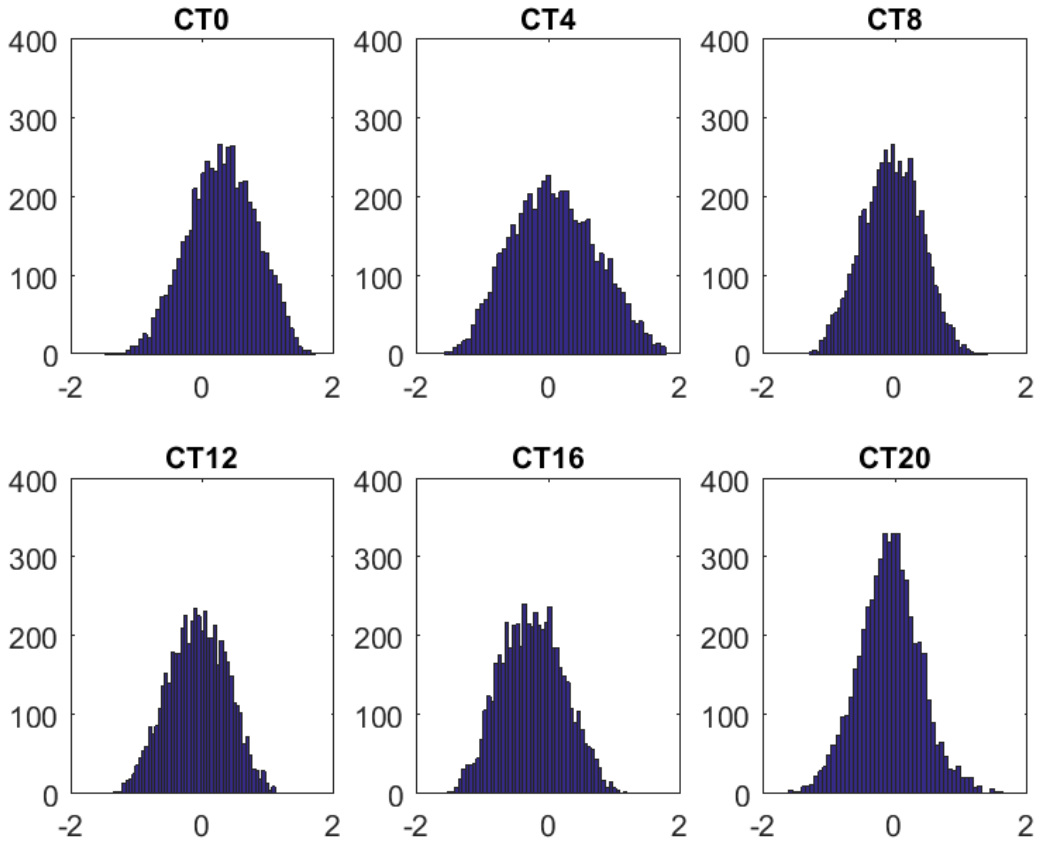


Figure 4.10: Error distributions for circadian genes at the different time points for the first replicate (y-axis indicates the number of genes, x-axis indicates magnitude of the error)

From Figure 4.8b and Figure 4.9b it is possible to observe respectively high and low expression values also at CT20, besides at CT0. This is due to the fact that CT0 and CT20 in the loading plot in Figure 4.7a are closely located, therefore it is not possible to sharply distinguish in this dataset genes that have a different behaviour from CT0 to CT20 (the samples have similar trend in correspondence to these variables).

Since multivariate dataset is modeled by the use of two LVs, the error of the model should be considered, in order to check if errors are normally distributed. The error distributions for circadian genes for the first replicate at the different time points are shown in Figure 4.10. From Figure 4.10 it is possible to observe that, at least for the first replicate, errors are normally distributed around zero for all time points. A normal error distribution is an indication that the unmodeled part of the data, i.e. errors, contains noise and other unmodeled features and not a systematic part of the information stored in the dataset.

It is also important to check the SSR as a measure of the discrepancy between the data and PCA

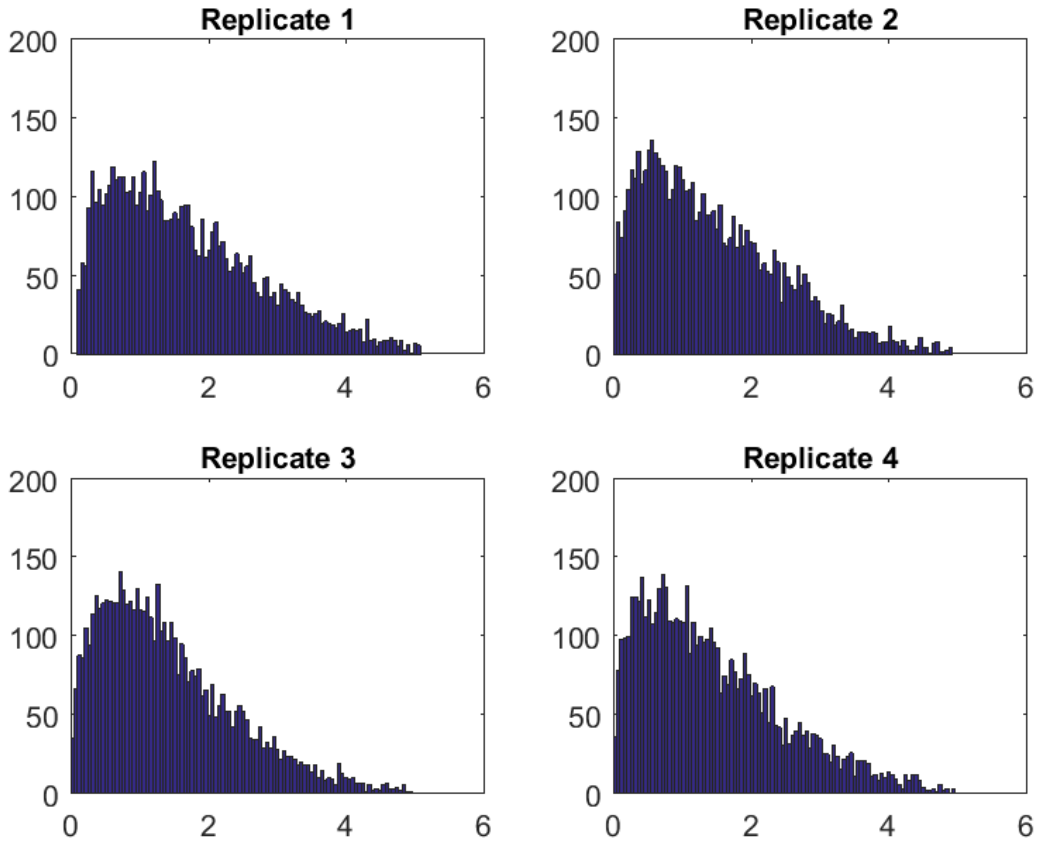


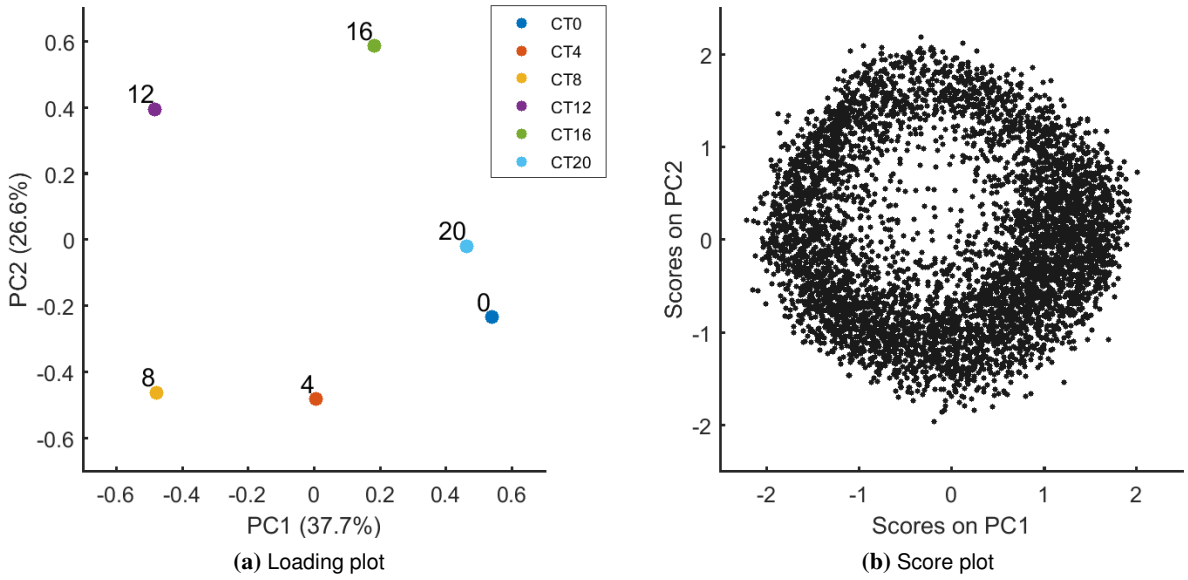
Figure 4.11: Distribution of sum of squared residuals of circadian genes calculated for all replicates based on the first two PCs (y-axis indicates the number of genes, x-axis indicates magnitude of the sum of squared residuals)

model. SSR is calculated for each gene as

$$SSR_{i,1} = \sum_{j=1}^{N} e_{i,j}^2 = \sum_{j=1}^{N} \left(x_{i,j}^{SNV} - \hat{x}_{i,j}^{SNV}\right)^2, \quad i = 1, \ldots, I \tag{4.20}$$

thus SSR summarizes for each sample the error of the model prediction over all variables. In Figure 4.11 SSR distribution is shown for all replicates. For all replicates, a chi-squared distribution can be observed: since this distribution is generated by the sum of independent, standard normal random variables, it can be concluded that all replicates have normal distribution of error for all time points, as for the first replicate as shown in Figure 4.10.

A very similar result in terms of model parameters has been obtained by performing PCA analysis on normalized dataset (Figure 4.12, compared to Figure 4.7). Also error and SSR distributions are comparable (Figure D.2 on page 120 and Figure D.3 on page 120 comparable to Figure 4.10 and Figure 4.11, respectively), meaning that the two pre-treatments lead to a similar model. Thus, this result shows that PCA does not depend on data pre-treatment, as it was expected.



(a) Loading plot **(b)** Score plot

Figure 4.12: PCA model calculated on circadian genes of a single replicate dataset after normalization. **(a)** Loading plot. **(b)** Score plot.

From now on, only results obtained by SNV pre-treatment will be shown since it is a more robust and used pre-treatment method.

In order to assess the experimental variability of the data, circadian genes of the four replicates have been considered. In particular, circadian genes of the other 3 replicates have been projected in the score plane built on the circadian genes of the first replicate, as in Figure 4.7b, by using Equation (4.16), where $\mathbf{X}^{\mathrm{v}}$ is the SNV of either $\mathbf{X}_2$, or $\mathbf{X}_3$, or $\mathbf{X}_4$, and $\mathbf{P}$ is the loading matrix containing only the first two LVs computed for $\mathbf{X}_1$. The result of the projection of all four dataset is shown in Figure 4.13. It is possible to observe that the genes distributions of $\mathbf{X}_2$, $\mathbf{X}_3$ and $\mathbf{X}_4$ are slightly shifted compared to the scores of $\mathbf{X}_1$. This is due to the fact that the means of the different datasets are slightly different, thus when they are projected the center of their distribution might not overlap with the origin of the score plane, which corresponds to the mean of $\mathbf{X}_1$.

From Figure 4.13, it is possible to quantitatively assess data variability by calculating the amplitude of the angle that includes all four replicates of the same gene (each point of this plot represents a gene). If data reproducibility is high, the angle distribution will be narrow and close to 0, while if data reproducibility is low, angle distribution will be broad, thus hindering for example genes grouping based on their phase. The result of distributions of amplitude of angle for circadian genes is shown in Figure 4.14a; it can be observed that the distribution is quite broad. The variability may be due to technical variability as well as to inter-organisms variability. Therefore, in order to reduce the variability among replicates, thus to identify an average behaviour among different organisms, the position of the barycenter of the four replicates of the same gene has been calculated and the replicate of the gene that was located furthest

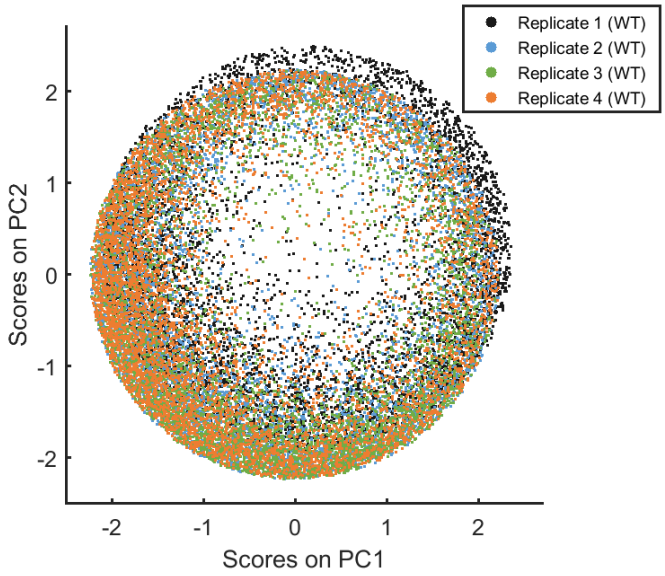


Figure 4.13: Score plot of circadian genes of all replicates (replicate 2, replicate 3 and replicate 4 are projected in the score plane of replicate 1 by the first two PCs)

from the barycenter was discarded. The result of angle distribution for circadian genes, keeping only the three replicates of the same gene that were closer to the barycenter, is shown in Figure 4.14b. It can be seen that the distribution is narrower than in Figure 4.14a and it has a more positive skewness than the one in Figure 4.14a.

Table 4.2 shows the variations of the amplitude of the angle at different confidence intervals (C. I.) and considering or not the relative position of the four replicates with respect to the barycenter. It can be observed that including the barycenter improves the precision in the location of the genes by reducing

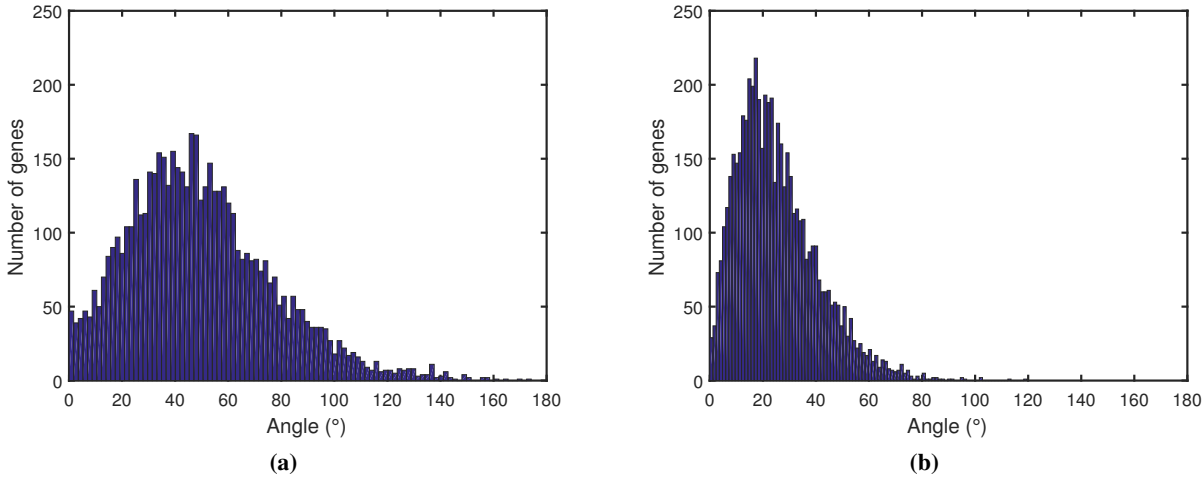


Figure 4.14: Distribution of circadian genes over the amplitude of the angle that replicates of the same genes form on the score plane.
(a) Distribution of number of genes vs. amplitude of angle that spans over all four projections of the same gene. **(b)** Distribution of number of genes vs. amplitude of angle that spans over the three projections of the same gene that are closer to the position of barycenter of the four replicates of the same gene.

Table 4.2: Amplitude of the angle that includes the replicates of the same gene in the score plane at different confidence intervals (C. I.) without and with barycenter correction

	90% C. I.	**95% C. I.**
Without barycenter correction (4 replicates)	90°	102°
With barycenter correction (3 replicates)	70°	82°

the uncertainty of 22%.

From the data it is possible to understand that data variability must be taken into account, therefore a PCA model has to be built on a dataset containing measurements from all datasets. This analysis will be addressed in Section 4.4.2.4.

4.4.2.2 Non-circadian genes of a single replicate

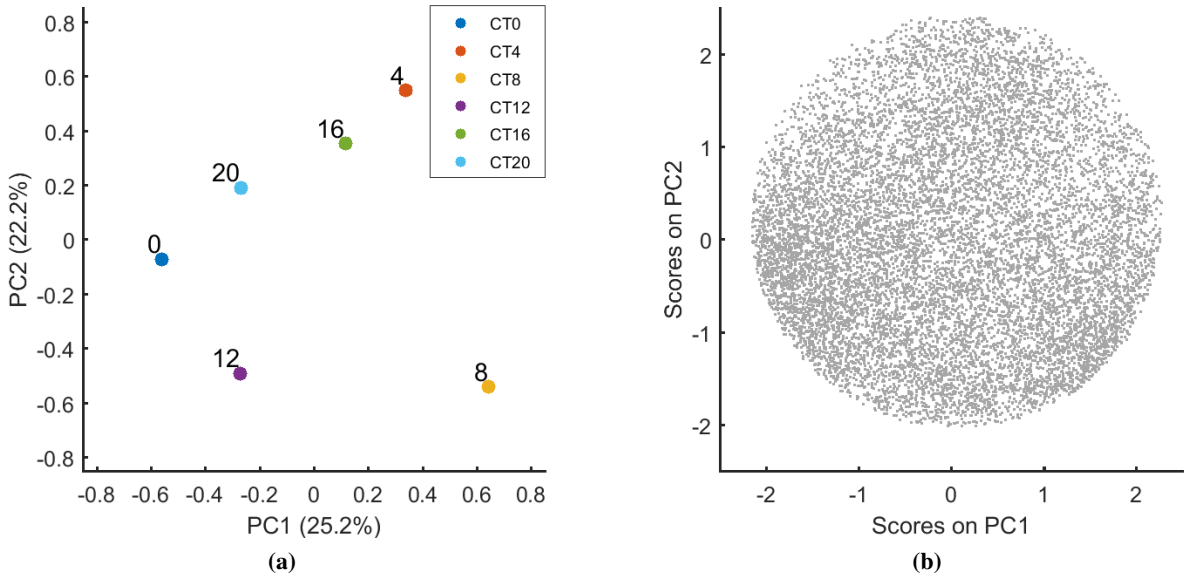
PCA model has also been calculated for non-circadian genes of $\mathbf{X}_1$, as they are identified by Yang et al. (2016). The result of model parameters calculated on SNV dataset is shown in Figure 4.15.

In Figure 4.15a the loading plot of the first two PCs is reported. It is possible to observe that the pattern of non-circadian genes is different from the circadian ones as it has been shown in Figure 4.7a, meaning that the dataset does not contain the same directions of variability. Each one of the two PCs accounts for a smaller percentage of variability compared to the variability accounted for each PC in the PCA model of circadian genes. This indicates that a clear pattern is not identifiable and more random behaviour is evident on how the genes are expressed over time.

In Figure 4.15b the score plot is reported; compared to Figure 4.7b, genes are almost uniformly distributed over all the circle area, occupying both the center and the border of the circle; as it has been shown before, the position of the gene indicates its specific temporal pattern. In particular, the temporal pattern of non-circadian genes located close to the center of the axis, selected as in Figure 4.16a, is shown in Figure 4.16b.

The temporal pattern of the genes located close to the axis origin is uncorrelated with respect to all variables, and the genes display a pseudo-random temporal profile. Therefore, it would be expected that non-circadian genes are located mostly close to the axis origin since they do not have a temporal pattern. However, since non-circadian genes are located in the scatter plot also in the border of the circle, their temporal profiles have been investigated.

In Figure 4.17, the temporal profiles of group of genes located in the border of the score plot in different positions with respect to the loadings are shown. Overall, comparing these temporal profiles with the ones in Figure 4.16b, it is possible to observe that when genes are located far from the axis origin

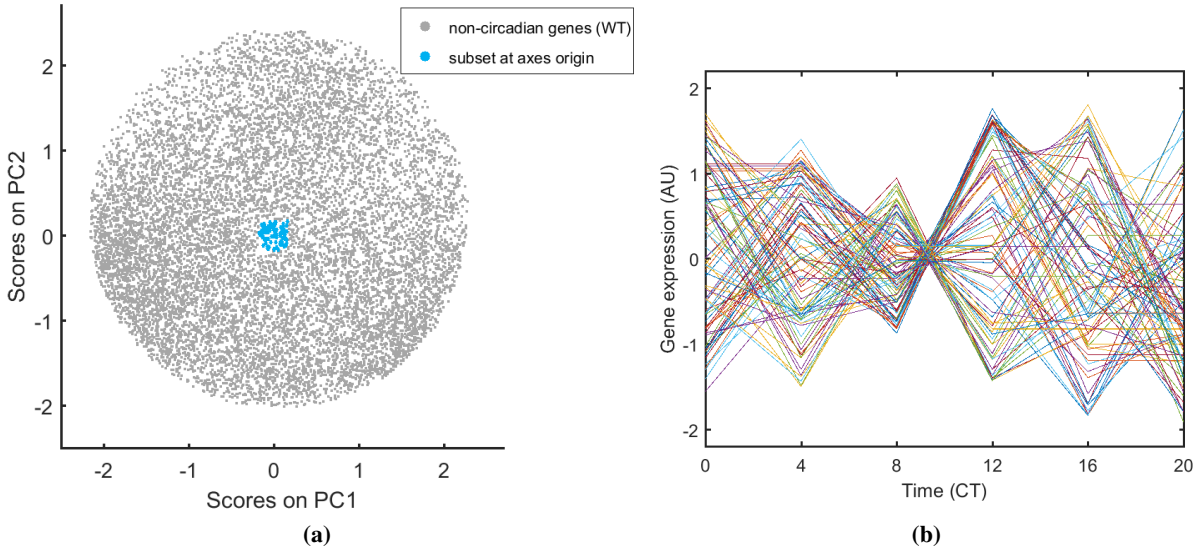


(a) (b)

Figure 4.15: PCA model calculated on non-circadian genes of a single replicate dataset after SNV.
(a) Loading plot. **(b)** Score plot.

they do have a clear temporal pattern. From the temporal profiles it is possible to identify some other temporal patterns, different from circadian ones, that might have for example a shorter period compared to circadian (i.e. ultradian genes).

Therefore, it can be understood that this method can also be used to identify other rhythmic patterns from transcriptomic dataset over time.



(a) (b)

Figure 4.16: Score plot and gene expression profile over time of non-circadian genes located around the axis origin.
(a) Score plot of PCA model calculated on non-circadian genes of one replicate after SNV with highlighted a group of genes located around the axis origin. **(b)** Gene expression over time of 113 genes highlighted in Figure 4.16b.

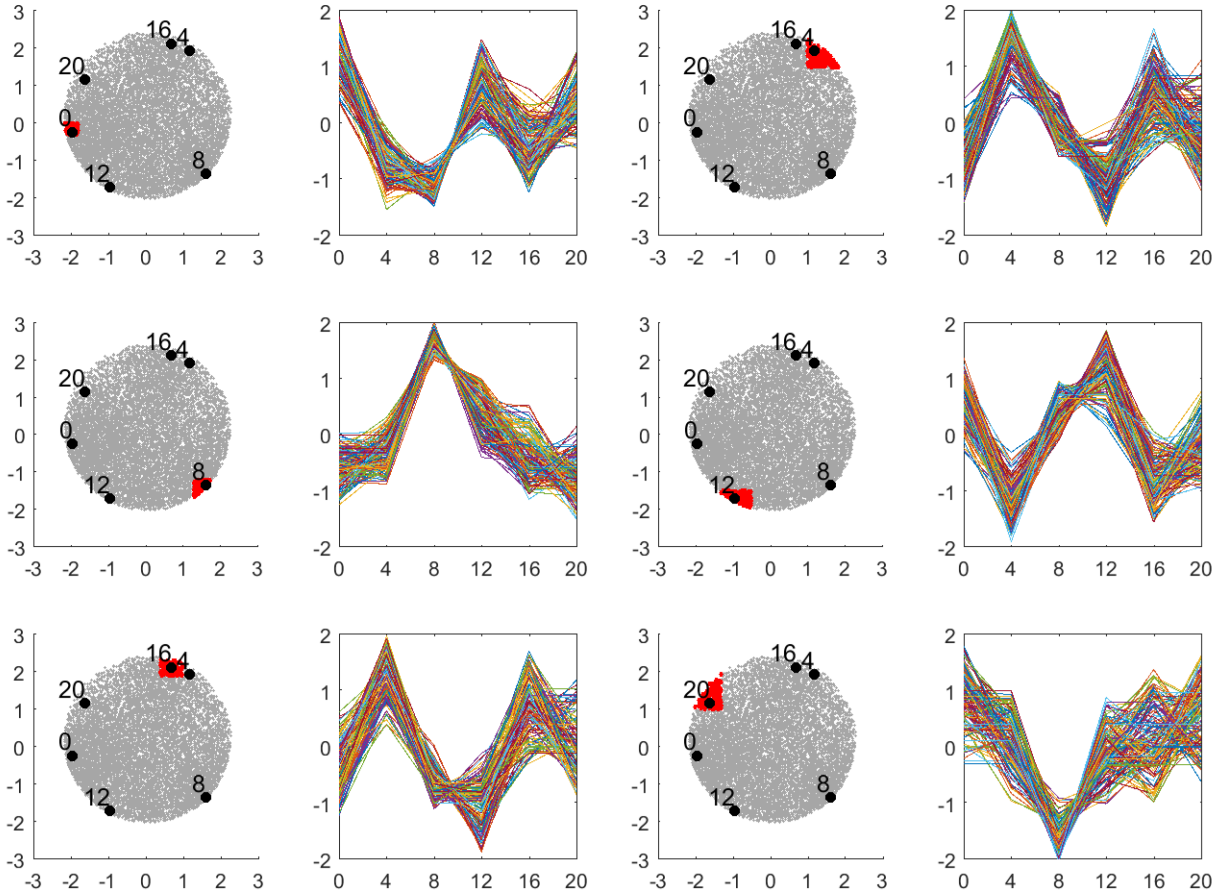


Figure 4.17: Score plots of non-circadian genes from PCA model calculated on one replicate, with highlighted groups of non-circadian genes located in correspondence of specific CT, whose temporal profile is shown in the corresponding temporal profile plot (groups are formed by 136 genes for CT0, 168 genes for CT4, 173 genes for CT8, 157 genes for CT12, 178 genes for CT16, 153 genes for CT20). The position of the loadings has been scaled not-uniformly for graphical representation (score plot: x-axis indicates scores on PC1, y-axis indicates scores on PC2; temporal profile plot: x-axis indicates time (CT), y-axis indicates gene expression [AU]).

4.4.2.3 Whole transcriptome of a single replicate

After characterizing PCA model for circadian genes and non-circadian genes separately, the model has been applied to study whole transcriptome dataset, concatenating vertically the two datasets. The result of model parameters calculated on SNV dataset is shown in Figure 4.18.

In Figure 4.18a the loading plot of the first two PCs is reported. As in the case of PCA model on circadian genes shown in Figure 4.7a, variables are located approximately in a circular pattern and they are ordered according to the temporal sequence of the time points. However, with respect to the axis origin, they are less uniformly separated (meaning that the angle by which two subsequent time points are separated is not constant over all pairs of subsequent time points). This makes the grouping of genes

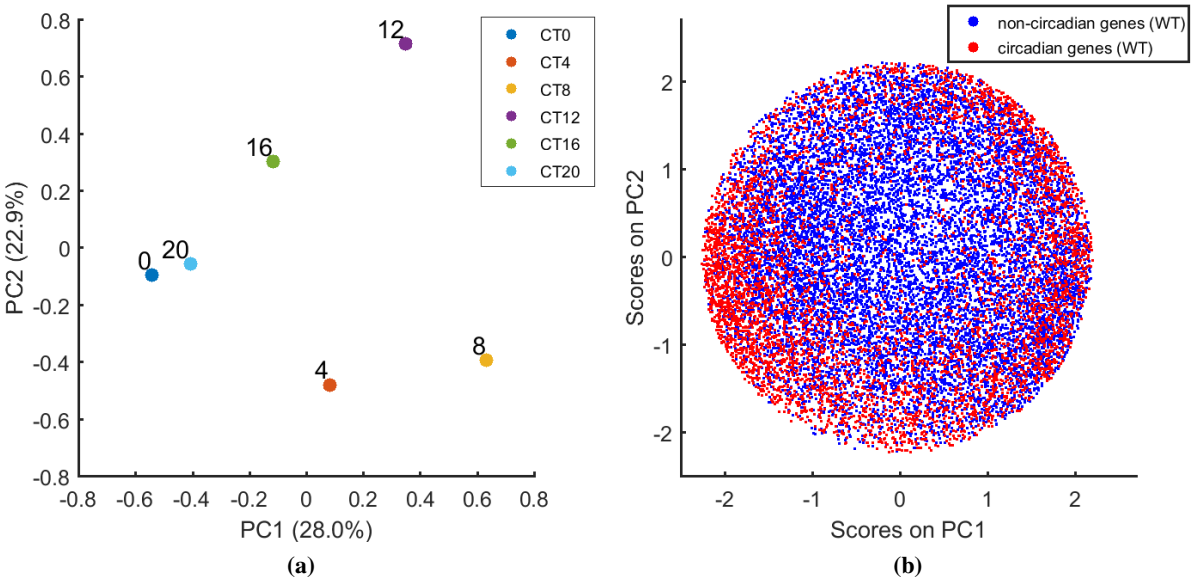


Figure 4.18: PCA model calculated on all genes of a single replicate dataset after SNV.
(a) Loading plot. **(b)** Score plot.

based on their phase less precise specially when two subsequent time points are located close to each other (for example CT0 and CT20). Moreover, each PC accounts for a smaller percentage of variability of the original dataset; PC1 and PC2 accounts for comparable variability. This is due to the fact that non-circadian genes included into the model disturb in some way the representation of the circadian ones. Only the first two PCs are considered in the PCA model, even though the third PC has a corresponding eigenvalue slightly above 1 (scree plot in Figure D.4 on page 121).

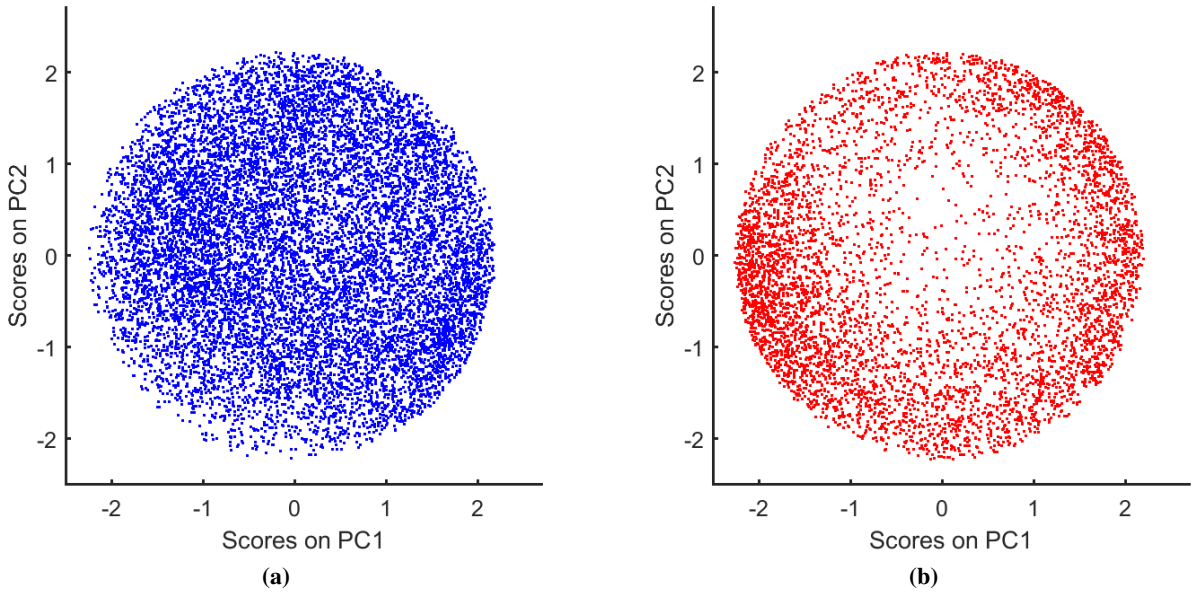


Figure 4.19: Score plot of PCA model calculated on all genes of one replicate after SNV.
(a) Score plot of non-circadian genes. **(b)** Score plot of circadian genes.

The score plot is shown in Figure 4.18b; it reports both circadian and non-circadian genes and it is possible to see that circadian genes are located mostly in the border, while non-circadian genes are located uniformly over the surface of the circle, as it has been observed for the PCA model on circadian and non-circadian genes separately. For a clearer representation, the two groups of genes are separately shown in Figure 4.19.

Figure 4.19 seems to indicate that circadian and non-circadian genes have the same temporal pattern since they co-localise in the plot, specially in the border of the score plot. However, variability among different individuals as well as technical variability should be taken into account in order to define a more robust model, since it has been shown that data variability cannot be neglected (Figure 4.14).

4.4.2.4 Whole transcriptome of multiple replicates

In this Section the results of PCA modeling on all replicate datasets, treated by hybrid unfolding, will be presented. As stated in Section 4.4.2.1, data variability is relatively high, thus it is very important to include all experimental replicates in PCA model to make it more robust.

Firstly, only the genes that are expressed (and detected) in all replicates are included in the PCA model; then a data matrix including all replicates is created by the method of hybrid unfolding as outlined in Section 4.3.5, to account for correlation and covariance between replicates as well as relationship between replicates. Then, the obtained data matrix is pre-treated by SNV and PCA is performed. The model parameters are shown in Figure 4.20.

In Figure 4.20a the correlation among the variables is reported, as it is explained according to PC1 and PC2. As in Figure 4.18a, the points are located in an approximately circular pattern, and the variables are ordered according to the temporal sequence they represent. This is due to the fact that circadian genes represent 28% of liver transcriptome, therefore their variability strongly affects the variability of the entire dataset. However, by considering all replicates, the amplitude of the angle by which subsequent time points are separated, is almost constant; in particular, the amplitude is $60° \pm 12.82°$, while it is $60° \pm 32.23°$ when only one replicate is considered as reported in Figure 4.18a. Thus, the standard deviation of the amplitude of the angles is 2.5 times smaller when all replicates are included in the model. This is very important since it will allow to distinguish more precisely between genes having different phases. Also in this case, PC1 and PC2 account for the same amount of variability of the data; compared to when a single replicate has been considered, the total explained variability by PCs is lower since a larger dataset is analysed.

In Figure 4.20b the relation among the genes is shown. In this Figure, as opposed to Figure 4.18b, a clearer spatial separation of non-circadian from circadian genes is visible, since circadian genes are located only in the outer region of the circle, while non-circadian genes are located in the inner area of

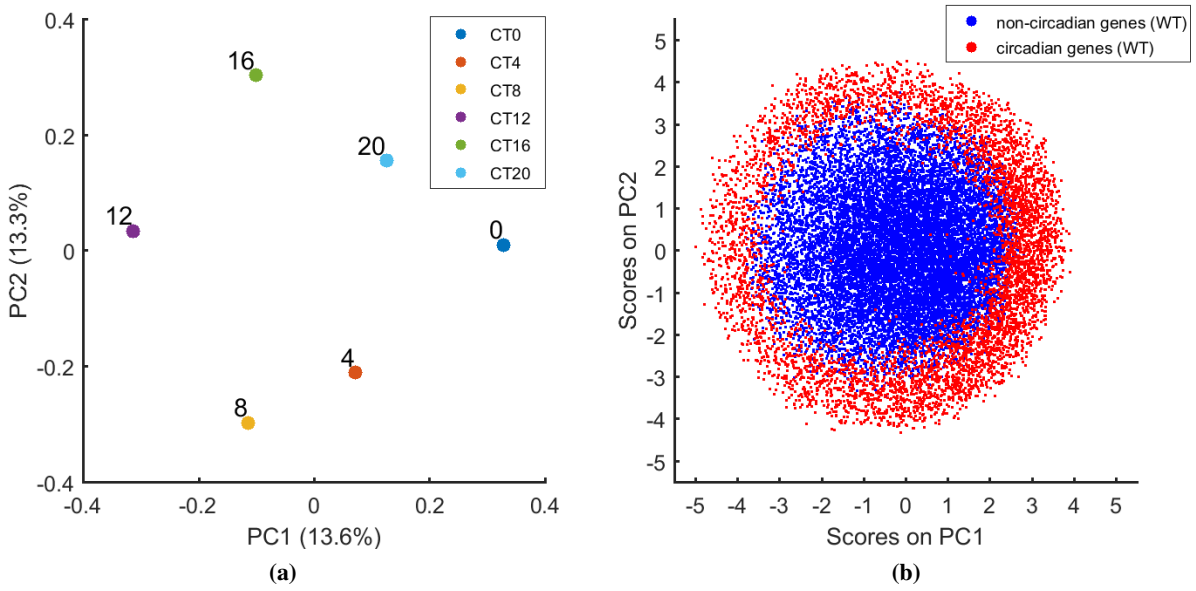


Figure 4.20: PCA model calculated on all replicates treated by hybrid unfolding and after SNV. **(a)** Loading plot. **(b)** Score plot.

the circle closer to the center of the plot. This result confirms the relevance of including all replicates in the analysis as well as that this method is in accordance to JTK_CYCLE classification of circadian genes.

The scree plot is shown in Figure 4.21: it reports the eigenvalue for its corresponding PC. It can be observed that PC1 and PC2 account for the same variance, and the same can be stated for PC3 and PC4. According to the rules explained in Section 4.3.4, the model should include the first 4 PCs.

In Figure 4.22, one of the most important properties of this method is reported, i.e. the identification of genes which are expressed with the same phase. In Figure 4.22a, a group of 394 genes is selected; they are located in a sector of 20° of amplitude centered around CT4. The temporal profile of these genes

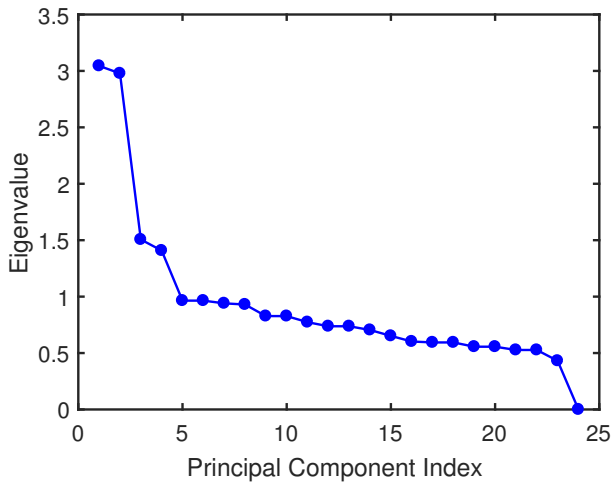
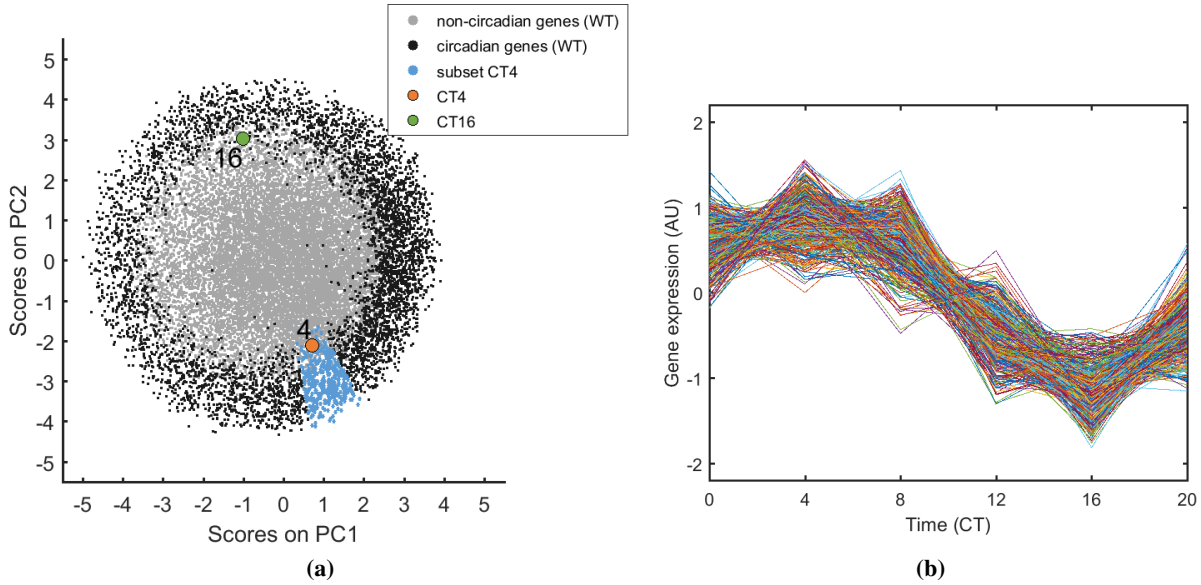


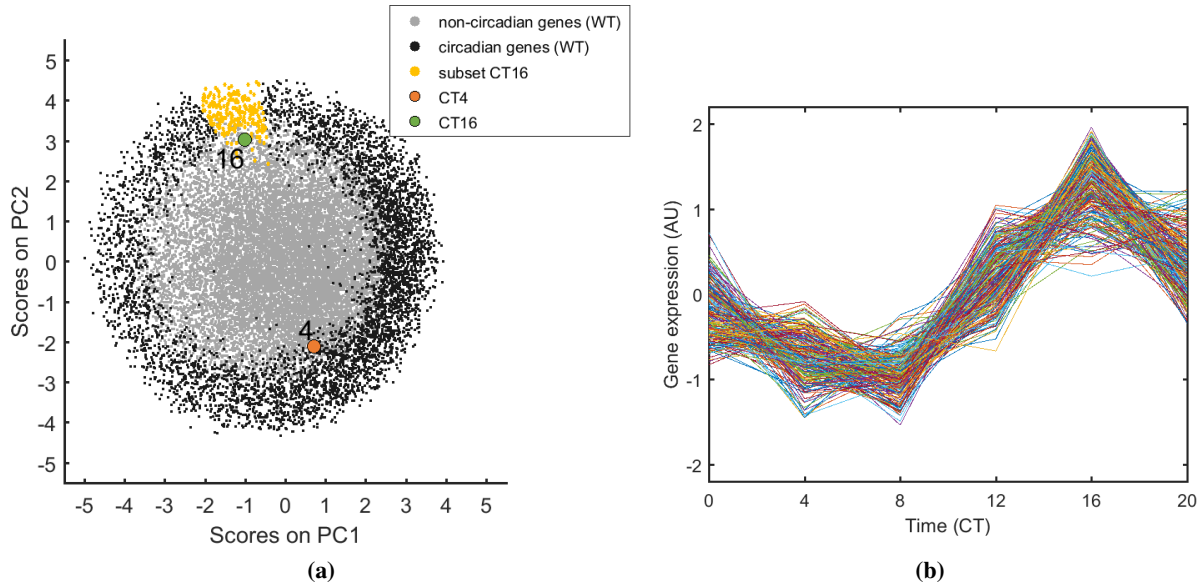
Figure 4.21: Scree plot for PCA model build on all replicates treated by hybrid unfolding



(a) (b)

Figure 4.22: Biplot and gene expression profile over time of genes having a peak at CT4.
(a) Biplot of PCA model calculated on all replicates treated by hybrid unfolding and after SNV, with highlighted a group of circadian genes located in correspondence of CT4. **(b)** Gene expression over time of 394 genes highlighted in Figure 4.22a. Gene expression corresponds to average of the SNV expression of the four replicates.

(represented as the average of the SNV expression of the four replicates) is shown in Figure 4.22b, where it is possible to observe that all genes display the same profile, having a peak of expression at CT4 and a trough at CT16 (CT16 is located in the opposite quadrant with respect to CT4 in the loading plot).



(a) (b)

Figure 4.23: Biplot and gene expression profile over time of genes having a peak at CT16.
(a) Biplot of PCA model calculated on all replicates treated by hybrid unfolding and after SNV, with highlighted a group of circadian genes located in correspondence of CT16. **(b)** Gene expression over time of 231 genes highlighted in Figure 4.23a. Gene expression corresponds to average of the SNV expression of the four replicates.

Similar considerations can be done by looking at Figure 4.23; in Figure 4.23a a group of 231 genes is selected, which are located within an angle of 20° centered at CT16. Their temporal profile is shown in Figure 4.23b, where it is possible to observe that these gene have the same phase of expression, having a peak of expression at CT16 and a trough at CT4. This temporal pattern is in anti-phase with the one shown in Figure 4.22b.

A finer characterization of circadian genes can be performed looking at the temporal profiles of circadian genes having the same phase but located progressively further from the axis origin, as shown in Figure 4.24a. By comparing Figure 4.24b, Figure 4.24c and Figure 4.24d it can be observed that the three groups of genes have the same amplitude of oscillation (the dataset has been treated by SNV); however, the furthest a group is located from the axis origin, the neatest its circadian oscillatory pattern is.

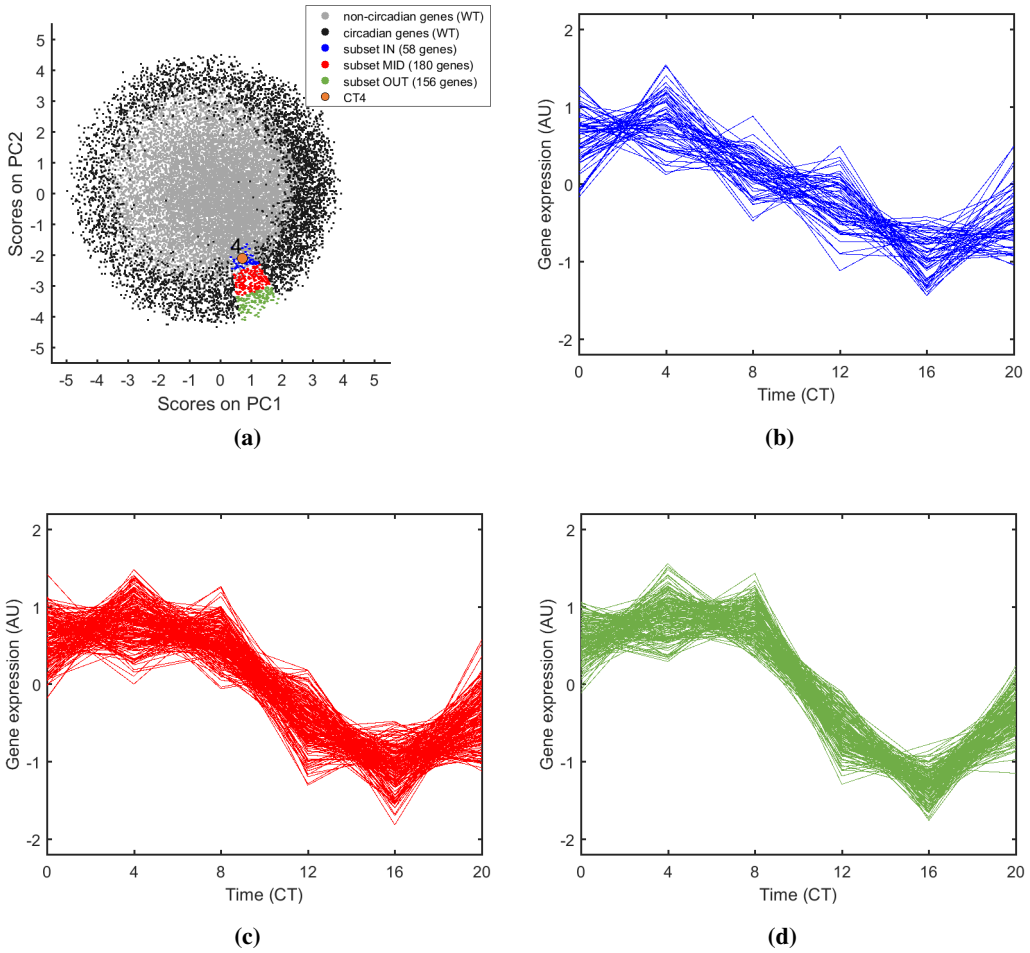


Figure 4.24: Biplot and gene expression profile over time of circadian genes having a peak at CT4. **(a)** Biplot of PCA model calculated on all replicates treated by hybrid unfolding and after SNV, with highlighted 3 groups of circadian genes located in correspondence of CT4 and progressively further from the axis origin. **(b)** Gene expression over time of 58 circadian genes highlighted in Figure 4.24a. Gene expression correspond to average of the SNV expression of the four replicates. **(c)** Gene expression over time of 180 circadian genes highlighted in Figure 4.24a. Gene expression correspond to average of the SNV expression of the four replicates. **(d)** Gene expression over time of 156 circadian genes highlighted in Figure 4.24a. Gene expression correspond to average of the SNV expression of the four replicates.

4.4.3 Applications of the method

In this Section three applications of the method will be described: transcriptome comparisons among different organisms, estimation of the circadian time of a sample, and enrichment of the genes expressed with the same phase.

4.4.3.1 Transcriptome comparisons

In this Section iKO phenotype will be considered in order to show how PCA model can be used to describe different transcriptome dataset. iKO genotype is the second considered mouse model by Yang et al. (2016), as explained in Section 4.4.1. In particular, in the paper it is stated that in iKO mice whole transcriptome is non-circadian, besides one single gene (*Erh*).

PCA model can be used to directly compare whole transcriptome data by projecting the transcriptome of another phenotype or another organ on the score plane of a reference organ or species. In this case, iKO mice transcriptome is projected on the score plane of WT mice, according to Equation (4.16), where $\mathbf{X}^{\mathbf{v}}$ is the dataset of iKO transcriptome over time and $\mathbf{P}$ is the loading matrix calculated for WT mice considering all replicates by hybrid unfolding.

The projection of iKO mice transcriptome is shown in Figure 4.25.

In Figure 4.25 it is possible to observe that iKO transcriptome almost overlap with non-circadian genes from WT mice, confirming the result stated in the paper. Moreover, *Erh* gene in iKO mice is the furthest located gene (among iKO genes) from the axis origin, coherently with the experimental result reported by Yang et al. (2016).

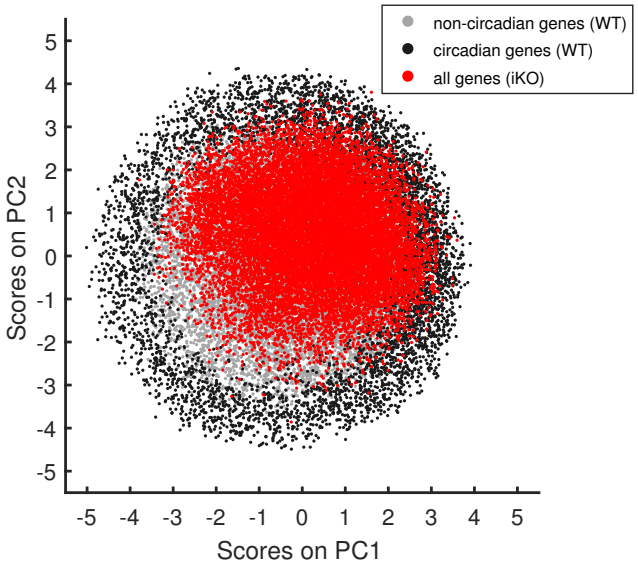


Figure 4.25: Score plot of hepatic whole transcriptome of WT organism and iKO organism (iKO transcriptome has been projected on WT transcriptome by PC1 and PC2)

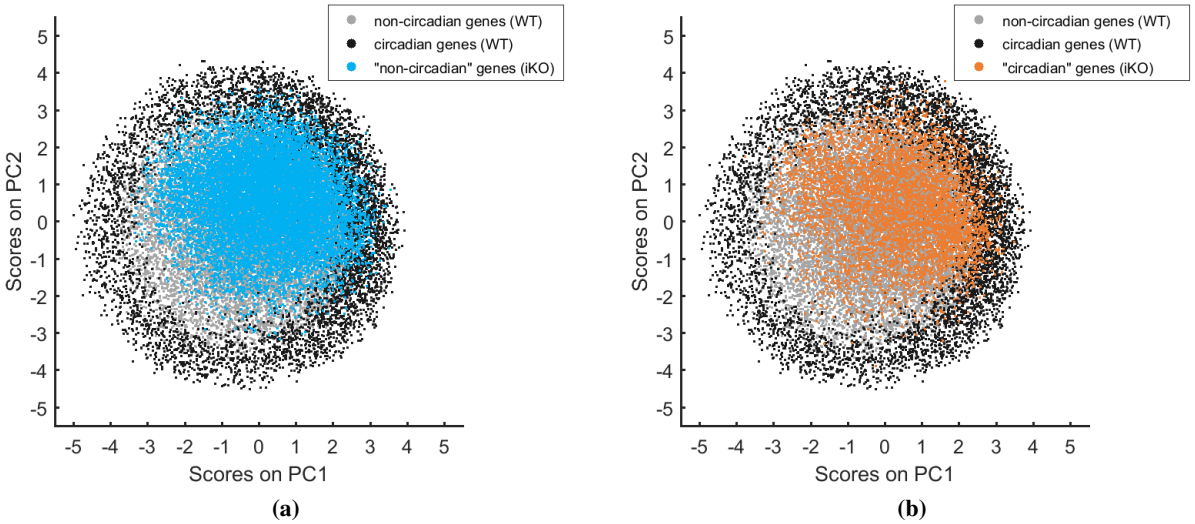


Figure 4.26: Score plot of hepatic whole transcriptome of WT mice and iKO mice.
(a) Score plot of hepatic whole transcriptome of WT mice and the projection of the subset of non-circadian genes (identified in WT mice) expressed by iKO mice. **(b)** Score plot of hepatic whole transcriptome of WT mice and the projection of the subset of circadian genes (identified in WT mice) expressed by iKO mice.

However, this method allows to do further considerations by looking at the position of the genes classified as circadian (in WT mice) in iKO mice compared to the position of the genes classified as non-circadian (in WT mice) in iKO mice. In Figure 4.26 these two parts of iKO transcriptome are shown separately; it is possible to observe that circadian genes completely lose their rhythmicity since they are almost uniformly distributed as non-circadian genes, without simply shrinking to the center of the plot but completely changing their distribution.

The change of expression in iKO mice of genes identified as circadian in WT mice can be also graphically represented as in Figure 4.27, where arrows indicate the change of expression of circadian genes in the two genotypes. The arrows originate from the position of a certain gene in WT mice and they point to the position of the same gene in iKO mice. Arrows are classified based on their normalized length into 11 groups, graphically represented by different colours.

Overall, it is possible to observe that there is a dominant and coordinated shift in the expression of circadian genes from the third to the first quadrant.

Moreover, genes could be grouped based on the magnitude and on the direction of their shift in expression to make some assumptions for their regulatory network. Figure 4.28a and Figure 4.28b show arrows corresponding to the second largest and largest spatial shift in the time of expression, respectively (according to the classification based on arrows normalized length).

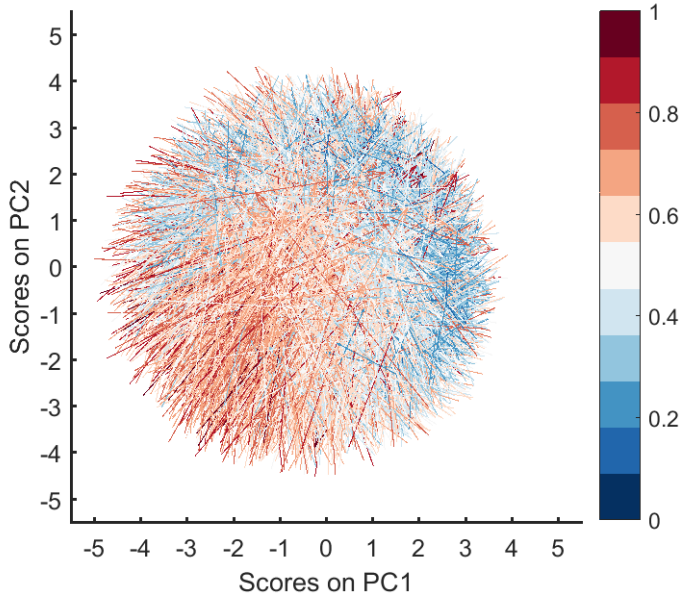


Figure 4.27: Score plot showing arrows which connect the same gene expressed in WT mice and in iKO mice. The arrows indicate the shift in expression of genes identified as circadian in WT mice. They originate from the position of the gene expressed in WT mice and they point towards the position of the gene expressed in iKO mice. The color bar indicates normalized arrow length.

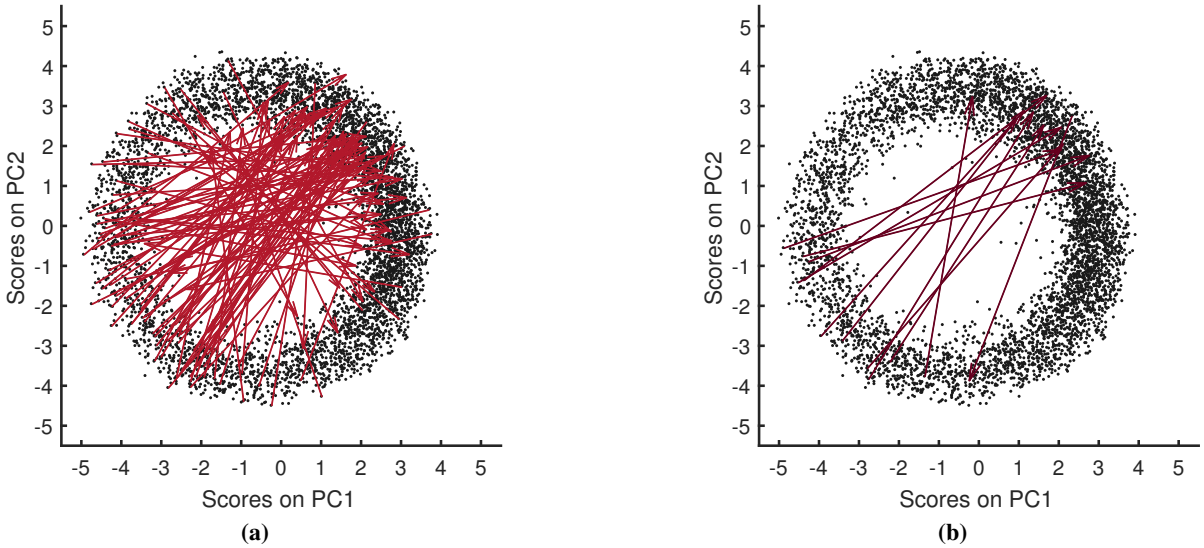


Figure 4.28: Score plot showing arrows which connect the same gene expressed in WT organism and in iKO organism.
(a) Score plot indicating the directions of temporal shift of 99 circadian genes (the ones having the second largest shift). The arrows indicate circadian genes (identified in WT organism) that display the largest shift in temporal expression. They originate from the position of the gene expressed in WT organism and they point towards the position of the gene expressed in iKO organism. **(b)** Score plot indicating the directions of temporal shift of 9 circadian genes (the ones having the largest shift). The arrows indicate circadian genes (identified in WT organism) that display the largest shift in temporal expression. They originate from the position of the gene expressed in WT organism and they point towards the position of the gene expressed in iKO organism.

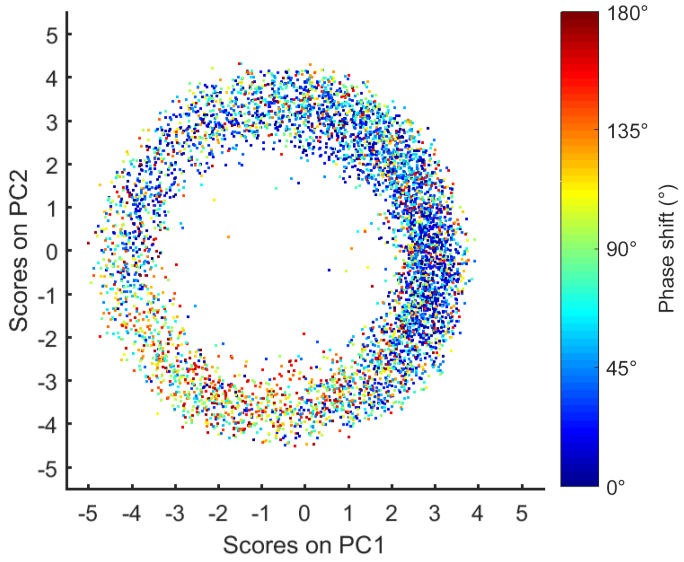


Figure 4.29: Score plot showing circadian genes of WT mice classified based on the phase shift they have in iKO mice

All genes in Figure 4.28b, for example, may be controlled by *Bmal1,* since when this gene is not expressed they change the time of their expression in a similar way; also genes in Figure 4.28a may be controlled by *Bmal1*, but they should also be further classified into groups having similar direction of the shift.

This method can also be used to classify the genes based on the amplitude of the phase shift they have between two biological conditions. Figure 4.29 shows circadian genes of WT mice depicted with different colors based on the amplitude of the phase shift: genes depicted in red are the ones that undergo the largest shift in temporal expression, thus they are the ones most likely to be under *Bmal1* control, while the ones that do not have phase shift may be *Bmal1* independent.

Thus, with this method different transcriptome can be projected on a calibration score plane and biological hypothesis can be defined based on changes at the transcriptome level.

4.4.3.2 Estimation of sample circadian time

PCA model can also be used to estimate the circadian time of a single sample. This is achieved by projecting on the score plane built on a calibration dataset another dataset, which could be formed by a single time point (i.e. all genes expressed by a tissue at a specific time point). This application can be useful when performing experiments, because it allows to reduce the number of samples.

The result is shown in Figure 4.30. The score plane has been built calculating PCA model on the transpose of $\mathbf{X}_1$ and $\mathbf{X}_2$, treated by hybrid unfolding. This is possible since PCA is a bi-linear modeling method. In this case the score plot will contain the time points, while the loading plot will show the correlation among genes. The transpose of $\mathbf{X}_3$ and $\mathbf{X}_4$, treated by hybrid unfolding, are projected by the

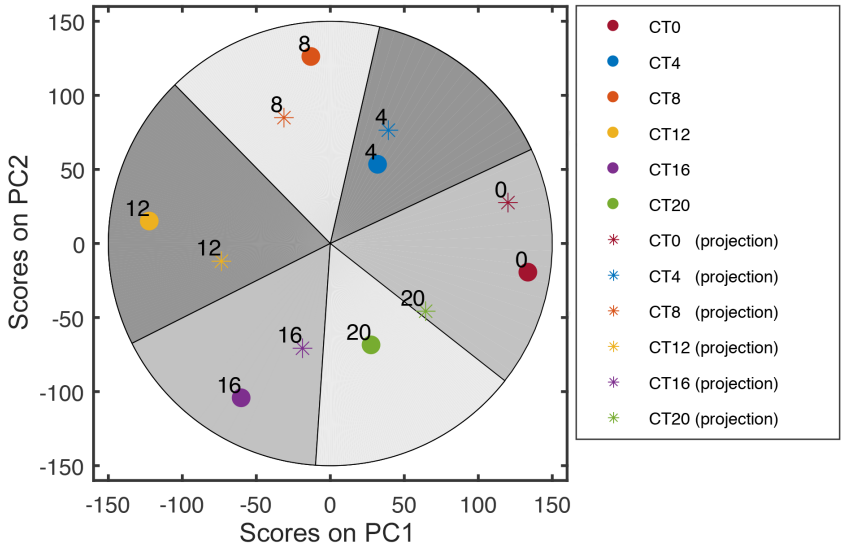


Figure 4.30: Score plot showing the scores of the calibration dataset (corresponding to a dataset formed by two replicates treated by hybrid unfolding) and the scores of the validation dataset (corresponding to a dataset formed by the other two replicates treated by hybrid unfolding).

loading (according to Equation (4.16)) on the score plane. To compare more projections, all time point are represented, but the same procedure could be applied also on a single time point. It is possible to observe that by the projection it is possible to precisely estimate the circadian time of a sample. The estimation of the circadian time has also been tried by projecting other RNA-seq datasets on the score plane built from this dataset, and it has been possible to estimate the circadian time also of other datasets.

Thus, this result shows that it is possible to estimate the circadian time of a sample given a robust calibration dataset.

4.4.3.3 Gene enrichment

Considering that genes can be grouped based on the peak of their expression, further considerations can be done characterizing, according to Gene Ontology (GO) categories, genes that are highly expressed in the same time-window. By GO, genes can be characterized according to three different aspects:

- Molecular Function (MF), to define molecular activities of gene products (for example catalytic activity, transporter activity, adenylate cyclase activity, etc.)
- Cellular Component (CC), to define the cellular compartment where gene products are active (for example nucleus, cytoplasm, ribosome, etc.)
- Biological Process (BP), to define pathways and larger processes made up of the activities of multiple gene products depending on their biological role (for example signal transduction, cellular physiological process, cell cycle, etc.).

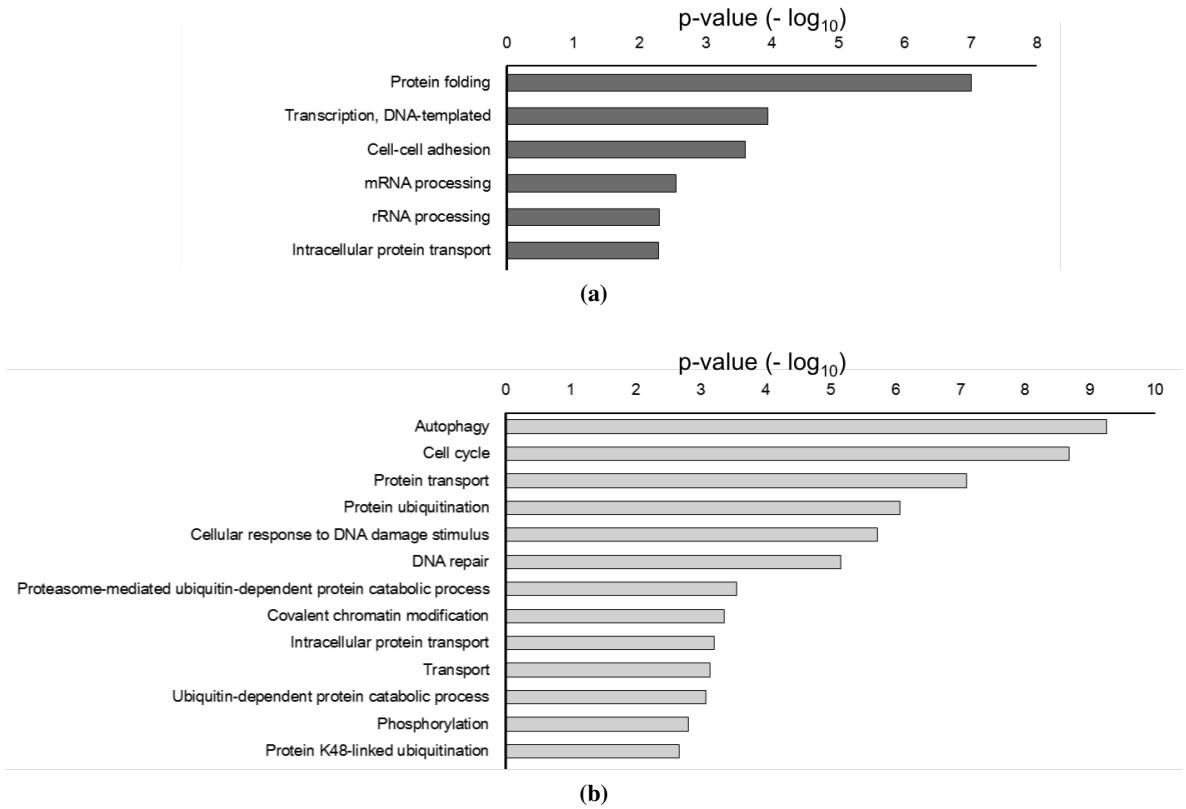


Figure 4.31: Biological processes enrichment.
(a) Biological processes enriched during the night. **(b)** Biological processes enriched during the day.

By this classification it is possible to understand if a specific category is enriched compared to a background, which in this case has been set to whole murine transcriptome. For example, it is possible to identify which BPs are active during the subjective night and compare them with the ones enriched during the subjective day. Figure 4.31a and Figure 4.31b show the biological processes enriched during night and day, respectively (Bonferroni-adjusted p-value$<1 \cdot 10^{-2}$).

It is possible to observe that during the active phase there is an enrichment of BPs involved in *de novo* protein synthesis, since it has been shown that external stimuli such as nutrition and exercise activate transcriptional and translational activities and that translation is circadian clock mediated (Atherton and Smith, 2012; Lipton et al., 2015). Instead, during the resting phase (i.e. day), the most enriched biological processes are the ones involved in clearance activities (such as ubiquitination), post-translational modifications and, in general, transformations involving proteins. The most significantly expressed process is autophagy, which is known to be active during fasting condition and it has been shown to be regulated in a circadian manner (Maiese, 2017; Vainshtein et al., 2014).

Thus, the method could be used to identify the temporal alternation of biological processes during the 24-hour day, and it might suggests new biological processes under circadian regulation.

4.4.4 Comparison with other algorithms

PCA model of whole transcriptome data is a different method compared to the algorithms described in Section 4.1.3. First of all, it is a completely new approach for transcriptomic data analysis since it is not based on fitting on cosine-based patterns or on spectra analysis methods but it is based on linear decomposition. It allows to describe large datasets by few variables (PCs) that account for the largest direction of variability of the data, maintaining the information embedded in the dataset. Actually, the method does not require to model the data on calculated time trajectories and it is entirely based on experimental data.

PCA model has been applied to an experimental dataset since the method does not require any reference. However, PCA could also be used to model a calculated reference dataset, formed by all possible temporal profiles. Then, by PCA model parameters, experimental datasets could be projected on the PCA plane built on the calculated dataset. Thus, the experimental temporal profile will be projected close to its most similar calculated profile. In this sense PCA model would be similar to model-based approaches as COSOPT, i.e. the closest an experimental profile of a gene is to a calculated cosine curve, the smallest the distance in the score plot between the gene and the point representing the calculated profile will be.

It has been shown that its robustness is increased by increasing the replicate number; the proposed method can "average" the characteristics among different replicates, discriminating what is circadian and what is not, independently from the individual differences. It is also computationally efficient and fast even for very large datasets, since it does not require iterations nor permutations.

However, its limitation is that it does not take into account the amplitude of transcripts but only their period and phase, since they have been considered the most important features of temporal expression (the information about amplitude is lost when data are pre-treated). If the aim is to analyse expression amplitudes, the data could be modeled by PCA without pre-treatment, making the amplitude the most important feature of the data besides period and phase.

The results described in Section 4.4.2.4 show that the developed method is in complete accordance to the classification of circadian and non-circadian genes done by JTK_CYCLE, which has been described as more reliable and efficient than COSOPT and Fisher's G-test due to its increased resistance to outliers as well as being computationally faster (Hughes et al., 2010). Moreover, the strong advantage of the proposed method is that it improves the visualization of whole transcriptome data, making it a very versatile approach to clearly distinguish between circadian and non-circadian genes, to group genes based on the phase of their expression and to compare different biological conditions.

4.5 Conclusion

The analysis method presented in this Chapter based on PCA model allows to visualize whole transcriptomic data at a glance and it clearly allows to identify circadian and non-circadian transcripts, as well as ultradian pattern of expression. By this method, transcripts can be grouped based on the phase of their expression, which may be useful in identification of group of genes belonging to the same regulatory network. Moreover, it allows to compare different biological conditions and to characterize their similarities and differences at the transcriptional level. The method can be also used to estimate the circadian time of a sample, thus it can be used to reduce time and experimental costs. Finally, it could also be used to identify processes that are under circadian control by enrichment analysis.

In general, transcriptomic approach, if coupled with a proper analysis method, is more comprehensive and powerful compared to conventional technologies based on qrtPCR and bioluminescence, since it allows to describe the evolution over time of all biological responses, thus helping in the definition of the gene network.

Chapter 5

Conclusion

Circadian clock crucial biological function is to maintain the physiological state of an organism and its alignment with the external environment, by acting as a driver of metabolic physiology, partitioning behavioural and metabolic processes according to time of day within each tissue. The continuous interaction between the endogenous circadian system and 24-hour environmental/behavioural cycles is ensured by the specific feature of the circadian clock of being entrainable to external cues, i.e. light/dark cycles and feeding/fasting schedules. Considering that food has been discovered as a potent synchronizer for peripheral clocks, and that altered feeding regimen are associated with several diseases, understanding how metabolites can entrain and affect circadian clock and clock-controlled genes is of paramount importance.

Therefore, this Ph.D. research work is aimed at the development of technologies and data analysis tools to investigate the entrainment of peripheral mammalian circadian clock to frequency-encoded metabolic stimuli.

First, technologies have been developed to investigate the effects of periodic metabolic entrainment showing that oscillatory periodic metabolic stimulations entrain *Per2* expression in murine fibroblasts. In particular, it has been shown that anti-phase cyclic metabolic stimulations are able to completely reset fibroblasts clock, leading to a 12 hour phase shift. This is relevant from a biological point of view, because it means that metabolic oscillations are able to synchronize the phase of peripheral clocks. Moreover, the key role of the starvation pathway has been identified, showing that starvation is necessary for circadian phase resetting.

Then, in order to develop an *in vitro* model that can resemble oscillations (physiological or pathological) at which peripheral tissues are exposed *in vivo*, a new automated microfluidic platform has been designed, tested and validated. This platform allows to perform frequency-encoded experiments with precise timing of the stimuli as well as with precise spatial control of the microenvironment inside culture chambers.

Finally, a data analysis method for transcriptomic time series has been developed to identify circadian genes in whole transcriptome data, to group genes based on the phase of their expression, to visualize transcriptome data at a glance and clearly identifying modifications at the transcriptome level from one biological condition to another one.

The future perspectives of this work concern the applications of the developed technologies, both the automated platform as well as the data analysis tool. They could be used for chronotherapy applications, which is the branch of pharmacology which studies the side-effects and efficacy of drugs based on their administration time, since it is known that side-effects and efficacy depend on the time at which drugs are taken. The automated microfluidic platform could be used for frequency-encoded experiments in which a drug is delivered at different timing: it could be used to evaluate the shift in the phase of circadian clock gene expression as well as the variations of other cellular markers (associated with apoptosis or inflammation, for example) caused by different timings of drug administration. The data analysis method could be used with the same target, i.e. understanding which is the administration time that minimizes side effects and maximizes therapeutic outcome, since the drug effects would be visible at the transcriptional level, thus analysed with the developed method. The analysis method could also be used to identify differences and similarities in circadian transcriptomes between organs and across different species, being a useful tool for broadening the understanding of clock-controlled genes and for suggesting new possible interconnections between genes, which will help in the definition of a comprehensive gene regulatory network.

Appendix A

Multilayer Soft Lithography: mold and platform fabrication

In this Appendix the complete workflow for the fabrication of the multilayer microfluidic platform used in Chapter 3 will be presented. It comprises platform design and photomasks, soft lithography protocol for control and flow molds and the protocol for platform fabrication. Multilayer microfluidic platform design, protocols and know-how have been based on previous work done by Zambon (2013).

A.1 Platform design

Platform design is performed in AutoCAD 2D. The design is a critical step since many constraints have to be considered, both for the present and for possible future applications. Considering the specific application, it is desired to have 4 independent conditions, each one in 4 independent biological replicates, thus a total of 16 culture chambers; the platform must have 2 inlets, and for each group of chambers, independently from the others, it must be possible to select which inlet should be used, the time interval between the use of one inlet and the subsequent one, and the overall number of changes. Moreover, the platform must work properly for 7 days, therefore it should be operated at the lowest possible pressure to reduce delamination risks.

Therefore, technical design decisions are taken as described in Table A.1.

Table A.1: Design characteristics of the multilayer microfluidic platform

Characteristic	Design decision	Note
Valve type	push-up	Lower actuation pressure than push-down valves (thus less mechanical stress)
Binding	plasma binding	The pattern of the chip should make one shot binding feasible (add mask alignment marks)
Minimum closing pressure	5 psi	
Flow channel width	170 μm	
Control channel width	200 μm	
Valve area	170 x 200 μm	This valve area should allow low actuation pressure
Flow channel height (rounded channels)	35 μm	
Flow channel height (squared channels)	250 μm	
Control channel height	35 μm	
Flow driving force	pressure	
Number of flow inlet	2	
Number of flow outlet	1	
Number of control inlet	7	
Number of chambers	16	Grouped in group of 4
Chamber length	10 mm	
Chamber width	2 mm	
Chamber surface	20 mm^2	The chamber should have this surface for imaging purposes
Chamber height	250 μm	The chamber must be below the 1:10 height-to-width recommended ratio
Chamber volume	4.6 μl	
Total valves number	35	

From the Table it is possible to understand that two different types of channels are present in the flow layer, created with two different photoresists: positive photoresist will be used in the flow layer where a valve will be formed.

The pattern of the flow layer and of the control layer are shown in Figure A.1 and in Figure A.2, respectively.

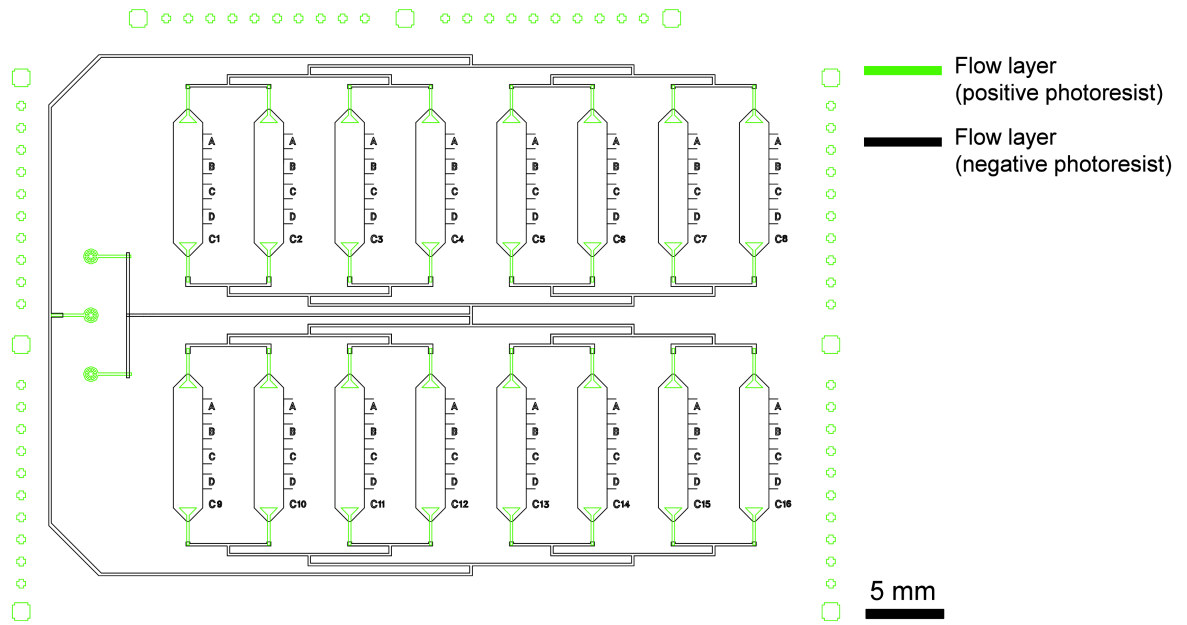


Figure A.1: Design of the flow layer, indicating the pattern that will be obtained with positive photoresist and the pattern that will be done with negative photoresist

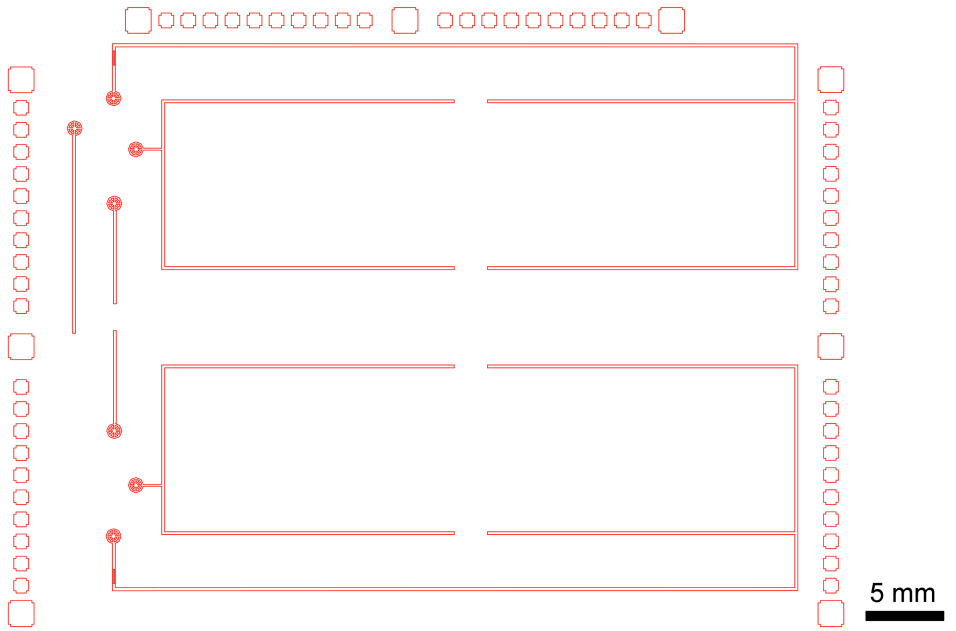


Figure A.2: Design of the control layer

A.2 Photomasks

Soft lithography technique requires the use of photomasks, which are transparent plastic films printed with the desired pattern. They are prepared from the AutoCAD drawing of the molds in Adobe Illustrator, which is used to paint with black specific areas. Prior to printing, flow photomasks must be scaled up by 1.5% to compensate for the shrinkage of PDMS after peeling. Their printing resolution has to be

sufficient to fabricate the smallest dimension of the pattern. In this case they have been printed at 2400 dpi resolution. They are placed over the mold before exposure to UV light to selectively direct UV light in specific parts of the mold, which are the transparent parts of the photomask. Depending on the desired cross-section and on the height of the channels, positive or negative photoresists can be used; negative photoresists are used for channels with a rectangular profiles, while positive photoresists are used for channels with rounded profiles. Negative photoresists can also be used for patterns with a smaller width-to-height ratio than positive photoresist. They have an opposite behaviour with respect to exposure to UV light. Negative photoresist polymerizes in the surfaces exposed to UV light, while it does not polymerize in areas not exposed to UV light. On the other hand, positive photoresist polymerizes in areas not exposed to UV light, while it does not cross-link in areas exposed to UV light. Therefore photomasks have to be designed accordingly to the photoresist that will be used.

In Figure A.3 and in Figure A.4, the two photomasks used for the flow layer are shown, since two different photoresists are used.

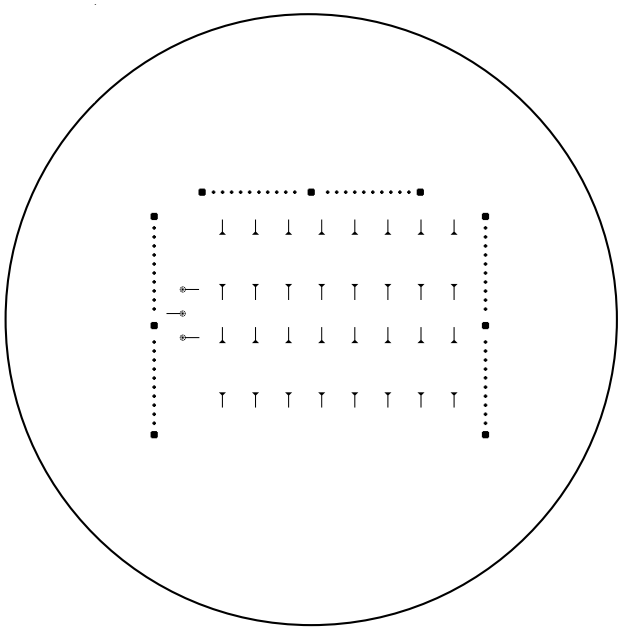


Figure A.3: Photomask of flow layer for SPR-220-7 (positive photoresist) used to fabricate one part of the flow layer through soft lithography

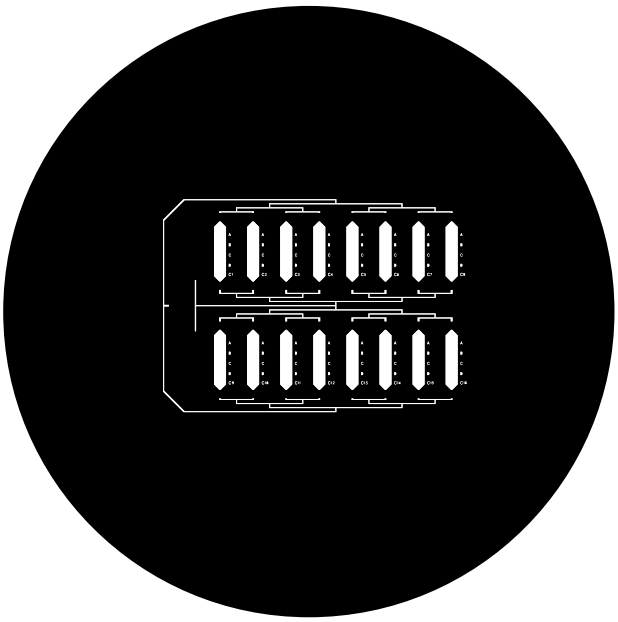


Figure A.4: Photomask of flow layer for SU-8 2100 (negative photoresist) used to fabricate one part of the flow layer through soft lithography

In Figure A.5 the photomask for control layer is shown.

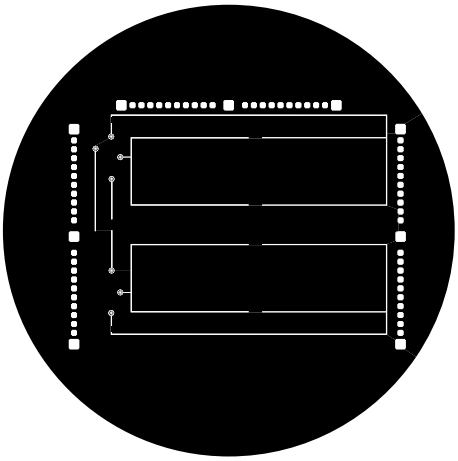


Figure A.5: Photomask used to fabricate the control layer through soft lithography

A.3 Soft lithography for flow mold

The protocol for flow mold fabrication is a two step procedure, since it involves first the positive photoresist and then the negative photoresist.

A.3.1 Positive Photoresist

The protocol has been developed based on photoresist datasheet.

A.3.1.1 Silicon wafer pre-treatment

A silicon wafer (SIEGERT WAFER, Germany) with a diameter equal to 4 in is cleaned rinsing consecutively with Acetone (Sigma-Aldrich), Methanol (Sigma-Aldrich) and distilled water to help the adhesion of the photoresist in the next steps. Then, the wafer is dried using compressed air and placed on a hot plate at 105 °C for at least 10 min to remove humidity. Finally, the wafer is treated with Hexamethyldisilazane (HMDS, Sigma-Aldrich) vapor for 10 min.

A.3.1.2 Photoresist deposition

The pre-treated wafer is inserted in a spin coater (WS-650-23NPP, Laurell) and it is secured in the chuck by vacuum generated by a pump (II-3-GD-T4 X, Edwards). At this point, one layer of SPR-220-7 positive photoresist (Rohm and Haas, Dow Corning) is distributed and spin-coated over the wafer to obtain a thickness of 15-20 μm; the spinning protocol is as follows:

1. 250 rpm for 15 s with an acceleration of 1064 rpm/s
2. 560 rpm for 105 s with an acceleration of 1064 rpm/s
3. 2000 rpm for 4 s with an acceleration of 9708 rpm/s

A.3.1.3 Soft bake

After another treatment of 10 min with Hexamethyldisilazane (HMDS, Sigma-Aldrich) vapor, soft bake is done placing the wafer on a hot plate at 90 °C for 10 min. This thermal step is meant to remove the internal stress of the photoresist and to partially evaporate the solvent contained in the photoresist.

Then photoresist deposition is repeated as described in Section A.3.1.2, to obtain a final positive photoresist thickness of 30-40 μm; soft bake is performed again, at 90 °C for 60 min.

A.3.1.4 Rehydration

The wafer is maintained in the dark for at least 3 hours to rehydrate the photoresist.

A.3.1.5 Exposure

In the exposure step an UV lamp is used (collimated UV light, 365 nm center wavelength). The photomask (Section A.2) has to be placed over the wafer before exposure, to selectively polymerize the photoresist (with positive photoresist, areas not treated by UV light polymerize). The needed exposure energy in this case is equal to 2000 mJ/cm^2. Since this value is very high, also the exposure time will be long: it is necessary to expose the photoresist for a time interval not longer than 60 s and to wait at least other 60 s between one exposure and the subsequent to avoid excessive heating of the material.

A.3.1.6 Development

The wafer is developed in a solvent (Microposit MF-319, Microchem) to dissolve not cross-linked polymer. A glass trough is filled with solvent and the wafer is completely immerged into it, then the glass trough is placed over a rocking shaker. The requested time for this step is about 15 min, then the development is stopped rinsing the wafer with distilled water and drying it with compressed air.

A.3.1.7 Hard bake

Hard bake is done placing the wafer over a hot plate for 10 min at 65 °C and then setting a ramp of 10 °C/h for 15 h to reach the maximum temperature of 190 °C. This specific thermal profile allows to obtain channels with a rounded profile (photoresist undergoes reflow at temperature below 140 °C). At the end of this last step, the wafer is left on the hot plate until it has reached room temperature to avoid thermal shock of the material.

A.3.2 Negative Photoresist

The protocol has been developed based on photoresist datasheet.

A.3.2.1 Silicon wafer pre-treatment

The pre-treatment of the wafer with the deposited and hard baked positive photoresist is performed as described in Section A.3.1.1.

A.3.2.2 Photoresist deposition

One layer of SU-8 2100 negative photoresist (Microchem) is spin coated over the wafer (where the positive photoresist has been spun before) to obtain a thickness of about 250 μm. To obtain this thickness, the spin protocol is as follows:

1. 500 rpm for 10 s with an acceleration of 100 rpm/s
2. 1000 rpm for 32 s with an acceleration of 300 rpm/s

Then the silicon wafer is left on a flat surface for 10 min at room temperature to better distribute the polymeric chains of photoresist.

A.3.2.3 Soft bake

Soft bake is done at 90 °C for 45 min on a hot plate. By soft bake, part of the solvent contained in the photoresist evaporates.

A.3.2.4 Exposure

The photomask (Section A.2) must be placed over the wafer and perfectly aligned with the positive photoresist layer with the help of device alignment marks; then the wafer is exposed to UV light. The needed exposure energy in this case is equal to 350 mJ/cm^2.

A.3.2.5 Post exposure bake

The post exposure bake is performed placing the wafer on a hot plate at 95 °C for 15 min.

A.3.2.6 Development

The development step requires a solvent to dissolve not cross-linked polymer from the silicon wafer. A glass trough is filled with Propylene Glycol Monomethyl Ether Acetate (Sigma-Aldrich) and the wafer is completely immerged into it, then the glass trough is placed over a rocking shaker. The development time in this case is equal to about 17 min. At the end of the development step, the wafer is rinsed thoroughly with Isopropyl Alcohol (Sigma-Aldrich) and subsequently dried with compressed air.

A.3.2.7 Hard bake

The last step of the soft lithography procedure is another heat treatment, called hard bake. It is meant to improve the mechanical and thermal characteristics of the mold and to ensure a good durability, and it is done placing the wafer on a hot plate for 10 min at 65 °C and then setting a ramp of 120 °C/h for 2 h to reach the maximum temperature of 160 °C. At the end of the thermal treatment, the wafer is left on the hot plate to gradually reach room temperature to avoid thermal shock of the material.

A.4 Soft lithography for control mold

Soft lithography of control mold is a two-step procedure since two layers of photoresist are deposited one on top of the other. Operating conditions and protocol have been taken from photoresist datasheet.

A.4.1 Blanket layer

The blanket layer provides a surface with uniform wetting properties to PDMS, which ensures a uniform PDMS spin-coating thickness when making the control layer.

A.4.1.1 Silicon wafer pre-treatment

The pre-treatment of a 3 in wafer is performed as described in Section A.3.1.1.

A.4.1.2 Photoresist deposition

One layer of SU-8 2005 negative photoresist (Microchem) is distributed and spin-coated over the wafer to obtain a thickness of about 5 μm. The spin protocol is as follows:

1. 500 rpm for 10 s with an acceleration of 100 rpm/s
2. 3000 rpm for 35 s with an acceleration of 300 rpm/s

After spinning, the silicon wafer is left on a flat surface for 5 min at room temperature to distribute spun photoresist.

A.4.1.3 Soft bake

Soft bake is performed at 95 °C for 5 min on a hot plate.

A.4.1.4 Exposure

The wafer covered with the blanket layer of spun photoresist is exposed without any photomasks at 105 mJ/cm^2.

A.4.1.5 Post exposure bake

Soft bake is performed at 95 °C for 3 min on a hot plate.

A.4.2 Channel layer

The channel layer is shaped with the pattern of the channels of the control layer.

A.4.2.1 Silicon wafer pre-treatment

The pre-treatment of the wafer with the blanket layer is performed as described in Section A.3.1.1.

A.4.2.2 Photoresist deposition

One layer of SU-8 2050 negative photoresist (Microchem) is distributed and spin-coated over the wafer to obtain a thickness of about 35 μm. The spin protocol is as follows:

1. 500 rpm for 10 s with an acceleration of 100 rpm/s
2. 4000 rpm for 45 s with an acceleration of 300 rpm/s

After spinning, the silicon wafer is left on a flat surface for 5 min at room temperature to distribute spun photoresist.

A.4.2.3 Soft bake

Soft bake is performed at 95 °C for 6 min on a hot plate.

A.4.2.4 Exposure

The wafer covered with the two layers of photoresists is exposed with the control layer photomask (Section A.2) at 160 mJ/cm^2.

A.4.2.5 Post exposure bake

Soft bake is performed at 95 °C for 6 min on a hot plate.

A.4.2.6 Development

The development is performed in Propylene Glycol Monomethyl Ether Acetate to dissolve not-cross-linked photoresist. The development time is 5 min. At the end of development, the wafer is washed thoroughly with Isopropyl Alcohol and subsequently dried with compressed air.

A.4.2.7 Hard bake

The Hard Bake is done placing the wafer on a hot plate for 10 minutes at 65 °C and then setting a ramp of 120 °C/h for 2 h to reach the maximum temperature of 160 °C. At the end of this last step, as usual, the wafer is left on hot plate until it has reached room temperature to avoid thermal shock of the material.

A.5 Platform fabrication

The microfluidic platform consists of two PDMS layers. The first has the channels where liquid flow occurs (flow layer); the second has dead-ended channels that control the valves (control layer). The fabrication procedure consists in the preparation of the two layers and then their binding.

At the beginning, each mold is treated with Chlorotrimethylsilane (Sigma Aldrich) vapour for 1 h at room temperature.

The flow layer is made by pouring uncured PDMS (Sylgard 184, Dow Corning), prepared with a 1:10 curing agent-to-base weight ratio, onto the flow mold to a thickness of 3 to 5 mm. After degassing under vacuum in a crystallization tank for 20 min and letting the mold on a flat surface for 1 h, PDMS is cured on the mold by baking it at 75 °C for 1.5 h on a flat hot plate. After curing, the thick PDMS layer is peeled off its mold and punched in correspondence to inlet and outlet holes with a dispense tip (Nordson EFD) having external diameter equal to 0.91 mm.

The control layer is made by spin coating a layer of PDMS (prepared with a 1:10 curing agent-to-base weight ratio) over the control mold to obtain a thickness of about 80 μm[1], which is achieved by a spin protocol with one step at 1300 rpm for 75 s with an acceleration of 300 rpm/s. Also the glass slide is coated with a layer of PDMS of about 10 μm to improve the adhesion to PDMS control layer and the resistance to pressure during automation operations. For glass slide, the spin protocol has one step at 2300 rpm for 75 s with an acceleration of 300 rpm/s. Curing of PDMS on glass slide and control layer is done on a hot plate with a heating ramp of 40 °C/h for 2 h, starting from room temperature and reaching the maximum temperature of 80 °C.

Flow layer is then attached by plasma treatment over the control layer, without peeling it off the mold. After a strong adhesion between control and flow layer is achieved, the control layer is cut and peeled off its mold. At this point, the layers are holed with the same needle used for the flow layer to create the inlets for the control layer, and then the chip is attached over the glass slide by plasma treatment. Finally, as usual, the microfluidic platform is placed on a hot plate at 80 °C for at least 10 minutes to help the complete adhesion of the control layer on the glass slide.

A.6 Platform automation

The automation of the microfluidic platform is performed by the lab-made instrumentation shown in Figure A.6.

It is composed of three parts:

1. Electronic unit for the generation of digital signals (NI USB-6501, National Instruments), in Figure A.6a and A.6e: this part has also a system of rechargeable batteries to guarantee charge during relocation of the platform from the incubator to the microscope

2. Pressurized vial system (Figure A.6a and A.6d): this part is composed of a pressure regulator (CDK Corporation) and 360° polycarbonate luer connectors (Cole Parmer)

3. Solenoid valves (24V, CDK Corporation), in Figure A.6a and A.6f: each of them is connected to a tube (PTFE, ID 0.032”, OD 0.056”, Cole Parmer) that will be filled with liquid (usually water) and connected to the control inlets of the microfluidic platform.

In Figure A.6b and A.6c it is possible to see an example of pressurized vial (Greiner Bio-One) and some particulars of its fabrication: it has two holes on the lid in which two needles (ID 21G, OD 0.8mm, Chemil) are inserted and glued with epoxy glue; these two needles are the inlet for pressurized air and

[1] Since the height of microfluidic structures for the control mold is equal to 35 μm, the membrane for the on-chip push-up valves will have a height of about 50 μm: this value ensures a good mechanical resistance of the membrane and a reasonable closing pressure for the on-chip valves.

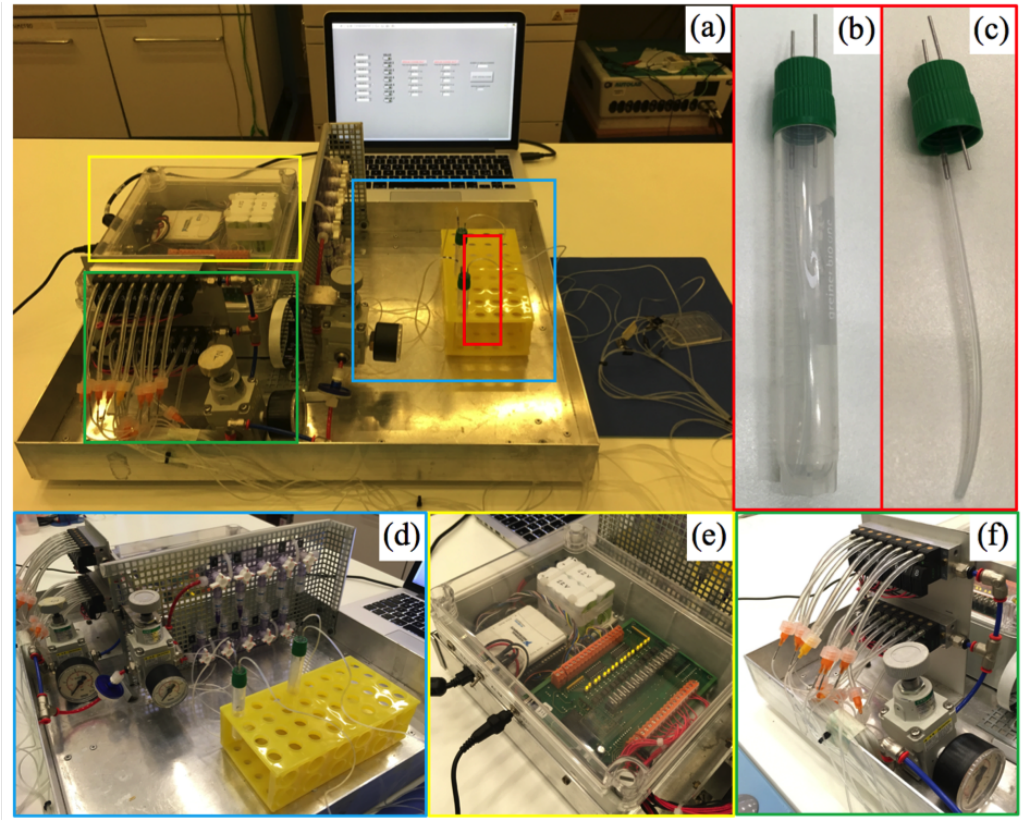


Figure A.6: Automation system. (a) Automation system, a computer to its control and the final configuration of an automated multilayer platform; (b), (c) Example of pressurized vial and some details of its fabrication; (d) Pressurized vials system; (e) Electronic unit for the generation of digital signals, batteries and current amplificatory; (f) Solenoid valves.

the outlet for the liquid inside the vial; the vial outlet has a short tube (PTFE, ID 0.032”, OD 0.056”, Cole Parmer) that connects the needle to the bottom of the vial, so when it is pressurized through the inlet the liquid purges the tube and comes out; on the other side of the vial outlet there is a long tube (PTFE, ID 0.022”, OD 0.042”, Cole Parmer) that ends with a needle to be easily inserted in the inlets or outlet of the flow layer in the platform.

The interface between the automation system and the computer is done using a custom software developed with LabVIEW.

Appendix B

Drop-seq

This Appendix contains the preliminary experimental results obtained regarding Drop-seq technology.

B.1 Drop-seq technology

Drop-seq is a recently published strategy for performing high-throughput single-cell RNA-seq (Macosko et al., 2015). The establishment of this technology has been started to allow single-cell RNA-seq of a large number of cells (from 10,000 to 100,000) at a very low price (approximately 1 Euro per cell), as opposed to conventional single-cell approaches where few cells are processed (from 20 to 100) at a much higher price per cell (approximately 200 Euro). The difference between the two approaches lays in the number of detected genes: by high-throughput approaches approximately 1,000 genes can be detected per cell, while with low throughput technology around 15,000 genes can be detected per cell.

Even though in high-throughput methods the number of detected genes is lower than in conventional methods, the identified genes are sufficient to assess cell heterogeneity in cell populations; therefore this method allow to discover new cell phenotypes and define transcriptionally distinct cell populations.

Drop-seq workflow is schematically represented in Figure B.1. It is based on a custom-designed microfluidic device that joins two aqueous flows, one flow contains single-cell suspension, and the other flow contains barcoded primer beads suspended in a lysis buffer. The aqueous flow is compartmentalized into discrete nanoliter-sized droplets by adding oil. The number of created droplets greatly exceeds the number of beads or cells injected, so that a droplet will generally contain zero or one cells, and zero or one beads. Millions of nanoliter-sized droplets are generated per hour, of which thousands contain both a bead and a cell. Immediately following droplet formation, the cell is lysed and releases its mRNAs, which then hybridize to the primers on the microparticle surface. The droplets are broken by adding a reagent to destabilize the oil-water interface, and the microparticles collected and washed. The mRNAs are then reverse-transcribed in bulk, forming single-cell transcriptomes attached to microparticles (STAMPs), and

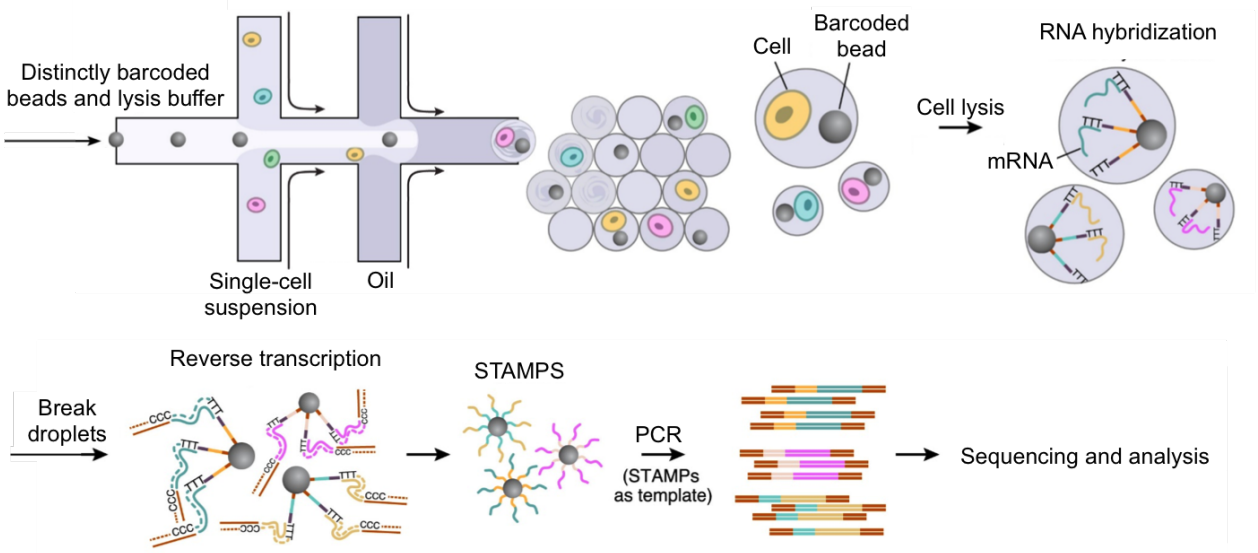


Figure B.1: Schematic of single-cell RNA-seq library preparation with Drop-seq. Adapted from Macosko et al. (2015)

template switching is used to introduce a PCR handle downstream of the synthesized cDNA. At the end, the samples are sequenced and it is possible to retrieve quantitative information of mRNA of each single-cell thanks to the structure of the barcoded primers that cover the surface of the beads, shown in Figure B.2a. Briefly, primers on all beads contain a common sequence (namely PCR handle) to enable PCR amplification after STAMP formation. Then, the cell barcode is a 12 bases sequence identical across all the primers on the surface of any one bead, but different from the cell barcodes on other beads to mark with the same sequence all transcripts of the same cell. Finally, unique molecular identifier (UMI), a 8 bases sequence different on each primer. to identify PCR duplicates.

The design of Drop-seq microfluidic chip is shown in Figure B.2b. It is possible to observe the 3 inlets for oil, cell and beads suspension and the outlet for collecting droplets. The geometrical structures in correspondence of the inlets are meant to distribute evenly cells and bead suspensions prior to enter

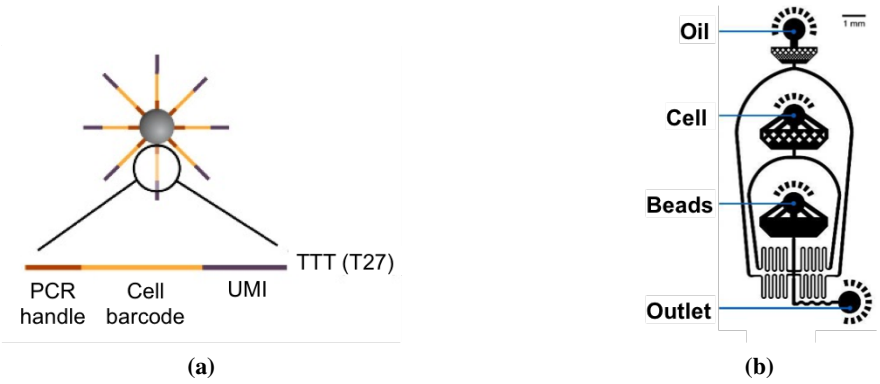


Figure B.2: **(a)** Sequence of primers on the barcoded primer bead; **(b)** pattern of Drop-seq microfluidic chip. Adapted from Macosko et al. (2015)

Table B.1: Operating conditions of Drop-seq experiments

Variable	Value
Oil flow rate	200 μl/min
Cell suspension flow rate	40 μl/min
Bead suspension flow rate	40 μl/min
Cell suspension concentration	620 cell/μl
Bead suspension concentration	300 bead/μl

into microchannels where they may clog and aggregate. Right after aqueous solutions co-flow, oil is added to form droplets. The serpentine channel before the outlet is used to mix the aqueous solutions contained in the droplets thanks to Dean vortices mixing.

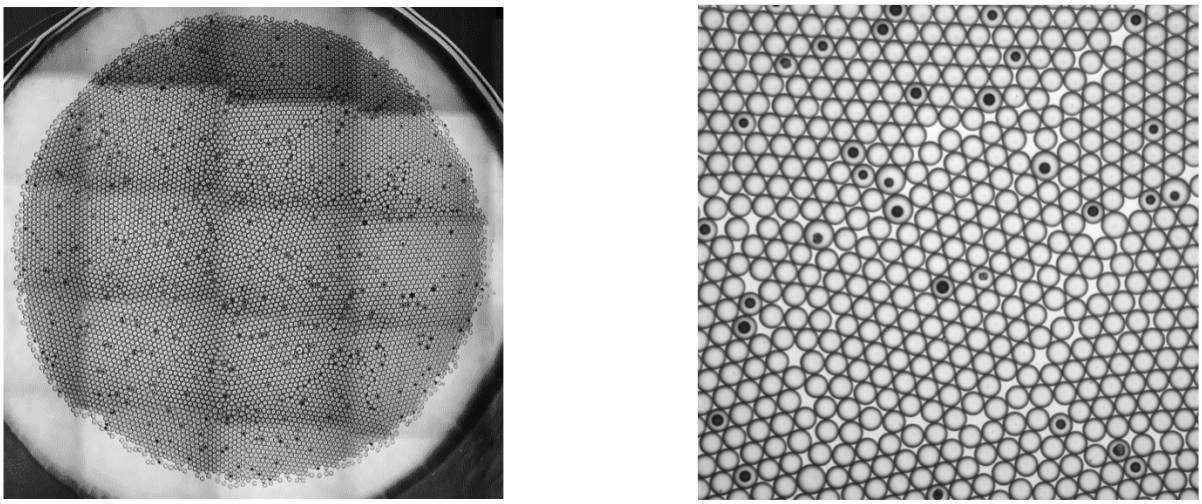
The operating parameters used for Drop-seq are reported in Table B.1; they have been modified compared to the published work to achieve higher efficiency (Section B.2).

B.2 Preliminary results

In this Section the preliminary results regarding bead encapsulation and STAMPs formation efficiency are presented.

B.2.1 Bead encapsulation efficiency

Preliminary experiments have been performed in order to test the efficiency of encapsulation of one single bead into a droplet. It is important to evaluate this efficiency and specially the percentage of droplet that contains more than one bead, which is one potential mode of failure. This experiment has been performed replacing cell suspension with the solution in which cells are normally resuspended. Figure B.3 shows images of beads encapsulated into droplets; the diameter of the droplets is 85 $\pm$ 1.5 μm.



(a) (b)

Figure B.3: Images showing beads encapsulated into droplets

The histogram in Figure B.4 shows the efficiency of encapsulation of one single bead into one droplet (5.5 ± 0.41 %) and the percentage of encapsulation of two beads or more (0.2 ± 0.08 %).

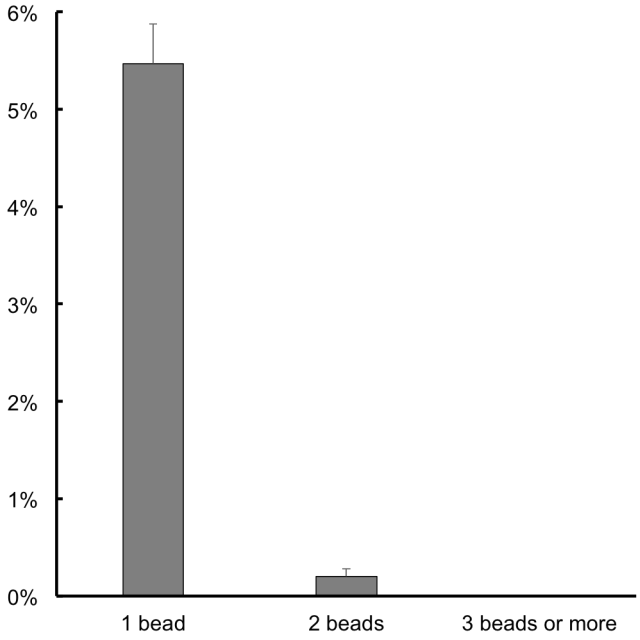
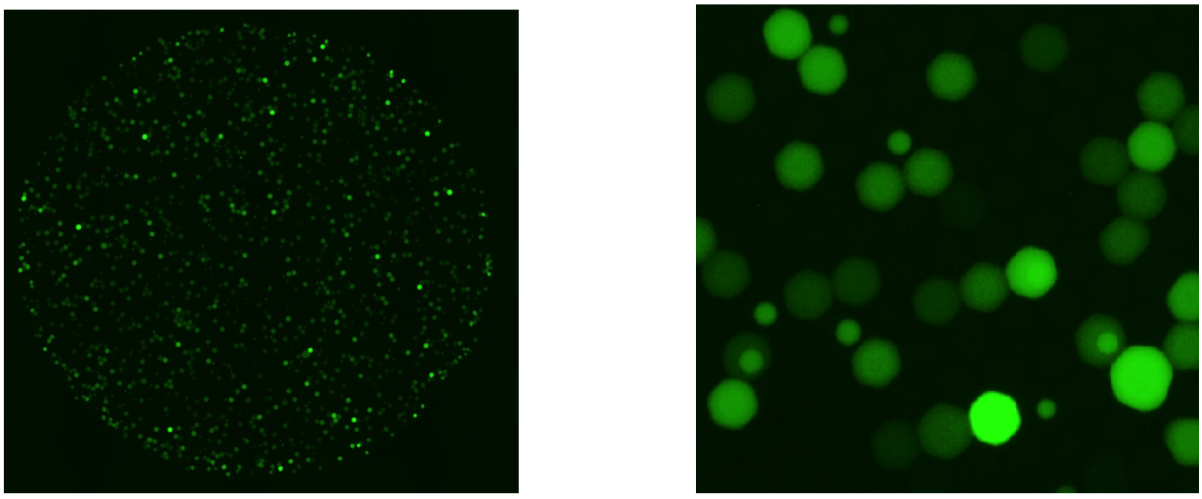


Figure B.4: Efficiency of bead encapsulation into droplets

B.2.2 STAMPs formation efficiency

The target event is the encapsulation of one bead and one cell into a droplet; this has been evaluated by marking cells with Calcein before performing a Drop-seq run since cells are lysed inside droplet and it is not possible to evaluate if and how many cells are present into a droplet by any other imaging means (note that beads have auto-fluorescence). The results is shown in Figure B.5.



(a) **(b)**

Figure B.5: Images showing lysed cells and beads encapsulated into droplets

The quantification of efficiency is shown in Figure B.6; the STAMPs formation efficiency is 0.5 $\pm$ 0.07 %. The STAMPs formation efficiency given by Macosko et al. (2015) is 0.26 %.

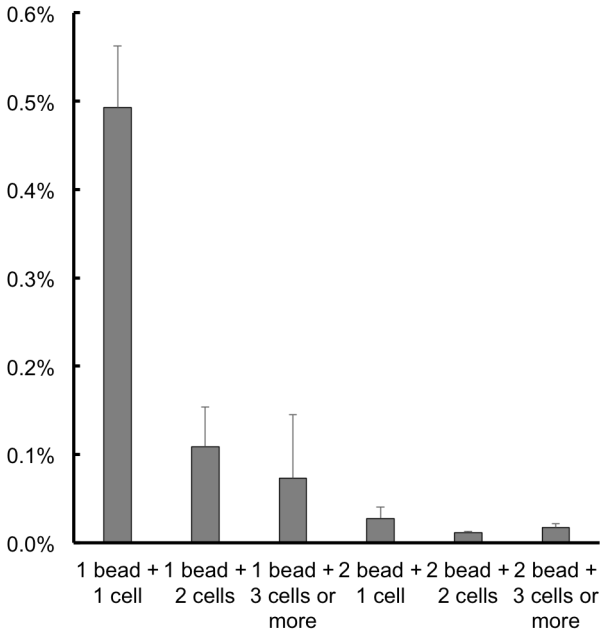


Figure B.6: Efficiency of STAMPs formation into droplets

Appendix C

RNA sequencing database

In this Appendix the database of published works using RNA sequencing for gene expression analysis over time is presented. Published works have been retrieved from publicly available databases (NCBI-SRA, EBI and DDBJ) until September 2017, found using the keyword "circadian" and filtering for datasets obtained by RNA-seq technology. Finally, only datasets obtained from murine and human samples are further characterized, since mouse is the model organism for *in vivo* studies of circadian rhythms, and human would be the final target to which circadian knowledge would be applied. The goal of this database is to give an overview of the available RNA-seq datasets, identifying which are the most sampled tissues and the most common decision for experimental as well as technical variables.

To achieve this goal, after selecting the datasets as explained above, specific information for each published work is collected: organism, tissue or cell line; sampling frequency; time interval of sampling; number of replicates; algorithm used for assessing circadian transcripts; stimulus used to synchronize the cell culture or the *in vivo* models; CT at which samples were collected; sequencing technical information.

The database regarding human samples is presented in Table C.1; the database regarding murine sample is presented from Table C.2 to Table C.6.

Table C.1: Database of published works using RNA-seq for gene expression analysis over time of human samples

Date	Journal	Paper	Organism, tissue, cell line	Sampling frequency	Interval	N replicates	Algorithm	Stimulus	Time points (TP)	Sequencing technical information
2015 Sept.	Cell metabolism	The pentose phosphate pathway regulates the circadian clock	U2OS Bmal1::Luc	4 h	48 h	3	RAIN, Fisher's G-test, ARSER	Dexamethasone shock (0.1 μM), sampling started 24 h after shock	CT24, CT28, CT32, CT36, CT40, CT44, CT48, CT52, CT56, CT60, CT64, CT68	Single-end 50-bp reads, 7M reads, HiSeq 2500
2015 Nov.	Science	Pancreatic β cell enhancers regulate rhythmic transcription of genes controlling insulin secretion	Pancreatic islet cells	4 h	24 h	3	JTK_CYCLE	1 h Forskolin shock (10 μM), ex-vivo synchronization, sampling started 40 h after shock	CT0, CT4, CT8, CT12, CT16, CT20	Paired-end 75-bp reads, at least 30M reads, NextSeq 500
2015 June	PNAS	Functional organization of the human 4D Nucleome	Human fibroblasts (CRL-2522)	8 h	56 h	3	JTK_CYCLE	1 h Dexamethasone shock (0.1 μM), sampling started after shock (before the shock, cells were kept in serum starvation for 48 h)	0h, 8h, 16h, 24h, 32h, 40h, 48h, 56h	Single-end 50-bp reads, extraction from 6-well plate per TP, HiSeq 2000
2015 Dec.	Genome Research	Ribosome profiling reveals an important role for translational control in circadian gene expression	U2OS PER2::LUC and U2OS treated with siRNAs for Bmal1	2 h	24 h	2	JTK_CYCLE	Dexamethasone shock (0.1 μM), sampling started 24 h after stimulus	CT0, CT2, CT4, CT6, CT8, CT10, CT12, CT14, CT16, CT18, CT20, CT22	Single-end 100-bp reads, 25M clean reads, HiSeq 2000

Table C.2: Database of published works using RNA-seq for gene expression analysis over time of murine samples (I)

Date	Journal	Paper	Organism, tissue, cell line	Sampling frequency	Interval	N replicates	Algorithm	Stimulus	Time points (TP)	Sequencing technical information
2017 Aug.	PNAS	Involvement of posttranscriptional regulation of Clock in the emergence of circadian clock oscillation during mouse development	Embrionic heart	4 h	44 h	1 (pool from 3-7 mice)	MetaCycle	LD12:12 conditions and were transferred to a constant dark condition for 36 h before sampling	CT0, CT4, CT8, CT12, CT16, CT20, CT24, CT28, CT32, CT36, CT40, CT44, CT48	Paired-end 100-bp reads, 38M clean reads, HiSeq 2500
2017 Aug.	Cell Reports	Time-Restricted Feeding Shifts the Skin Circadian Clock and Alters UVB-Induced DNA Damage	Back skin	4 h	1 day	1 (pool of 2-4 mice)	JTK_CYCLE, ARSER, Lomb-Scargle, sinusoid curve fitting	LD12:12, midday feeding (CT5-CT9); LD12:12, early day feeding (CT0-CT4); LD12:12, early night feeding (CT12-CT16)	CT0, CT4, CT8, CT12, CT16, CT20, CT24, CT28	35M clean reads,HiSeq 2500
2017 Feb.	Nature Medicine	Dissociation of muscle insulin sensitivity from exercise endurance in mice by HDAC3 depletion	WT and HDAC3-depleted muscles, muscle, tibialis anterior	/	/	3	/	LD12:12	CT10	HiSeq 2000
2017 Jan.	PLoS one	Normalisation against Circadian and Age-Related Disturbances Enables Robust Detection of Gene Expression Changes in Liver of Aged Mice	Liver	/	/	4	/	LD12:12	Pool of CT2, CT6, CT10, CT14, CT18 and CT22 at different ages	Forward sequences (35 bp), AB 5500xl Genetic Analyzer

Table C.3: Database of published works using RNA-seq for gene expression analysis over time of murine samples (II)

Date	Journal	Paper	Organism, tissue, cell line	Sampling frequency	Interval	N replicates	Algorithm	Stimulus	Time points (TP)	Sequencing technical information
2016 Dec.	Cell	Microbiota Diurnal Rhythmicity Programs Host Transcriptome Oscillations	Intestinal Epithelium	6 h	2 days	2	JTK_CYCLE	LD12:12	CT12, CT18, CT0, CT6, CT12, CT18, CT0, CT6, CT12	NextSeq 500
2016 Nov.	Cell	CRY2 and FBXL3 Cooperatively Degrade c-MYC	Primary WT and Cry2-/- MEFs	4 h	1 day	4	/	Dexamethasone 21-phosphate sodium salt (1 M)	CT0, CT4, CT8, CT12, CT16, CT20	Single-end 100 bp, reads, 6M clean reads, HISeq Analyzer 2000
2016 Oct.	Journal of the American Society of Nephrology	Nephron-Specific Deletion of Circadian Clock Gene Bmal1 Alters the Plasma and Renal Metabolome and Impairs Drug Disposition	WT and Bmal1 KO in kidney, kidney, renal tubular cells	/	/	3	/	LD12:12 food ad libitum	CT4, CT16	30M clean reads, HiSeq 2000
2016 May	PLoS genetics	Long-Range Chromosome Interactions Mediated by Cohesin Shape Circadian Gene Expression	WT and Smc3-/- MEF	/	/	1	/	1 h Dexamethasone shock (0.1 μM), then medium serum-free	CT20, CT32	100 bp read, HiSeq 2000

Table C.4: Database of published works using RNA-seq for gene expression analysis over time of murine samples (III)

Date	Journal	Paper	Organism, tissue, cell line	Sampling frequency	Interval	N replicates	Algorithm	Stimulus	Time points (TP)	Sequencing technical information
2016 Feb.	Scientific Reports	Oscillating primary transcripts harbor miRNAs with circadian functions	Liver WT and miR-378 over-expression	/	/	2	/	LD12:12, food ad libitum. Sampling started after 48 h in darkness	CT10, CT22	/
2016 Feb.	Molecular Cell Biology	Identifying Novel Transcriptional Regulators with Circadian Expression	NIH3T3	4 h	32 h	2	JTK_CYCLE, RAIN, Arser, CBNLR	Dexamethasone shock (0.1 μM)	0h, 4h, 8h, 12h, 16h, 20h, 24h, 28h, 32h	50-bp reads, poly(A) RNA sequencing, nonstrand specific, 33M clean reads, Illumina HiSeq 2000
2015 Dec.	Genome Research	Ribosome profiling reveals the rhythmic liver translatome and circadian clock regulation by upstream open reading frames	Liver	2 h	24 h	2 (pool of 2 livers)	AIC (custom)	LD12:12, food ad libitum for 2 weeks, sampling done in LD12:12	CT0, CT2, CT4, CT6, CT8, CT10, CT12, CT14, CT16, CT18, CT20, CT22	21–60 bp, 23M clean reads, Illumina HiSeq 2500
2015 Dec.	Scientific Reports	Circadian and light-driven regulation of rod dark adaptation	Eye (posterior sclera, choroid, RPE, and retina)	/	/	2	/	/	1	50 bp, HiSeq 2500 sequencer

Table C.5: Database of published works using RNA-seq for gene expression analysis over time of murine samples (IV)

Date	Journal	Paper	Organism, tissue, cell line	Sampling frequency	Interval	N replicates	Algorithm	Stimulus	Time points (TP)	Sequencing technical information
2015 Dec.	Cell reports	KLF15 Establishes the Landscape of Diurnal Expression in the Heart	WT and KLF15 KO, Heart (apex)	4 h	24 h	4	JTK_CYCLE	LD12:12	CT2, CT6, CT10, CT14, CT18, CT22	Paired-end 100 bp reads, Illumina HiSeq 2500
2015 Nov.	PNAS	Circadian and feeding rhythms differentially affect rhythmic mRNA transcription and translation in mouse liver	Liver WT and Bmal1 KO in Restricted Feeding and Ad Libitum feeding	2 hours	1 day	4	Harmonic regression	LD12:12, food ad libitum	CT0, CT2, CT4, CT6, CT8, CT10, CT12, CT14, CT16, CT18, CT20, CT22	Paired-end
2015 Nov.	Science	Pancreatic β cell enhancers regulate rhythmic transcription of genes controlling insulin secretion	Pancreatic islets β cells	4 h	2 days	3	JTK_CYCLE	Forskolin shock (ex vivo synchronization), sampling started 40 h after shock	CT0, CT4, CT8, CT12, CT16, CT20, CT24, CT28, CT32, CT36, CT40, CT44	Paired-end 100-bp reads, at least 30M reads, HiSeq 2000
2015 July	Cell	The Regulatory Factor ZFHX3 Modifies Circadian Function in SCN via an AT Motif-Driven Axis	SCN	/	1 day	3	/	LD12:12 food ad libitum	CT3, CT15	Paired-end 50-bp reads,HiSeq2000

Table C.6: Database of published works using RNA-seq for gene expression analysis over time of murine samples (V)

Date	Journal	Paper	Organism, tissue, cell line	Sampling frequency	Interval	N replicates	Algorithm	Stimulus	Time points (TP)	Sequencing technical information
2014 Nov.	PNAS	A circadian gene expression atlas in mammals: implications for biology and medicine	Adrenal gland, aorta, brainstem, brown fat, cerebellum, heart, hypothalamus, kidney, liver, lung, skeletal muscle, and white fat	6 h	2 days	1 (pool from 3 mice)	JTK_CYCLE	LD12:12 for 1 week, then DD, starting at CT18 with samples	CT22, CT28, CT34, CT40, CT46, CT52, CT58, CT64	Paired-end 100-bp reads, 30M clean reads, Illumina HiSeq 2000
2014 May	Elife	MicroRNAs shape circadian hepatic gene expression on a transcriptome-wide scale	Dicer KO and WT Liver	4 h	1 day	1 (pool from 3-4 mice)	JTK_CYCLE and harmonic regression	LD12:12, food ad libitum for 1 month	CT0, CT4, CT8, CT12, CT16, CT20	Single-end 100-bp reads, Illumina HiSeq 2500
2012 Nov.	Elife	Nascent-Seq reveals novel features of mouse circadian transcriptional regulation	WT and Bmal 1 KO Liver	4 h	1 day	2	Custom	LD12:12	CT2, CT6, CT10, CT14, CT18, CT22	101bp, 25M clean reads, HiSeq2000
2012 Oct.	Science	Transcriptional architecture and chromatin landscape of the core circadian clock in mammals	Liver	4 h	2 days	1 (pool from 3 mice)	COSOPT, JTK_CYCLE,ARSER	LD12:12 for 1 week, then constant darkness, first sample after 36h	CT0, CT4, CT8, CT12, CT16, CT20, CT24, CT28, CT32, CT36, CT40, CT44	Paired-end, 55M total reads, AB SOLiD 4 System

Appendix D

Additional graphs of PCA model

This Appendix contains additional material of PCA model of whole transcriptome data.

D.1 Circadian genes of a single replicate

Figure D.1 shows the scree plot for PCA model build on circadian genes of one replicate after SNV. According to the eigenvalue-greater-than-one rule, only the first two PCs are selected for the PCA model.

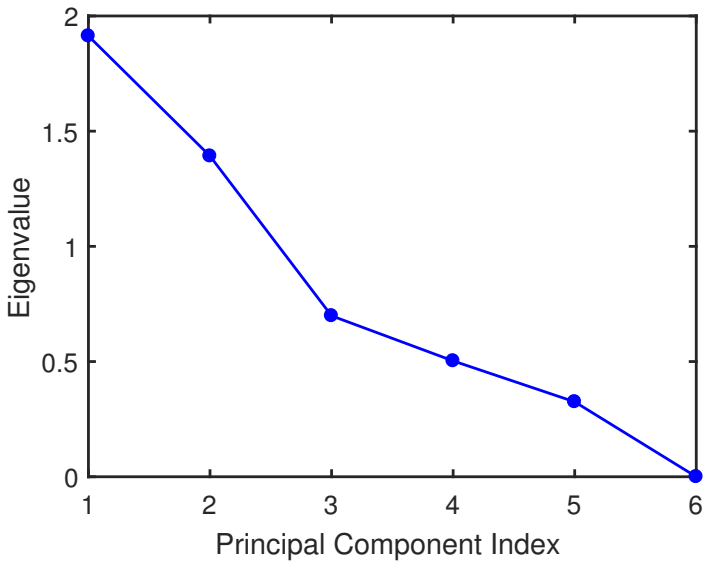


Figure D.1: Scree plot for PCA model build on circadian genes of one replicate after SNV

D.2 Normalization

Figure D.2 shows the error distribution for normalized dataset; the distributions are normal for all time points, similar to what is observed for SNV dataset.

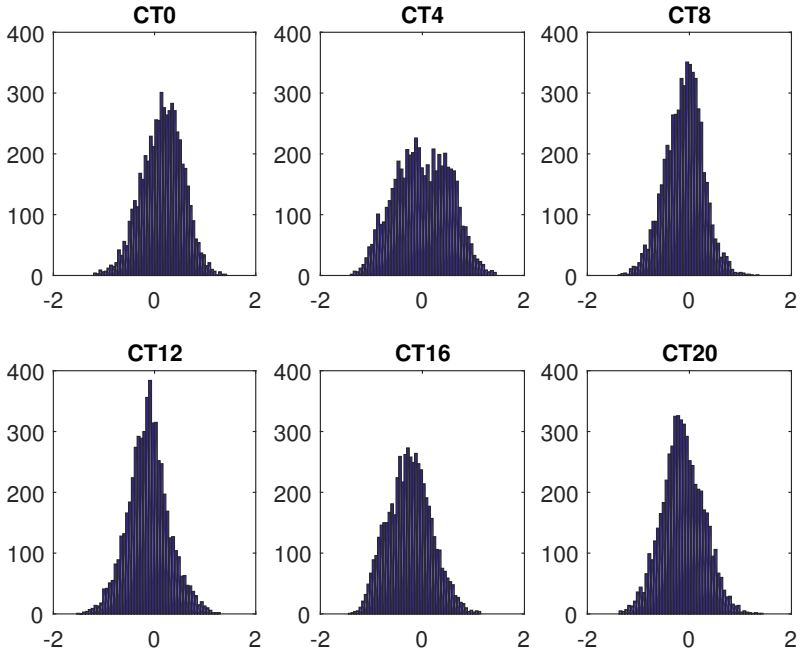


Figure D.2: Error distributions for circadian genes at the different time points for the first replicate (y-axis indicates the number of genes, x-axis indicates magnitude of the error)

FigureD.3 shows SSR distribution for all replicates. For all replicates, a chi-squared distribution can be observed: since this distribution is generated by the sum of independent, standard normal random variables, it can be concluded that all replicates have normal distribution of error for all time points, as for the first replicate as shown in Figure D.2 .

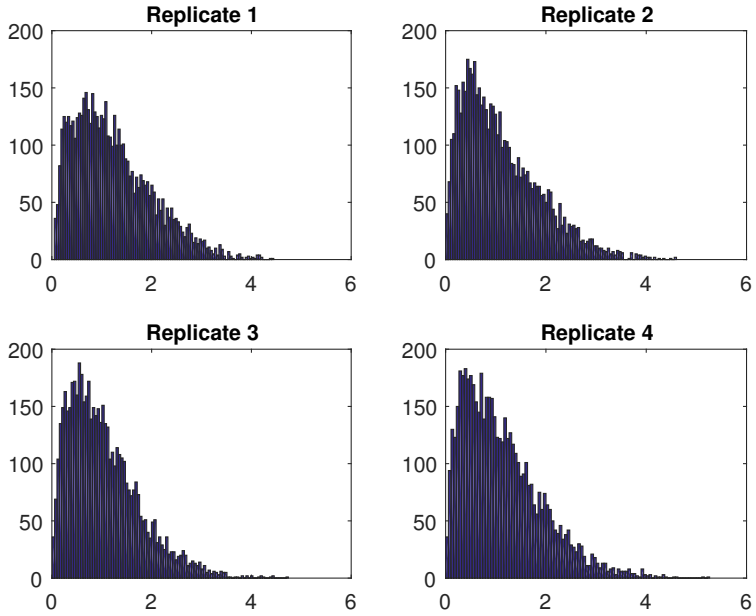


Figure D.3: Distribution of sum of squared residuals of circadian genes calculated for all replicates based on the first two PCs (y-axis indicates the number of genes, x-axis indicates magnitude of the sum of squared residuals)

D.3 Whole transcriptome of a single replicate

Figure D.4 shows the scree plot for PCA model build on whole transcriptome of one replicate after SNV. Only the first two PCs are selected for the PCA model, even though the third one has a value slightly above 1.

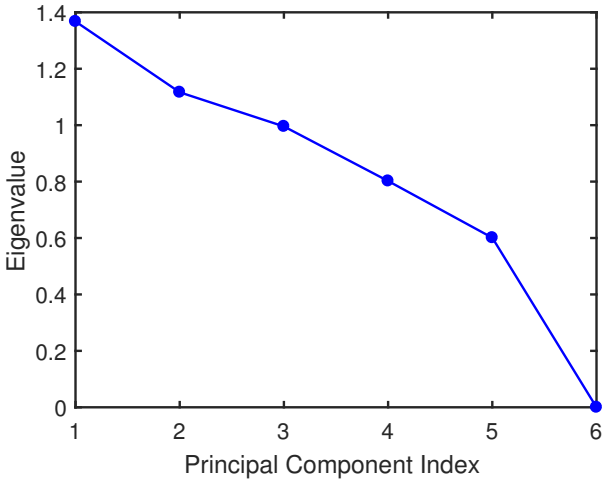
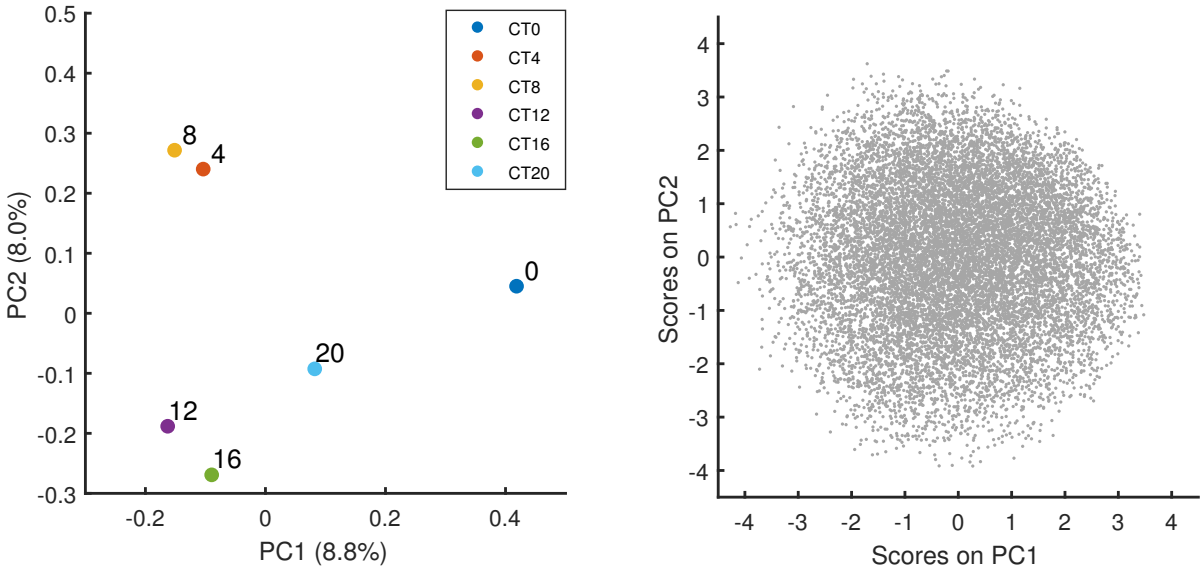


Figure D.4: Scree plot for PCA model build on all genes of one replicate after SNV

D.4 Non-circadian genes of multiple replicates

Figure D.5 shows PCA model parameters calculated on non-circadian genes of all replicates treated by hybrid unfolding. In Figure D.5a it is possible to see that the loadings are not arranged following a neat temporal pattern, meaning that there is not a clear correlation between variables.

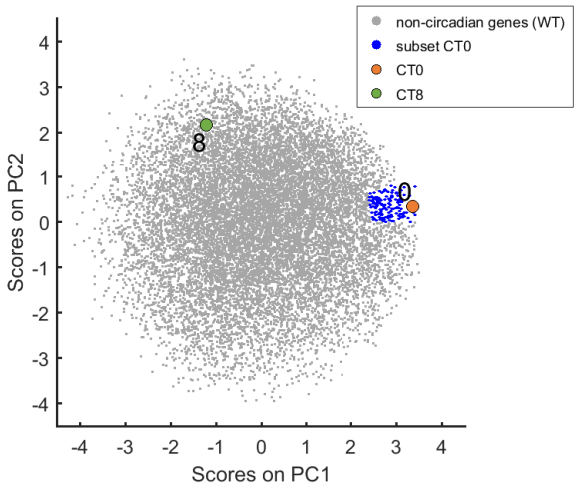


(a) Loading plot

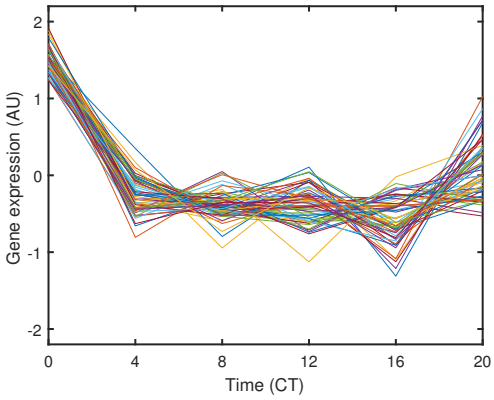
(b) Score plot

Figure D.5: Loading plot and score plot of PCA model calculated on non-circadian genes of all replicates treated by hybrid unfolding and after SNV

Figure D.6 and Figure D.7 show the temporal patterns of groups of genes located in correspondence to CT0 and CT8; it is possible to see that they do not have a clear circadian pattern and, if they may have it, it is comparable to an ultradian rhythm.

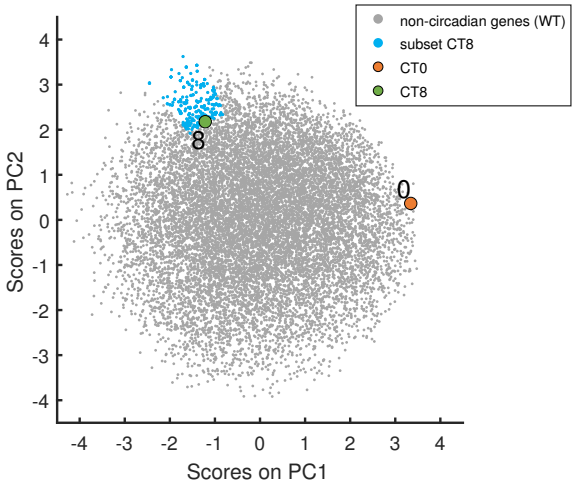


(a) Score plot of PCA model calculated on non-circadian genes of all replicates treated by hybrid unfolding and after SNV, with highlighted a group of 61 genes located in correspondence of CT0

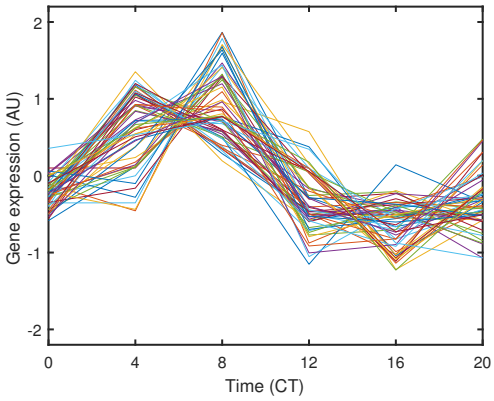


(b) Gene expression over time of 61 genes highlighted in Figure D.6b. Gene expression corresponds to average of the SNV expression of the four replicates

Figure D.6: Score plot and gene expression profile over time of non-circadian genes located in correspondence of CT0



(a) Score plot of PCA model calculated on non-circadian genes of all replicates treated by hybrid unfolding and after SNV, with highlighted a group of 56 genes located in correspondence of CT8



(b) Gene expression over time of 56 genes highlighted in Figure D.7a. Gene expression corresponds to average of the SNV expression of the four replicates

Figure D.7: Score plot and gene expression profile over time of non-circadian genes located in correspondence of CT8

Appendix E

Biological protocols

This Appendix describes in detail the biological protocols that have been used.

E.1 Cell culture

Murine fibroblasts PER2::LUC are expanded in 10 mm Petri dishes in high-glucose Dulbecco Modified Eagle Medium (DMEM 41965, Thermo Fisher Scientific) at 90% in volume and 10% of Fetal Bovine Serum (FBS, Thermo Fisher Scientific). Cells are maintained in humidified 5% CO2 incubator at 37 °C. Culture medium is replaced every 3 days. When cells reach confluence, they are trypsinized with Trypsin-EDTA 0.25% (Life Technology) and replated either in new dishes or into microfluidic chip or they are thawed.

Before cell seeding, monolayer microfluidic chip was sterilized by autoclave processing; after its sterilization, it is moved into a 35 mm Petri dish under sterile hood and injected with cold Fibronectin at 25 μg/ml (Santa Cruz Biotechnology) and kept at room temperature for 30 min. Cell suspension is then prepared in order to have a cell seeding of 700 cell/mm^2 and injected into the microfluidic channels[1]. The bottom of the dish is covered by PBS 1X –, to maintain proper humidity. For bioluminescence experiments, cells are seeded 16 h prior to the first stimuli of the stimulation protocol and then moved to the dark acquisition chamber of the luminometer. For other experiments, cells are kept in the incubator and medium is replaced every 12 h by manual pipetting.

As far as multilayer platform is concerned, cell seeding is performed in a different way compared to monolayer platform. The first step consists in filling seven tubes (PTFE, ID 0.032”, OD 0.056”, Cole

[1] After calculating the absolute number of cells in the cell suspension, the resuspension volume V_{res} in which the cells should be resuspended in order to reach the desired cell surface density is

$$V_{res} = \frac{N_{cells} \cdot h_{channel}}{\rho}$$

where N_{cells} is the absolute number of cells in the cell suspension, $h_{channel}$ is the height of the microfluidic channel and ρ is the desired surface density of cells.

Parmer) connected to seven independent solenoid valves with Milli-Q water containing 1% of Penicillin Streptomycin (Thermo Fisher Scientific) to avoid formation of fungi or molds; then they are inserted in the seven inlets of the control layer of the platform. Then, the minimum closing pressure for all on-chip valves is determined. After this, the chip is placed under the sterile hood and sterilization procedure starts (multilayer platforms are not sterilized in autoclave since high temperature makes valve membrane very fragile): both sides of the chip are exposed to UV light for 20 min. Then the chip is placed in a Petri dish and coating with Fibronectin is performed by filling a pressurized vial with Fibronectin and injecting it into the channels and chambers of the microfluidic platform, from the outlet hole of the platform. The flow pressure is set at 5 psi. During this step (called debubbling) all air contained in chambers and channels is let to permeate through PDMS, in order to wet all surfaces with coating (it takes at least 30 min). When all bubbles have been removed, seeding is performed putting cell suspension inside a pressurized vial and injecting it in one group of chambers at a time for 15 s (the minimum required volume for cell suspension for seeding all chambers is 400 μl). In this case the cell culture medium contains also Penicillin Streptomycin, to prevent and reduce bacterial contaminations. After seeding, two vials containing cell culture media are connected to the inlets of the chip, and a tube is connected to the outlet of the platform. Then the platform is placed in an incubator and kept for 7 days with automatic media changes every 12 h.

E.2 LIVE/DEAD Cell viability assay

Cell viability assay is performed using a commercial kit (LIVE/DEAD Viability/Citotoxicity Kit, Invitrogen). Cells are stained using a staining solution which contains:

- HOECHST 33342: used in volume ratio 1:1000, it is a blue fluorescent dye which marks cell nuclei (excitation wavelength is equal to 361 nm while the emission one is equal to 461 nm)
- Calcein: used in volume ratio of 1:4000, it is a green fluorescent dye which marks the living cell cytoplasm (excitation wavelength is equal to 495 nm while the emission one is equal to 515 nm)
- Ethidium bromide: used in volume ratio of 1:500, it is a red fluorescent dye which marks cells in the final stages of apoptosis (excitation wavelength is equal to 526 nm while the emission one is equal to 605 nm).

After a wash with PBS 1X –, cells are stained with the staining solution (three markers are diluted in DMEM); cells are then incubated for 45 min at 37 °C in dark. Prior to imaging, microfluidic channels are washed with PBS 1X ++ to remove the staining solution.

E.3 Bioluminescence assay

Bioluminescence acquisition is performed into the dark acquisition chamber of the luminometer (Xenogen IVIS Imaging System 100 Series). The chamber is not provided with CO_2 supply, therefore the medium needs to have a specific composition to keep the pH of the medium in the suitable range for cells. Table E.1 details components present in bioluminescence recording medium of all experiments.

Table E.1: Composition of recording medium for bioluminescence acquisitions

Component	Concentration	Volume fraction
FBS	-	1 %
B27 supplement (- Insulin)	-	2 %
Penicillin Streptomycin	-	1 %
$NaHCO_3$	0.089 mM	-
Sodium Pyruvate	1 mM	-
HEPES	10 mM	-
L-Glutamine	4 mM	-
Luciferin	1 mM	-

FBS and B27 are added to the medium to keep cells viable, without making cells to enter cell cycle from G0 phase (since cells are confluent, they are blocked in G0 phase, so they do not replicate). Preliminary experiments show that Luciferin concentration at 1 mM is not limiting the reaction rate. $NaHCO_3$, Sodium Pyruvate and HEPES are added as a buffer to control medium pH.

E.4 Experiments with inhibitors

PER2::LUC fibroblasts are seeded at 100% confluence in 96-well tissue culture-treated polystyrene microplate with black walls and clear bottom in DMEM and 10% FBS. 16 hours after seeding, cell pre-treatment begin after washing with PBS 1X –. The pre-treatment is performed in DMEM with low glucose content (2 mM) and 3% B-27 Supplement minus insulin (Life Technologies). The pre-treatment allows for completion of three entire circadian periods. After pre-treatment and a PBS 1X – wash, cells are treated with inhibitors which are added to DMEM with 100 nM insulin (Sigma-Aldrich), 1% FBS, 2% B27 Supplement minus insulin and 1 mM D-Luciferin. The tested inhibitors are PI-103 (Santa Cruz) at 100 nM, Rapamycin (Sigma-Aldrich) at 100 nM, StemMACS CHIR99021 (Miltenyi Biotec) at 20 μM, Resveratrol (Sigma Aldrich) at 100 μM and Tumor Necrosis Factor-α (Sigma-Aldrich) at 20 ng/mL. After stimulations with inhibitors, bioluminescence recording is performed for four days in the temperature

controlled acquisition chamber of Xenogen IVIS Imaging System 100 Series.

Appendix F

Simvastatin Rapidly and Reversibly Inhibits Insulin Secretion in Intact Single-Islet Cultures

Valentina Scattolini[1,2], Camilla Luni[2,3,4], Alessandro Zambon[2,4], Silvia Galvanin[2,4], Onelia Gagliano[2,4], Catalin Dacian Ciubotaru[5], Angelo Avogaro[1], Fabio Mammano[5,6], Nicola Elvassore[2,4] and Gian Paolo Fadini[1,2]

[1] Department of Medicine, University of Padova, Via Giustiniani 2, 35129 Padua, Italy

[2] Venetian Institute of Molecular Medicine, Via Orus 2, 35128 Padua, Italy

[3] Shanghai Institute for Advanced Immunochemical Studies, ShanghaiTech University, 99 Haike Road, Shanghai 201210, China

[4] Department of Industrial Engineering, University of Padova, Via Marzolo 9, 35131 Padua, Italy

[5] CNR Institute of Cell Biology and Neurobiology, 00015 Monterotondo, Italy

[6] Department of Physics, University of Padova, Via Marzolo 8, 35131 Padua, Italy

Diabetes Therapy

Volume 7(4): 679-693

10.1007/s13300-016-0210-y

F.1 Abstract

Introduction Epidemiological studies suggest that statins may promote the development or exacerbation of diabetes, but whether this occurs through inhibition of insulin secretion is unclear. This lack of understanding is partly due to the cellular models used to explore this phenomenon (cell lines or pooled islets), which are non-physiologic and have limited clinical transferability.

Methods Here, we study the effect of simvastatin on insulin secretion using single-islet cultures, an optimal compromise between biological observability and physiologic fidelity. We develop and validate a microfluidic device to study single-islet function ex vivo, which allows for switching between media of different compositions with a resolution of seconds. In parallel, fluorescence imaging provides real-time analysis of the membrane voltage potential, cytosolic Ca^{2+} dynamics, and insulin release during perfusion under 3 or 11 mM glucose.

Results We found that simvastatin reversibly inhibits insulin secretion, even in high-glucose. This phenomenon is very rapid (<60 s), occurs without affecting Ca^{2+} concentrations, and is likely unrelated to cholesterol biosynthesis and protein isoprenylation, which occur on a time span of hours.

Conclusions Our data provide the first real-time live demonstration that a statin inhibits insulin secretion in intact islets and that single islets respond differently from cell lines on a short time scale.

F.2 Introduction

Statins, hydroxy-methyl-glutaryl-CoA (HMG-CoA) reductase inhibitors, are widely used to lower cholesterol and prevent cardiovascular disease [1, 2]. As diabetes is characterized by high cardiovascular risk, international guidelines recommend that most diabetic patients should receive statin therapy [3–5]. Statins effectively reduce cardiovascular morbidity and mortality in patients with diabetes [6], but epidemiological studies suggest they increase the risk of developing diabetes in healthy individuals [7–10] and may worsen glycemia in diabetic patients [11]. Pathophysiologic studies in humans indicate that statins may affect both insulin sensitivity and insulin secretion [12]. In vitro studies show that statins impair insulin secretion by affecting multiple pathways [13], including cellular cholesterol synthesis, membrane fluidity, and isoprenylation of proteins [14]. Though their cardiovascular benefit exceeds the diabetes hazard [15], the mechanisms whereby statins accelerate the onset of diabetes remain unclear, and whether statins truly exert any action on insulin secretion is debated [16, 17]. This is in part because cellular models used to explore this phenomenon, either cell lines [13, 14] or pooled islets [14], are non-physiologic and lack clinical transferability. Cell lines may behave differently from mature beta

cells or have incomplete glucose sensor or insulin secretion machinery, whereas pooled islet cultures are poorly suitable to provide information on dynamic insulin secretion because islet response may be heterogeneous and asynchronous.

A single islet is the minimum fully functional unit of the endocrine pancreas. A precise understanding of islet response to exogenous stimuli is better captured in vitro at the single-islet level, devoid of complex and not fully understood islet–islet interactions. Microfluidics is a technology suitable for cell culture at micrometer scale. Because a single islet has an approximate size of 50–500μm, it is particularly suitable for microfluidic culture. Accordingly, a number of microfluidic devices have been developed for understanding single islet behaviour, as recently reviewed [18]. However the advantages of microfluidic technology go well beyond the size of these devices [19]. For example, Benninger et al. used a microfluidic device to generate precise glucose concentration gradients in a single islet microenvironment to study cell–cell gap junctions in the regulation of coordinated insulin release [20]. Microfluidic devices favor the accumulation of endogenously secreted factors and were previously used to detect islet amyloid polypeptide secretion from as few as ten islets [21] and glucose flux in live myoblasts [22], besides insulin [23].

Here, we study the effects of simvastatin on insulin secretion from single murine islets. A microfluidic system able to temporally control the single-islet soluble microenvironment is presented. The setup has been technically and biologically validated, before analyzing simvastatin action on intact isolated islets.

F.3 Methods

F.3.1 Microfluidic setup

The setup was produced by standard photo- and soft-lithographic techniques. Details on the fabrication of the islet culture and the multi-inlet chips are given in the Electronic Supplementary Material (ESM).

F.3.2 Animals and Pancreatic Islet Extraction and Preparation

The protocol followed the Guide for the Care and Use of Laboratory Animals, eighth edition (2011) (http://grants.nih.gov/grants/ olaw/guide-for-the-care-and-use-of-laboratory- animals.pdf), and was compliant with local and national laws. For murine islet isolation, we optimized a previously described protocol [24]. Details are given in the ESM.

F.3.3 Microfluidic System Running Conditions

The multi-inlet microfluidic system was connected to medium vials of different compositions. Details are given in ESM. This system allowed switching between up to nine different medium compositions by remotely controlling valve opening and closing. The complete system is displayed in Figure F.1.

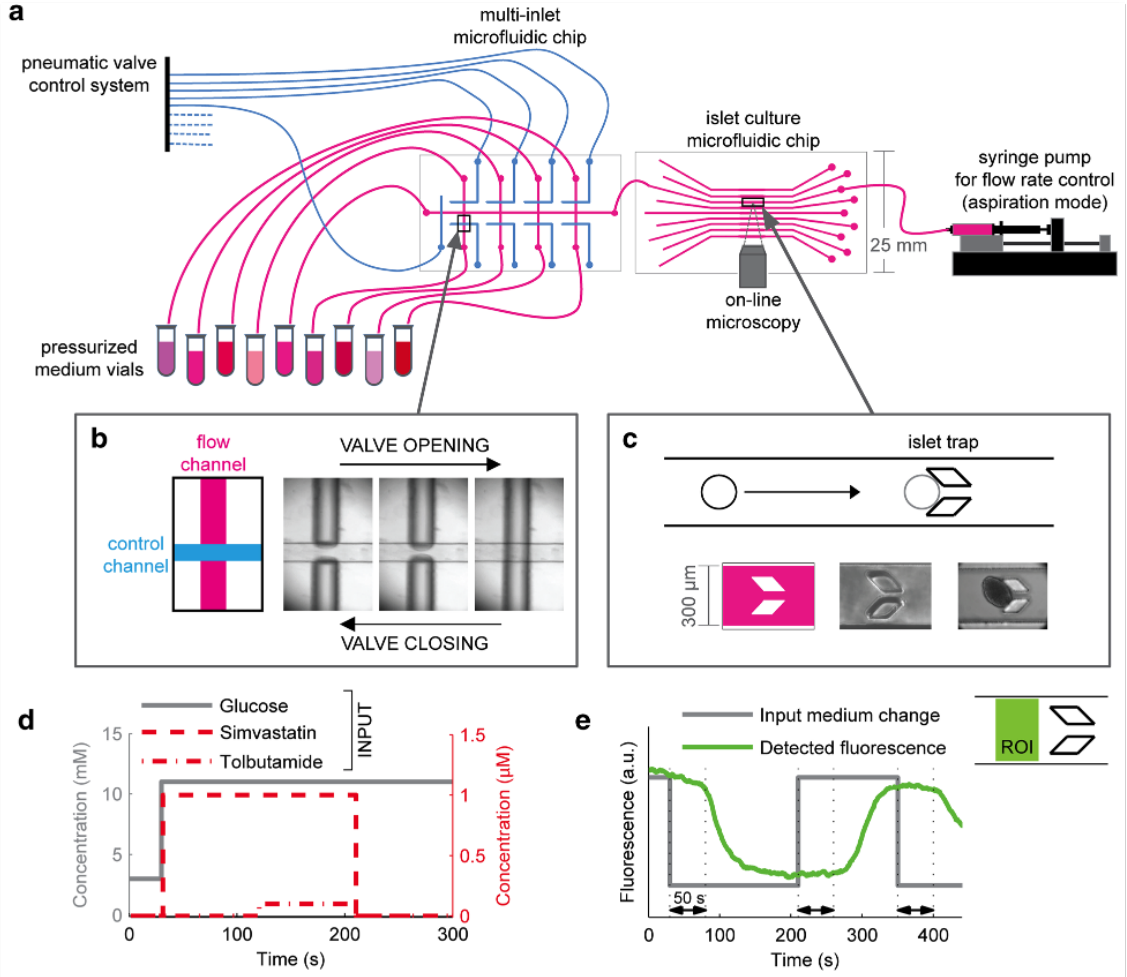


Figure F.1: Microfluidic setup design and validation. **a** Full setup for single islet microscopy observation under controlled dynamic biochemical stimulation. A multi-inlet microfluidic chip contains a system of pneumatic valves that are used to temporally control the medium composition entering the islet culture chamber. The pneumatic valve system is actuated by means of electrovalves (not shown) and is responsible for each valve opening and closing according to the mechanism shown in **b**: pressurizing the control channel prevents medium flow through the flow channel. A single islet is placed in each channel of the islet culture microfluidic chip, where it is precisely positioned because of the presence of a mechanical trap that stops its flow along the channel, as shown in **c**. The islet culture microfluidic chip is a transparent device that is put under the microscope for online observation. Once the specific valve is open, medium flows through the channel because it is contained in a pressurized vial; to accurately control its flow rate, a syringe pump working in aspiration is placed downstream of the culture system. **d** The system described in **a-c** makes possible the biochemical stimula- tion of a single-islet with temporal accuracy of seconds. Medium composition is changed over time with minimal external perturbation. **e** Validation of the dynamic input system using medium with and without fluorescein. The theoretical curve (red) is compared with the actual medium fluorescence (green) detected in the region of interest (ROI) indicated. A delay of approximately 50s was obtained because of the distance between the valve system and the ROI

F.3.4 Islet Staining for Imaging of Viability, Ca^{2+}, Membrane Potential, and Zn^{2+}

Islets were loaded with 1 mg/ml Hoechst 33342 (Sigma-Aldrich) for 10 min. PE-conjugated Annexin-V (BD Biosciences) was added in the perfusion medium. Prior to microfluidic device insertion, islets were loaded with 3-μM Fura Red acetoxymethyl (AM) (ThermoFisher) or 10 μM Fluo-4 AM (ThermoFisher) for 50 min in RPMI medium at 37 °C for intracellular Ca^{2+} detection; 2-μM FluoZin-3 (ThermoFisher) was added at the same concentration to the different media used for islet stimulation. Extracellular Zn^{2+} concentration detection was used as an indirect measure for insulin secretion detection, as previously described [21]. To avoid biases due to slight differences in the FluoZin-3 concentration or media autofluorescence, before islet insertion into the microfluidic device, a background profile using different media was established at the beginning of each acquisition and then subtracted to the islet response profile. For voltage imaging, islets were incubated for 20min at 37 °C in RPMI with Vf2.1.Cl (200 nM, kindly provided by Roger Y. Tsien from University of California at San Diego) [25] and Pluronic F–127 (0.1% w/v) (Sigma-Aldrich).

F.3.5 Confocal Imaging

Fluorophores were excited by light from a 460-nm LED (M470L2, Thorlabs) passing through a T510LPXR filter (Olympus) and directed onto the sample through a full mirror (Chroma) while 512 fluorescence emission was selected by a BA495-540HQ filter (Chroma). All fluorescence images were acquired with a 209 water immersion objective (NA 0.96, Olympus) and projected on a Complementary Metal-Oxide Semiconductor (CMOS) cooled camera (PCO), controlled by in-house developed software. Image sequences of fluorescence were acquired continuously at three frames per second with 100 ms exposure time.

F.3.6 Two-Photon Imaging

Two-photon acquisitions were performed with a modular multiphoton microscope (Bergamo II, Thorlabs) equipped with an 8-kHz resonant scanner, extended-field-of-view collection optics, four independent detection channels in the backward direction, and laser-scanned Dodt contrast in the forward direction. The microscope was coupled to two synchronized pulsed laser beams generated by a Ti:sapphire pump laser (Chameleon Ultra 2, Coherent) and an optical parametric oscillator (Camelion Compact OPO). Two-photon microscopy at 800nm excitation was used to visualize the intracellular calcium (Fluo-4, green), nuclei (Hoechst 33342, blue), and islet structure simultaneously (transmitted light, gray).

F.3.7 Data Analysis

The analysis software, developed in-house in Matlab environment (Release 14, The MathWorks, Inc., Natick, MA, USA), was designed to plot averaged time responses from specific regions of interest (ROIs) as well as to create false-color animations reporting time-dependent fluorescence intensity variations. False-color images were generated using the hue-saturation-value (HSV) visualization algorithm: hue was used to represent fluorescence changes; saturation was set to 1.0; the value (brightness) carried pixel intensity from a reference image that was either updated on a frame-by-frame basis or obtained as an average over a specified number of frames. The resulting frames so constructed were converted to ordinary RGB images and exported in the Audio Video Interlaced (AVI) proprietary video format of Microsoft Video for Windows applications [26]. FluoZin-3 and Fura Red signals were analyzed independently by playing back images in single-channel mode. ROIs used for Fluo-4 and Fura Red signal analysis were drawn covering the islets, while ROIs for FluoZin-3 signal analysis were drawn outside the islet areas. When Fura Red is excited at 488 nm, its fluorescence emission decreases as Ca^{2+} concentration increases [27]. For this reason, changes in intracellular Ca^{2+} were expressed as (F_0/F) - 1 ratios [27]. Fluo-4 and FluoZin-3 signals were measured as relative changes of fluorescence emission intensity: $\Delta F/F_0$, where F_0 is the pre-stimulus basal fluorescence; $\Delta F = F - F_0$, F is fluorescence at time t.

F.3.8 Compliance with Ethics Guidelines

The institutional review board of the Venetian Institute of Molecular Medicine approved the study. The protocol followed the Guide for the Care and Use of Laboratory Animals, eighth edition (2011) (http://grants.nih.gov/grants/ olaw/guide-for-the-care-and-use-of-laboratory- animals.pdf). All institutional and national guidelines for the care and use of laboratory animals were followed.

F.4 Results

F.4.1 Microfluidic Single-Islet Culture Setup

The precise dynamic perturbation of a single-islet environment with the online measurement of its response is prevented in conventional culture systems, but can be robustly achieved using microfluidic devices. To this aim, we designed a microfluidic system where a single islet can be trapped without mechanical damage in a defined position, dynamically stimulated with media of multiple compositions, and contemporarily observed under a microscope (Figure F.1a–c). Each islet was cultured under constant flow perfusion, a condition that better mimics the highly vascularized environment where islets reside in vivo.

The microfluidic device includes an integrated pneumatic valve system (Figure F.1b) that is remotely controlled and enables switching the medium entering the culture chamber in few seconds, while minimizing other culture perturbations (Figure F.1d). The time required for the new medium to reach the islet position within the channel was estimated using buffer with or without fluorescein. At the 2μl/min flow rate used in all the experiments, there was an approximate 50 s lag between the medium valve switch and medium change at the islet position (Figure F.1e).

The size and transparency of the microfluidic device make possible the online microscopic observation of the islet response. Overall, the described setup provides maximal information in terms of the spatial and temporal resolution of the collected data.

F.4.2 Biological Validation of the Microfluidic Device

The canonical glucose-stimulated insulin secretion (GSIS) pathway in the β-cell fraction of pancreatic islets involves some well-established sequential steps: glucose is transported into β-cells through plasma membrane glucose transporters and is metabolized, thereby elevating the intracellular ATP/ADP ratio; this event causes the closure of cell-surface ATP-sensitive K^+ channels, leading to cell membrane depolarization; plasma membrane voltage-dependent Ca^{2+} channels are consequently opened, increasing cytosolic Ca^{2+} concentration and triggering the exocytosis of insulin [28].

To interrogate islet behaviour via microscopy, we used multiple fluorescent dyes probing cell viability and specific steps of GSIS pathway. We added to the flowing culture media PE-labeled Annexin V, a fluorophore for detecting early apoptosis. We also loaded the islets, prior to their insertion into the device, with dyes enabling the detection of the intracellular Ca^{2+} concentration (Fluo-4 or Fura Red) and cell membrane potential (Vf2.1Cl). We used both confocal and two-photon microscopy.

After each islet insertion into the device, we always observed its morphology to verify the integrity of the external membrane (Figure F.2a, b; Online Video 1). Islets that did not show a visibly intact morphology, for example, due to some variability in the extraction process, were discarded. This occurred in less than 5 % of the cases and was usually associated with the presence of some dead cells, assessed by annexin V, not functionally responsive, for instance in terms of intracellular Ca^{2+} oscillations under high-glucose stimulation (Figure F.2c, d).

To verify that the microfluidic system did not impair islet functional response, we perfused the islets with high-glucose medium (11 mM), which is stimulatory for insulin secretion. We used islets from mutant mice that express GFP under the insulin promoter in order to identify the b-cells within the islet. In Figure F.3a, we show persistent oscillations of the intracellular Ca^{2+} at 11 mM. As expected, the oscillations were restricted to β-cells and were not an artifact of the culture system or method of analysis. Moreover, β-cells within the same islet showed a properly synchronized response (Figure F.3b).

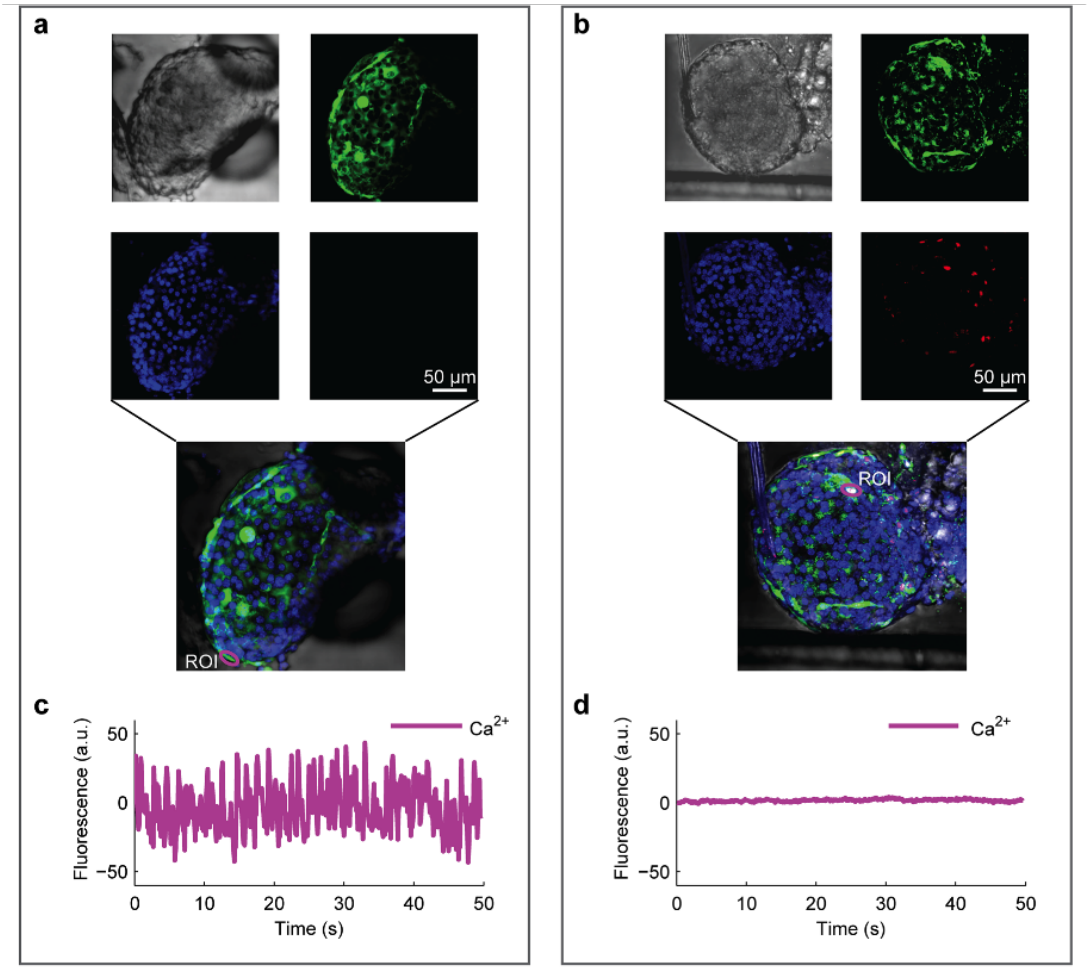


Figure F.2: Two-photon imaging of islet morphology and viability in microfluidics. Islets were pre-loaded with Hoechst 33342 and Fluo-4 for nuclear identification and intracellular Ca^{2+} detection, respectively. Viability was detected by adding annexin V in the flowing medium. **a**, **b** Images of single islets, intact (**a**) or subjected to mechanical stress (**b**), within a microfluidic channel (see video 1). From top left: morphology (transmitted light), nuclei (blue), intracellular Ca^{2+} (green), and mortality (red). **c**, **d** Plots of intracellular Ca^{2+} oscillations in the same islets shown in **a**, **b**, respectively, constantly perfused with buffer containing stimulatory glucose concentration (11 mM). Zero-centered raw data. Annexin-V^{+} cells from the stressed islets do not display oscillatory Ca^{2+} activity, as detected by Fluo-4

To verify islet structural integrity also in terms of cell–cell electrical coupling, we then loaded the islets with Vf2.1Cl, a voltage-sensitive fluorescent dye, and performed imaging at high speed. Under high-glucose conditions, we were able to observe the spatial propagation of membrane potential waves throughout the islet (Figure F.3c; Online Video 2), as previously reported [29].

As a last verification of functionality, we loaded the islets with probes for detection of both membrane potential and intracellular Ca^{2+}. Imaging at high speed, we observed the two waves propagating at the same frequency, but with a small delay of intracellular Ca^{2+} compared to membrane voltage (Figure F.3e, f; Video 3). This result is consistent with Ca^{2+} regulation occurring along the pathway of insulin secretion and confirms that each step of the pathway is in place and functional within the microfluidic culture system.

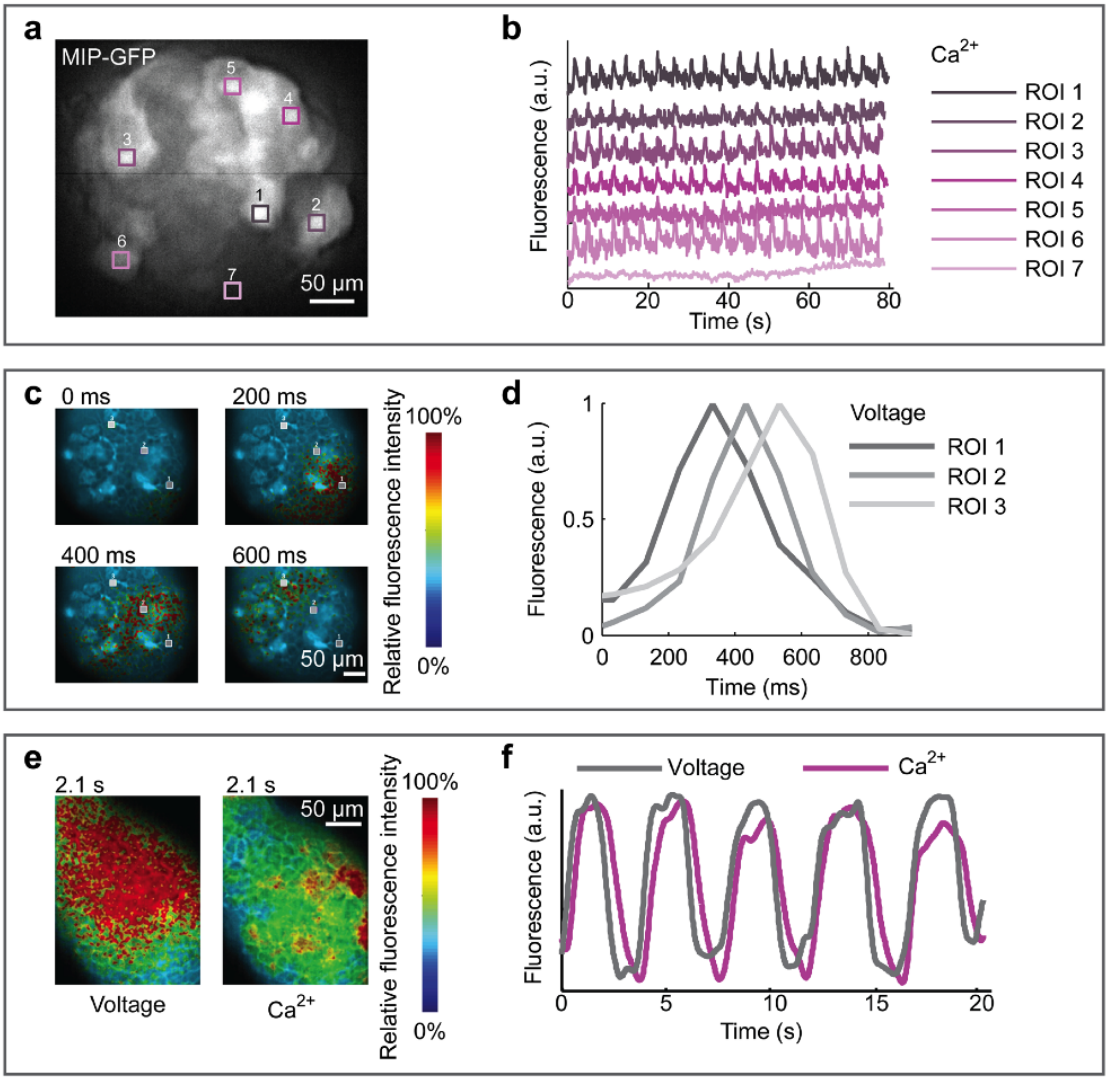


Figure F.3: Validation of islet functional response at 11 mM glucose. **a** Green fluorescence intensity of an islet isolated from a Mouse Insulin Promoter (MIP)-GFP mouse, pre-loaded with Fura Red for intracellular Ca^{2+} detection. **b** Plots of synchronous intracellular Ca^{2+} oscillations of cells in the regions of interest (ROIs) indicated in **a**. ROI 7 refers to non β-cells (GFP-), which do not show Ca^{2+} oscillatory activity in response to high glucose concentration. **c** Membrane potential of an islet from a wild-type mouse loaded with the voltage dye Vf2.1.Cl and imaged at the indicated time points during constant perfusion at high glucose. The false-colored image sequence shows fluorescence intensity. **d** Plot of signal intensity inside ROIs indicated in **c**. Oscillation phase shift describes the spatial propagation of electrical activity from cell to cell within the islet (see also Video 2). **e** Simultaneous recording of intracellular Ca^{2+} and membrane potential in an islet from a wild-type mouse, pre-loaded with Fura Red and Vf2.1.Cl. The false-colored image sequence shows fluorescence intensity. **f** Plot derived from data shown in **e**. Intracellular Ca^{2+} and membrane potential waves propagate with the same periodicity. Ca^{2+} propagation occurs with slightly delayed phase (see also Video 3)

F.4.3 Single-Islet Dynamic Stimulation

Next, we performed dynamic stimulations of the islets, fully taking advantage of the developed microfluidic setup that makes automatically switching between media of different compositions possible. For these experiments, we also added a cell-impermeant Zn^{2+}-sensitive fluorophore (FluoZin-3) to the

flowing culture media.

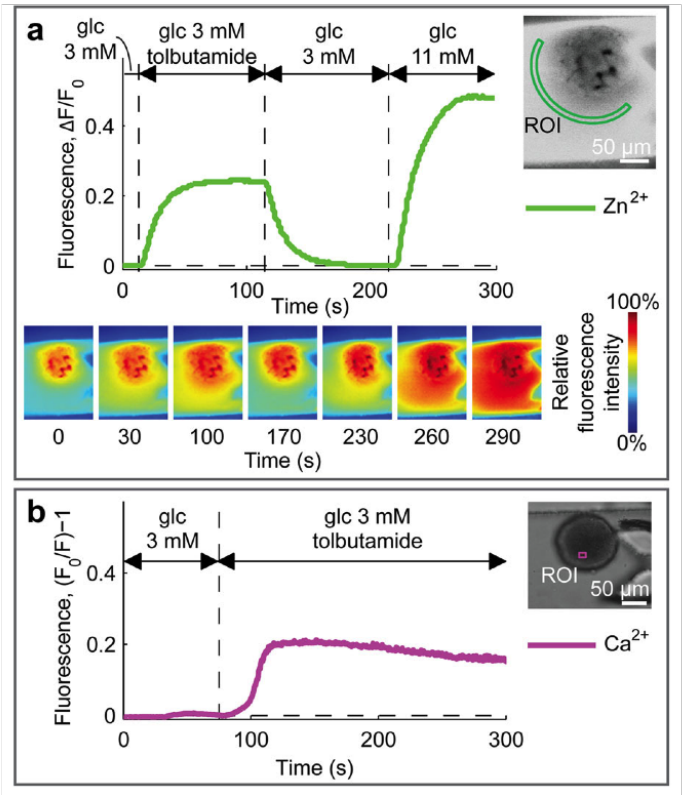


Figure F.4: Physiologic and pharmacologic validation of islet functionality. **a** The islet shown in the inset was dynamically perfused with different media, as indicated, all containing the FluoZin-3 Zn^{2+}-sensitive fluorophore. This dye indicates Zn^{2+} secretion from the islet, as an indirect measurement of insulin secretion [21]. The plot shows changes of fluorescence intensity in the ROI shown in the inset. The false-colored images at the bottom show the spatial fluorescence intensity at different time points (see Video 4). **b** The islet shown in the inset was pre-loaded with FuraRed and dynamically perfused with different media, as indicated. Fluorescence intensity in the inset ROI was quantified in the plot that shows intracellular Ca^{2+} concentration

Zn^{2+} measurement was used to indirectly evaluate the dynamics of insulin secretion, as Zn^{2+} is present within the intracellular insulin-containing granules undergoing exocytosis [30].

We analyzed single-islet response both upon switching between non-stimulatory (3 mM) and stimulatory (11 mM) glucose concentrations and in the presence of tolbutamide, an ATP-sensitive K^+-channel blocker that induces insulin secretion even at non-stimulatory glucose concentrations [31]. As expected, tolbutamide produced a sudden increase of insulin secretion (Figure F.4a; Online Video 4) and of intracellular Ca^{2+} concentration (Figure F.4b).

Altogether, these data demonstrate islet functionality within the microfluidic system, validate the detection method for analyzing single islets with high spatio-temporal resolution in response to stimulations, and give a demonstration of the high potential of this system for performing accurate dynamic studies.

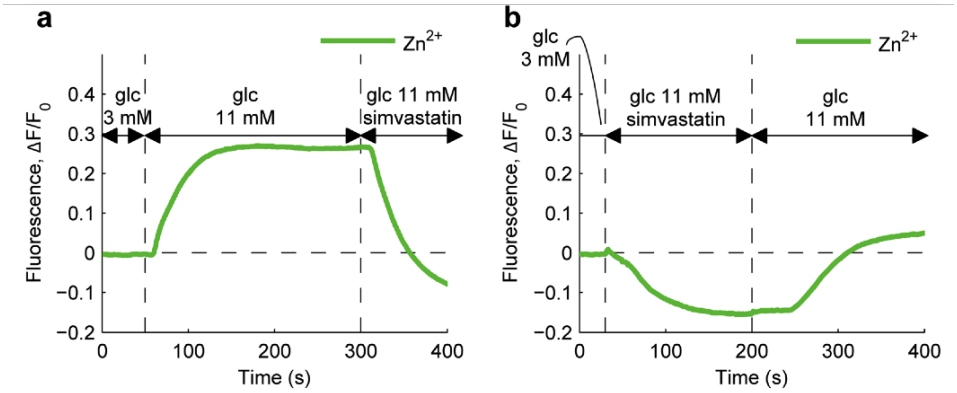


Figure F.5: Effects of simvastatin on glucose-induced insulin secretion. **a** Dynamics of insulin secretion, indirectly measured by FluoZin-3 fluorescence intensity, during islet perfusion with media of different compositions, as indicated. **b** Same as **a**, inverting the order of simvastatin stimulation at high-glucose concentration

F.4.4 Effects of Simvastatin on Insulin Secretion

In a previous study on MIN-6, a murine insulinoma cell line, simvastatin was shown to reduce insulin secretion by targeting ATP-sensitive K^+-channels, voltage-gated Ca^{2+}-channels, muscarinic M3 receptors, and GPR40 in a time span of approximately an hour [13]. On a shorter time scale (minutes), the same study demonstrated that simvastatin suppresses cytosolic Ca^{2+} regulation in response to both stimulatory glucose concentrations and tolbutamide in MIN-6 cells. We used our validated microfluidic system to analyze short-term effects of simvastatin on the insulin secretion pathway using intact single pancreatic islets as a more physiologically relevant in vitro model.

After testing different concentrations of simvastatin from 1 to 10 μM (data not shown), we used a 1 μM concentration in all experiments, as the lowest concentration of drug that elicited an effect on islet function. We then analyzed the ability of simvastatin to modify insulin secretion by single islets in microfluidics. As shown in Figure F.5a and Online Video 5, simvastatin rapidly blocked insulin secretion even in the presence of a stimulatory glucose concentration (11 mM). The same inhibitory effect on secretion was observed independently of the sequence in which stimuli were applied (Figure F.5a, b). Such a rapid and reversible effect of simvastatin on single islets has never been shown before and was detected taking advantage of the small-scale transparent dynamic setup we developed.

To compare our data to those previously obtained with MIN-6 cells, we coupled the measurement of insulin secretion with that of cytosolic Ca^{2+}. Unlike in Yaluri et al. [13], we found that cytosolic Ca^{2+} was not affected by simvastatin in single islets, despite its inhibition of insulin secretion at stimulatory glucose concentrations (Figure F.6a; Online Video 6). This suggests that simvastatin acts on the late stages of insulin granule trafficking, downstream of cytosolic Ca^{2+} regulation. However, when tolbutamide was concomitantly used to induce insulin secretion, simvastatin did not prevent a cytosolic Ca^{2+} increase and insulin secretion, at both low (3 mM) and high (11 mM) glucose concentrations (Figure

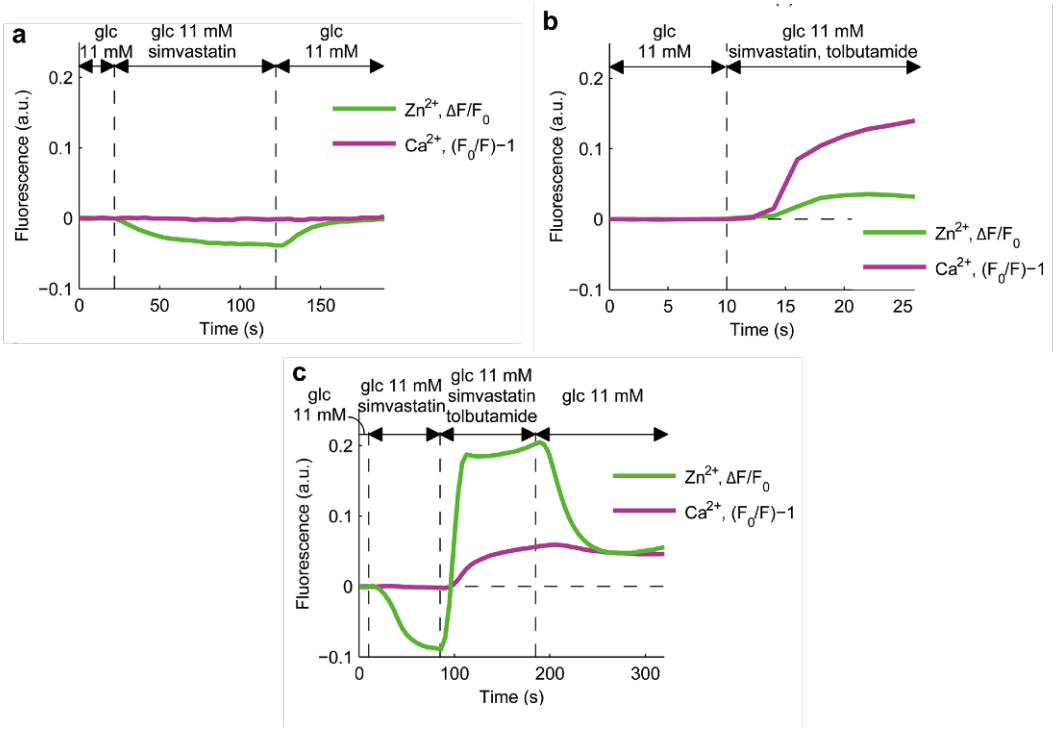


Figure F.6: Effects of simvastatin on insulin secretion and intracellular Ca^{2+} concentration at 11 mM glucose. **a-c** Plots show the simultaneous recording of FluoZin-3 and Fura Red fluorescence intensity of islets pre-loaded with Fura Red and perfused with media of different compositions, as indicated

F.6b). Similarly, islet treatment with simvastatin, prior to addition of tolbutamide, did not prevent the rise in cytosolic Ca^{2+} concentration and insulin secretion (Figure F.6c).

F.5 Discussion

In this study, we developed and validated a microfluidic platform optimized for functional analysis of single pancreatic islets by live confocal and two-photon imaging. Microfluidic systems are increasingly used to study physiological phenomena in vitro, with several advantages [32, 33]. The possibility to analyze a single islet, the fundamental unit of endocrine pancreas, preserves intercellular nteractions within the islet and avoids the confounding complexity of whole animal physiology and inter-islet communication as in organo-culture. By using a miniaturized multichannel system we could remotely control the switch of the perfusion medium entering the microfluidic chamber with a resolution of seconds and analyze live how a single islet responds to different exogenous stimuli. In vivo imaging of the endocrine pancreas is the most suitable physiological model to assess insulin secretion. However, this is poorly standardized, highly invasive, and laborious. In particular, it is difficult to deliver precisely localized stimuli, while dynamically visualizing β-cell response. Ex vivo intact islets are fully functional units, whereas β-cell lines are immortalized and often not pure: the commonly used MIN6 cell line contains a mix of pancreatic endocrine cells [34]. Furthermore, primary islets can be quickly isolated and immedi-

ately used for imaging, whereas b-cell lines are difficult to maintain in culture (Table F.1).

Table F.1: A comparison of different models for beta cell imaging

Methods	Usefulness	Limitations
Cell line	Controlled experimental conditions	Difficult to maintain in culture
	Reproducibility	Immortalized
		No pure beta cell line (e.g., MIN6)
		No interaction with alpha/delta cells
Pooled islet culture	Controlled experimental conditions	Batch-to-batch variability
	Quick isolation (2 h)	3 days of vitality after isolation
	Preserved islet heterogeneity integrity	Average response from asynchronous islets
	High signal-to-noise ratio	
Intact single islet	Controlled experimental conditions	3 days of vitality after isolation
	Quick isolation (2 h)	Islet-to-islet variability
	Preserved islet heterogeneity integrity	
	Quasi-physiological model	
Animal model	High-fidelity physiological model	Difficult visualization
		No controlled conditions
		High variability

Using confocal and two-photon imaging, we showed normal functionality of single islets within the microfluidic device in terms of synchronized membrane voltage and intracellular Ca^{2+} regulation, as well as glucose- and tolbutamide-stimulated insulin secretion.

As a clinically relevant application of this setup, we explored the effects of the HMG-CoA reductase inhibitor, simvastatin, on Ca^{2+} dynamics and insulin secretion. Several reports suggest that statins may increase the risk of developing diabetes by acting on insulin secretion, insulin resistance, or both [11, 12, 35, 36]. However, whether statins truly inhibit insulin secretion is unclear.

Pooling together five murine islets, Źúniga-Hertz et al. demonstrate that simvastatin interferes with isoprenylation of proteins involved in insulin granule exocytosis on a 2-h time scale and with cholesterol biosynthesis on a 24-h time scale [14]. Shorter time effects of simvastatin on the GSIS pathway have only been demonstrated in vitro using a murine β-cell line, MIN-6, and only as for its action on cytosolic Ca^{2+} dynamics [13].

Here, we extended the results previously obtained with MIN-6 cells to single, freshly isolated murine

islets. While Yaluri et al. [13] found that simvastatin affects Ca^{2+} regulation, in our results single islet cytosolic Ca^{2+} is not perturbed by simvastatin. Moreover, for the first time, we were able to demonstrate that simvastatin affects insulin secretion of single islets, a result previously shown only by pooling more islets together, a less physiological model because of the in vivo islet–islet coordination [37], and analyzing insulin secretion only on much longer time scales. The rapidity of the simvastatin effect on insulin secretion suggests its involvement in regulatory mechanisms other than cholesterol biosynthesis and post-translational modifications, which require more time to occur.

Ultimately, we showed that single islets respond to simvastatin acutely and reversibly, blocking the normal insulin secretion induced by high glucose. As there is no effect of simvastatin on Ca^{2+} intracellular concentration and dynamics, we suggest that the inhibitory effect on insulin secretion may be exerted at the levels of insulin granule trafficking. Moreover, tolbutamide can effectively counteract the inhibitory effect of simvastatin on insulin secretion, suggesting that simvastatin's action is reversible not only by its withdrawal, but also by Ca^{2+} overload and membrane depolarization. Reversibility of the effect has clinical implications and likely reflects the relatively benign effect of statins on diabetes onset or progression. Limitations of this study are inherent to the short time course of observation and the lack of a dose-dependent response. Table F.2 highlights the advancement over the state of the art made by our study.

Table F.2: Advancement over state of the art

What was previously known?	Clinical trials and epidemiological studies suggest that statins increase the risk of developing diabetes in healthy individuals and may worsen glycemia in diabetic patients
	In vitro studies show that statins impair insulin secretion by affecting multiple pathways, including cellular cholesterol synthesis, membrane fluidity, and isoprenylation of proteins
What this study adds?	We have used a new methodological approach to dynamically study the inhibitory effect of simvastatin on insulin secretion of intact islets. Data show this effect is rapid and reversible
What are the implications?	Rapidity of inhibition suggests mechanisms unrelated to cholesterol synthesis, but possibly based on electrophysiology or granule docking/fusion modulation
	Reversibility of inhibition has clinical implications and may suggest a relatively benign effect of statins on beta cell function
Which are the next steps?	To better dissect the molecular mechanisms of insulin secretion inhibition by statins
	To identify concomitant treatments that may counteract the effects of statins on insulin secretion

F.6 Conclusion

In conclusion, we report that simvastatin rapidly and reversibly impairs insulin release from intact islets ex vivo. Further work is needed to understand the underlying mechanisms of action of simvastatin at the second-to-minute scale. The microfluidic and imaging techniques herein described will be suitable to dissect the molecular and electrophysiological processes underlying this statin effect. The possibility to use concomitant treatments will help to clarify whether the effect of simvastatin can be counteracted by other glucose-lowering agents with mechanisms of action different from sulfonylureas, such as GLP-1 receptor agonists or even SLGT2 inhibitors.

F.7 References

1. Collins R, Reith C, Emberson J, Armitage J, Baigent C, Blackwell L, et al. Interpretation of the evidence for the efficacy and safety of statin therapy. Lancet Lond Engl. 2016;S0140-6736-5.

2. Anderson TJ, Grégoire J, Pearson GJ, Barry AR, Couture P, Dawes M, et al. 2016 Canadian Cardiovascular Society guidelines for the management of dyslipidemia for the prevention of cardiovascular disease in the adult. Can J Cardiol. 2016;S0828-282-2.

3. Dong N, Xie Z, Dai J, Wang W, Sun R, Zhan Y, et al. Statin-induced improvements in vulnerable plaques are attenuated in poorly controlled diabetic patients with coronary atherosclerosis disease: a serial optical coherence tomography analysis. Acta Diabetol. 2016. doi:10.1007/s00592-016-0902-9.

4. Garber AJ, Abrahamson MJ, Barzilay JI, Blonde L, Bloomgarden ZT, Bush MA, et al. Consensus statement by the American Association of Clinical Endocrinologists and American College of Endocrinology on the comprehensive type 2 diabetes management algorithm–2016 Executive summary. Endocr Pract Off J Am Coll Endocrinol Am Assoc Clin Endocrinol. 2016;22:84–113.

5. Chamberlain JJ, Rhinehart AS, Shaefer CF, Neuman A. Diagnosis and management of diabetes: synopsis of the 2016 American Diabetes Association standards of medical care in diabetes. Ann Intern Med. 2016;164:542–52.

6. Colhoun HM, Betteridge DJ, Durrington PN, Hitman GA, Neil HAW, Livingstone SJ, et al. Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study (CARDS): multicentre randomised placebo-controlled trial. Lancet Lond Engl. 2004;364:685–96.

7. Swerdlow DI, Preiss D, Kuchenbaecker KB, Holmes MV, Engmann JEL, Shah T, et al. HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials. Lancet Lond Engl. 2015;385:351–61.

8. Navarese EP, Buffon A, Andreotti F, Kozinski M, Welton N, Fabiszak T, et al. Meta-analysis of impact of different types and doses of statins on new-onset diabetes mellitus. Am J Cardiol. 2013;111:1123–30.

9. Preiss D, Seshasai SRK, Welsh P, Murphy SA, Ho JE, Waters DD, et al. Risk of incident diabetes with intensive-dose compared with moderate-dose statin therapy: a meta-analysis. JAMA. 2011;305:2556–64.

10. Sattar N, Preiss D, Murray HM, Welsh P, Buckley BM, de Craen AJM, et al. Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. Lancet Lond Engl. 2010;375:735–42.

11. Erqou S, Lee CC, Adler AI. Statins and glycaemic control in individuals with diabetes: a systematic review and meta-analysis. Diabetologia. 2014;57:2444–52.

12. Cederberg H, Stančáková A, Yaluri N, Modi S, Kuusisto J, Laakso M. Increased risk of diabetes with statin treatment is associated with impaired insulin sensitivity and insulin secretion: a 6 year follow-up study of the METSIM cohort. Diabetologia. 2015;58:1109–17.

13. Yaluri N, Modi S, López Rodríguez M, Stančáková A, Kuusisto J, Kokkola T, et al. simvastatin impairs insulin secretion by multiple mechanisms in MIN6 cells. PLoS One. 2015;10:e0142902.

14. Zúñiga-Hertz JP, Rebelato E, Kassan A, Khalifa AM, Ali SS, Patel HH, et al. Distinct pathways of cholesterol biosynthesis impact on insulin secretion. J Endocrinol. 2015;224:75–84.

15. Ridker PM, Pradhan A, MacFadyen JG, Libby P, Glynn RJ. Cardiovascular benefits and diabetes risks of statin therapy in primary prevention: an analysis from the JUPITER trial. Lancet Lond. Engl. 2012;380:565–71.

16. Blackburn DF, Chow JY, Smith AD. Statin use and incident diabetes explained by bias rather than biology. Can J Cardiol. 2015;31:966–9.

17. Robinson JG. Statins and diabetes risk: how real is it and what are the mechanisms? Curr Opin Lipidol. 2015;26:228–35.

18. Wang Y, Mendoza-Elias J, McGarrigle J, Nourmohammadzadeh M, Wang Q, Li Z, et al. Application of microfluidic technology for studying islet physiology and pathophysiology. Micro Nanosyst. 2013;5:216–23.

19. Gagliano O, Elvassore N, Luni C. Microfluidic technology enhances the potential of human pluripotent stem cells. Biochem Biophys Res Commun. 2016;473:683–7.

20. Benninger RKP, Hutchens T, Head WS, McCaughey MJ, Zhang M, Le Marchand SJ, et al. Intrinsic islet heterogeneity and gap junction coupling determine spatiotemporal $Ca2^{+}$ wave dynamics. Biophys J. 2014;107:2723–33.

21. Lomasney AR, Yi L, Roper MG. Simultaneous monitoring of insulin and islet amyloid polypeptide secretion from islets of langerhans on a microfluidic device. Anal Chem. 2013;85:7919–25.

22. Zambon A, Zoso A, Luni C, Frommer WB, Elvassore N. Determination of glucose flux in live myoblasts by microfluidic nanosensing and mathematical modeling. Integr Biol. 2014;6:277–88.

23. Dishinger JF, Reid KR, Kennedy RT. Quantitative monitoring of insulin secretion from single islets of langerhans in parallel on a microfluidic chip. Anal Chem. 2009;81:3119–27.

24. Stull ND, Breite A, McCarthy R, Tersey SA, Mirmira RG. Mouse islet of Langerhans isolation using a combination of purified collagenase and neutral protease. J Vis Exp. 2012. doi:10.3791/4137.

25. Miller EW, Lin JY, Frady EP, Steinbach PA, Kristan WB, Tsien RY. Optically monitoring voltage in neurons by photo-induced electron transfer through molecular wires. Proc Natl Acad Sci USA. 2012;109:2114–9.

26. Ciubotaru CD, Bastianello S, Beltramello M, Pozzan T, Mammano F. Multi-modal imaging of cytosolic and mitochondrial Ca2?. In: Adlassnig K-P, Bracale M, editors. Anaheim: ACTA Press; 2005. p. 99–102.

27. Kurebayashi N, Harkins AB, Baylor SM. Use of fura red as an intracellular calcium indicator in frog skeletal muscle fibers. Biophys J. 1993;64:1934–60.

28. Henquin JC. Regulation of insulin secretion: a matter of phase control and amplitude modulation. Diabetologia. 2009;52:739–51.

29. Santos RM, Rosario LM, Nadal A, Garcia-Sancho J, Soria B, Valdeolmillos M. Widespread synchronous $[Ca^{2+}]$i oscillations due to bursting electrical activity in single pancreatic islets. Pflugers Arch. 1991;418:417–22.

30. Dunn MF. Zinc-ligand interactions modulate assembly and stability of the insulin hexamer-a review. Biometals Int J Role Met Ions BiolBiochem Med. 2005;18:295–303.

31. Trube G, Rorsman P, Ohno-Shosaku T. Opposite effects of tolbutamide and diazoxide on the ATP-dependent K^{+} channel in mouse pancreatic beta-cells. Pflugers Arch. 1986;407:493–9.

32. Salieb-Beugelaar GB, Simone G, Arora A, Philippi A, Manz A. Latest developments in microfluidic cell biology and analysis systems. Anal Chem. 2010;82:4848–64.

33. Wang D, Bodovitz S. Single cell analysis: the new frontier in "omics". Trends Biotechnol. 2010;28:281–90.

34. Nakashima K, Kanda Y, Hirokawa Y, Kawasaki F, Matsuki M, Kaku K. MIN6 is not a pure beta cell line but a mixed cell line with other pancreatic endocrine hormones. Endocr J. 2009;56:45–53.

35. Livingstone SJ, Looker HC, Akbar T, Betteridge DJ, Durrington PN, Hitman GA, et al. Effect of atorvastatin on glycaemia progression in patients with diabetes: an analysis from the Collaborative Atorvastatin in Diabetes Trial (CARDS). Diabetologia. 2016;59:299–306.

36. Swerdlow DI, Sattar N. A dysglycaemic effect of statins in diabetes: relevance to clinical practice? Diabetologia. 2014;57:2433–5.

37. Pedersen MG, Bertram R, Sherman A. Intra- and inter-islet synchronization of metabolically driven insulin secretion. Biophys J. 2005;89:107–19.

Bibliography

G. Asher and U. Schibler. Crosstalk Between Components of Circadian and Metabolic Cycles in Mammals. *Cell Metabolism*, 13(2):125–137, 2011.

P. J. Atherton and K. Smith. Muscle protein synthesis in response to nutrition and exercise. *The Journal of Physiology*, 590(5):1049–1057, 2012.

A. Ballesta, P. F. Innominato, R. Dallmann, D. A. Rand, and F. A. Lévi. Systems Chronotherapeutics. *Pharmacological Reviews*, 69(2):161–199, 2017.

A. Balsalobre. Resetting of Circadian Time in Peripheral Tissues by Glucocorticoid Signaling. *Science*, 289(5488):2344–2347, 2000.

K.G. Baron and K. J. Reid. Circadian Misalignment and Health. *International Review of Psychiatry*, 26 (2):139–154, 2014.

J. Bass. Circadian Topology of Metabolism. *Nature*, 491(7424):348–356, 2012.

J. Bass and J. S. Takahashi. Circadian integration of metabolism and energetics. *Science (New York, N.Y.)*, 330(6009):1349–54, 2010.

D. Bell-Pedersen, V. M. Cassone, D. J. Earnest, S. S. Golden, P. E. Hardin, T. L. Thomas, and M. J. Zoran. Circadian Rhythms from Multiple Oscillators: Lessons from Diverse Organisms. *Nature Reviews Genetics*, 6(7):544–546, 2005.

M. M. Bellet, R. Orozco-Solis, S. Sahar, K. Eckel-Mahan, and P. Sassone-Corsi. The Time of Metabolism: NAD+ , SIRT1, and the Circadian Clock. *Cold Spring Harbor Symposia on Quantitative Biology*, LXXVI:31–38, 2011.

R. Bumgarner. DNA microarrays: Types, Applications and their Future. *Current Protocols Molecular Biology*, 6137(206):1–17, 2013.

A. J. Burnham and R. Viveros. Frameworks for Latent Variable Multivariate Regression. *Journal of chemometrics*, 10(1):31–45, 1996.

J. Camacho, J. Picó, and A. Ferrer. Data understanding with PCA: Structural and Variance Information plots. *Chemometrics and Intelligent Laboratory Systems*, 100(1):48–56, 2010.

I. Chaves, G. T. J. Van Der Horst, R. Schellevis, R. M. Nijman, M. G. Koerkamp, F. C. P. Holstege, M. P. Smidt, and M. F. M. Hoekman. Insulin-FOXO3 signaling modulates circadian rhythms via regulation of clock transcription. *Current Biology*, 24(11):1248–1255, 2014.

C. Cordella. PCA: The Basic Building Block of Chemometrics. In I. S. Krull, editor, *Analytical Chemistry*. InTech (Rijeka), 2012.

F. Damiola, N. Le Minli, N. Preitner, B. Kornmann, F. Fleury-Olela, and U. Schibler. Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. *Genes and Development*, 14(23):2950–2961, 2000.

F. Dang, X. Sun, X. Ma, R. Wu, D. Zhang, Y. Chen, Q. Xu, Y. Wu, and Y. Liu. Insulin post-transcriptionally modulates Bmal1 protein to affect the hepatic circadian clock. *Nature Communications*, 7(23):12696, 2016.

T. Danino, O. Mondragón-Palomino, L. Tsimring, and J. Hasty. A synchronized quorum of genetic clocks. *Nature*, 463(7279):326–330, 2010.

A. Deckard, R. C. Anafi, J. B. Hogenesch, S. B. Haase, J. Harer, and A. Valencia. Design and analysis of large-scale biological rhythm studies: A comparison of algorithms for detecting periodic signals in biological data. *Bioinformatics*, 29(24):3174–3180, 2013.

T. A. Duncombe, A. M. Tentori, and A. E. Herr. Microfluidics: reframing biological enquiry. *Nature Reviews Molecular Cell Biology*, 16(9):554–567, 2015.

K. Eckel-Mahan and P. Sassone-Corsi. Metabolism and the Circadian Clock Converge. *Physiological Reviews*, 93(1):107–135, 2013.

K. L. Eckel-Mahan, V. R. Patel, S. De Mateo, R. Orozco-Solis, N. J. Ceglia, S. Sahar, S. A. Dilag-Penilla, K. A. Dyar, P. Baldi, and P. Sassone-Corsi. Reprogramming of the Circadian Clock by Nutritional Challenge. *Cell*, 155(7):1464–1478, 2013.

L. Eriksson, E. Johansson, N. Kettaneh-Wold, J. Trygg, C. Wikström, and S. Wold. *Multi- and Megavariate Data Analysis. Part I. Basic Principles and Applications*. Umetics AB, Umeå (Sweden), 2006.

P. Facco, A. C. Santomaso, and M. Barolo. Artificial vision system for particle size characterization from bulk materials. *Chemical Engineering Science*, 164:246–257, 2017.

O. Gagliano. *Development of Microfluidic Cell Culture Technology for Studying the Control of the Circadian Clock for the Study of Type 2 Diabetes by Metabolic Cycles*. Phd thesis, University of Padova (Italy), 2015.

C. B. Green, J. S. Takahashi, and J. Bass. The Meter of Metabolism. *Cell*, 134(5):728–742, 2008. doi: 10.1016/j.cell.2008.08.022.

A. Gupta, B. Hepp, and M. Khammash. Noise Induces the Population-Level Entrainment of Incoherent, Uncoupled Intracellular Oscillators. *Cell Systems*, 3(6):521–531, 2016.

S. Halldorsson, E. Lucumi, R. Gómez-Sjöberg, and R. M.T. Fleming. Advantages and challenges of microfluidic cell culture in polydimethylsiloxane devices. *Biosensors and Bioelectronics*, 63:218–231, 2015.

Y. Hamaguchi, Y. Tahara, H. Kuroda, A. Haraguchi, and S. Shibata. Entrainment of mouse peripheral circadian clocks to <24 h feeding/fasting cycles under 24 h light/dark conditions. *Scientific Reports*, 5 (1):14207, 2015.

R. Hara, K. Wan, H. Wakamatsu, R. Aida, T. Moriya, M. Akiyama, and S. Shibata. Restricted feeding entrains liver clock without participation of the suprachiasmatic nucleus. *Genes to Cells*, 6(3):269–278, 2001.

T. Hirota, T. Okano, K. Kokame, H. Shirotani-Ikejima, T. Miyata, and Y. Fukada. Glucose Down-regulates Per1 and Per2 mRNA Levels and Induces Circadian Gene Expression in Cultured Rat-1 Fibroblasts. *Journal of Biological Chemistry*, 277(46):44244–44251, 2002.

M. E. Hughes, L. DiTacchio, K. R. Hayes, C. Vollmers, S. Pulivarthy, J. E. Baggs, S. Panda, and J. B. Hogenesch. Harmonics of circadian gene transcription in mammals. *PLoS Genetics*, 5(4), 2009.

M. E. Hughes, J. B. Hogenesch, and K Kornacker. JTK_CYCLE: an efficient non-parametric algorithm for detecting rhythmic components in genome-scale datasets. *Journal of Biological Rhythms*, 25(5): 372–380, 2010.

Y. P. Hung and G. Yellen. Live cell imaging of cytosolic NADH-NAD+ redox state using a genetically encoded fluorescent biosensor. *Methods Molecular Biology*, 1071:83–95, 2014.

S. I. Imai and L. P. Guarente. NAD+ and sirtuins in aging and disease. *Trends in Cell Biology*, 24(8): 464–471, 2014.

J.E. Jackson. *A user's guide to Principal Component Analysis*. John Wiley & Sons, New York (U.S.A.), 1991.

R.A. Johnson and D.W. Wichern. *Applied Multivariate Statistical Analysis*. Pearson Education, Upper Saddle RIver, NJ (U.S.A.), 6 edition, 2007.

A. Kohsaka, A. D. Laposky, K. M. Ramsey, C. Estrada, C. Joshu, Y. Kobayashi, F. W. Turek, and J. Bass. High-Fat Diet Disrupts Behavioral and Molecular Circadian Rhythms in Mice. *Cell Metabolism*, 6(5): 414–421, 2007.

B. Kornmann, O. Schaad, and H. Reinke. Regulation of Circadian Gene Expression in Liver by Systemic Signals and Hepatocyte Oscillators Regulation of Circadian Gene Expression in Liver by Systemic Signals and Hepatocyte Oscillators. LXXII:319–330, 2007.

M. Lagouge, C. Argmann, Z. Gerhart-Hines, H. Meziane, C. Lerin, F. Daussin, N. Messadeq, J. Milne, P. Lambert, P. Elliott, B. Geny, M. Laakso, P. Puigserver, and J. Auwerx. Resveratrol Improves Mitochondrial Function and Protects against Metabolic Disease by Activating SIRT1 and PGC-1α. *Cell*, 127(6):1109–1122, 2006.

K. K. Lee, C. H. Ahn, and C. I. Hong. Circadian Rhythms in Neurospora Crassa on a Microfluidic Device for Real-Time Gas Perturbations. 2:1247–1250, 2013.

F. Lévi, A. Okyar, S. Dulong, P. F. Innominato, and J. Clairambault. Circadian Timing in Cancer Treatments. *Annual Review of Pharmacology and Toxicology*, 50(1):377–421, 2010.

J. Li, G. R. Grant, J. B. Hogenesch, and M. E. Hughes. *Considerations for RNA-seq analysis of circadian rhythms*, volume 551. Elsevier Inc., 1 edition, 2015.

J. O. Lipton, E. D. Yuan, L. M. Boyle, D. Ebrahimi-Fakhari, E. Kwiatkowski, A. Nathan, T. Güttler, F. Davis, J. M. Asara, and M. Sahin. The circadian protein BMAL1 regulates translation in response to S6K1-mediated phosphorylation. *Cell*, 161(5):1138–1151, 2015.

R. López-Negrete de la Fuente, S. García-Muñoz, and L. T. Biegler. An Efficient Nonlinear Programming Strategy for PCA Models with Incomplete Data Sets. *Journal of Chemometrics*, (2009):n/a–n/a, 2010.

C. Luni, E. Serena, and N. Elvassore. Human-on-chip for therapy development and fundamental science. *Current Opinion in Biotechnology*, 25:45–50, 2014.

E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, M. Shekhar, K.and Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, and S. A. McCarroll. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 161(5):1202–1214, 2015.

K. Maiese. Moving to the Rhythm with Clock (Circadian) Genes, Autophagy, mTOR, and SIRT1 in Degenerative Disease and Cancer. *Current Neurovascular Research*, 14(3):48–56, 2017.

A. P. Mansur, A. Roggerio, M. F. S. Goes, S. D. Avakian, D. P. Leal, R. C. Maranhão, and C. M. C. Strunz. Serum Concentrations and Gene Expression of Sirtuin 1 in Healthy and Slightly Overweight Subjects After Caloric Restriction or Resveratrol Supplementation: A randomized trial. *International Journal of Cardiology*, 227:788–794, 2017.

K.V. Mardia, J.T. Kent, and J.M. Bibby. *Multivariate Analysis*. Academic Press Limited, London (U.K.), 1979.

P. McNamara, S. S. Beom, R. D. Rudic, A. Sehgal, D. Chakravarti, and G. A. FitzGerald. Regulation of CLOCK and MOP4 by nuclear hormone receptors in the vasculature: A humoral mechanism to reset a peripheral clock. *Cell*, 105(7):877–889, 2001.

J. Melin and S. R. Quake. Microfluidic Large-Scale Integration: The Evolution of Design Rules for Biological Automation. *Annual Review of Biophysics and Biomolecular Structure*, 36(1):213–231, 2007.

Natascia Meneghetti. *Implementing "Quality by Design" in the Pharmaceutical Industry: a Data-driven Approach*. Phd thesis, University of Padova (Italy), 2016.

J. A. Mohawk, C. B. Green, and J. S. Takahashi. Central and Peripheral Circadian Clocks in Mammals. *Annual Review of Neuroscience*, 35(1):445–462, 2012.

C. J. Morris, T. E. Purvis, K. Hu, and F. A. J. L. Scheer. Circadian Misalignment Increases Cardiovascular Disease Risk Factors in Humans. *Proceedings of the National Academy of Sciences*, 113(10):E1402–E1411, 2016.

E. S. Musiek and D. M. Holtzman. Mechanisms linking circadian clocks, sleep, and neurodegeneration. *Science*, 354(6315):1004–1008, 2016.

E. Nagoshi, C. Saini, C. Bauer, T. Laroche, F. Naef, U. Schibler, Q. E. Ansermet, C. Boveresses, and C. Epalinges. Circadian Gene Expression in Individual Fibroblasts: Cell-Autonomous and Self-Sustained Oscillators Pass Time to Daughter Cells. *Cell*, 119(5):693–705, 2004.

Y. Nakahata, M. Kaluzova, B. Grimaldi, S. Sahar, J. Hirayama, D. Chen, L. P. Guarente, and P. Sassone-Corsi. The NAD+ -Dependent Deacetylase SIRT1 Modulates CLOCK-Mediated Chromatin Remodeling and Circadian Control. *Cell*, 134:329–340, 2008.

P. Nomikos and J. F. MacGregor. Monitoring Batch Processes Using Multiway Principal Component Analysis. *AIChE Journal*, 40(8):1361–1375, 1994.

H. Nonaka, N. Emoto, K. Ikeda, H. Fukuya, M. S. Rohman, S. B. Raharjo, K. Yagita, H. Okamura, and M. Yokoyama. Angiotensin II induces circadian gene expression of clock genes in cultured vascular smooth muscle cells. *Circulation*, 104(15):1746–1748, 2001.

H. Oike and M. Kobori. Resveratrol Regulates Circadian Clock Genes in Rat-1 Fibroblast Cells. *Bioscience, Biotechnology, and Biochemistry*, 72(11):3038–3040, 2008.

Matteo Ottavian. *Latent Variable Modeling to Assist Product Quality Characterization in the Food and Pharmaceutical Industries*. Phd thesis, University of Padova (Italy), 2014.

J Ouellette. A New Wave of Microfluidic Devices. *The Industrial Physicist*, pages 14–17, 2003.

S. Panda. Circadian physiology of metabolism. *Science*, 354(6315):317–322, 2016.

C. L. Partch, C. B. Green, and J. S. Takahashi. Molecular Architecture of the Mammalian Circadian Clock. *Trends in Cell Biology*, 24(2):90–99, 2014.

J. Qian and F. A. J. L. Scheer. Circadian System and Glucose Metabolism: Implications for Physiology and Disease. *Trends in Endocrinology and Metabolism*, 27(5):282–293, 2016.

K. M. Ramsey, J. Yoshino, C. S. Brace, D. Abrassart, Y. Kobayashi, B. Marcheva, H.-K. Hong, J. L. Chong, E. D. Buhr, C. Lee, J. S. Takahashi, S.-I. Imai, and J. Bass. Circadian Clock Feedback Cycle Through NAMPT-Mediated NAD+ Biosynthesis. *Science*, 324(5927):651–654, 2009. ISSN 0036-8075. doi: 10.1126/science.1171641. URL `http://www.sciencemag.org/cgi/doi/10.1126/science.1171641`.

S. Sahar and P. Sassone-Corsi. Metabolism and Cancer: the Circadian Clock Connection. *Nature Reviews Cancer*, 9(12):886–896, 2009.

M. Sato, M. Murakami, K. Node, R. Matsumura, and M. Akashi. The Role of the Endocrine System in Feeding-Induced Tissue-Specific Circadian Entrainment. *Cell Reports*, 8(2):393–401, 2014.

M. Straume. DNA Microarray Time Series Analysis: Automated Statistical Assessment of Circadian Rhythms in Gene Expression Patterning Introduction. In *Methods in Enzymology*, volume 383, pages 149–166. 2004.

J. S. Takahashi, H.-K. Hong, C. H. Ko, and E. L. McDearmon. The Genetics of Mammalian Circadian Order and Disorder: Implications for Physiology and Disease. *Nature Reviews Genetics*, 9(10):764–775, 2008.

S.-W. Teng, S. Mukherji, J. R. Moffitt, S. de Buyl, and E. K. O'Shea. Robust Circadian Oscillations in Growing Cyanobacteria Require Transcriptional Feedback. *Science*, 340(6133):737–740, 2013.

T. Thorsen, S. J. Maerkl, and S. R. Quake. Microfluidic Large-Scale Integration. *Science*, 298(October): 5593, 2002.

Emanuele Tomba. *Latent Variable Modeling Approaches to Assist the Implementation of Quality-by-Design Paradigms in Pharmaceutical Development and Manufacturing*. Phd thesis, University of Padova (Italy), 2013.

X. Tong and L. Yin. Circadian Rhythms in Liver Physiology and Liver Diseases. *Comprehensive Physiology*, 3(2):917–940, 2013.

M. A. Unger. Monolithic Microfabricated Valves and Pumps by Multilayer Soft Lithography. *Science*, 288(5463):113–116, 2000.

A. Vainshtein, P. Grumati, M. Sandri, and P. Bonaldo. Skeletal muscle, autophagy, and physical activity: The ménage à trois of metabolic regulation in health and disease. *Journal of Molecular Medicine*, 92 (2):127–137, 2014.

C. Vollmers, S. Gill, L. DiTacchio, S. R. Pulivarthy, H. D. Le, and S. Panda. Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression. *Proceedings of the National Academy of Sciences*, 106(50):21453–21458, 2009.

D. Welsh, S. H. Yoo, A. C: Liu, J. S. Takahashi, and S. A. Kay. Bioluminescence Imaging of Individual Fibroblasts Reveals Persistent, Independently Phased Circadian Rhythms of Clock Gene Expression. *Current Biology*, 14:2289–2295, 2004.

S. Wichert, K. Fokianos, and K. Strimmer. Identifying periodically expressed transcripts in microarray time series data. *Bioinformatics*, 20(1):5–20, 2003.

D. Yamajuku, T. Inagaki, T. Haruma, S. Okubo, Y. Kataoka, S. Kobayashi, K. Ikegami, T. Laurent, T. Kojima, K. Noutomi, S. Hashimoto, and H. Oda. Real-time monitoring in three-dimensional hepatocytes reveals that insulin acts as a synchronizer for liver clock. *Scientific Reports*, 2(1):439, 2012.

S. Yamazaki and J. S. Takahashi. Real-Time Luminescence Reporting of Circadian Gene Expression in Mammals. *Methods in Enzymology*, 393:288–301, 2005.

G. Yang, L. Chen, G. R. Grant, G. Paschos, W. L. Song, E. S. Musiek, V. Lee, S. C. McLoughlin, T. Grosser, G. Cotsarelis, and G. A. FitzGerald. Timing of Expression of the Core Clock Gene Bmal1 Influences its Effects on Aging and Survival. *Science translational medicine*, 8(324):324ra16, 2016.

R. Yang and Z. Su. Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation. *Bioinformatics*, 26:168–174, 2010.

A. Zambon. *Development of Microfluidic Cell Culture Technology for the Study of Type 2 Diabetes*. Phd thesis, University of Padova (Italy), 2013.

A. Zambon, A. Zoso, O. Gagliano, E. Magrofuoco, G. P. Fadini, A. Avogaro, M. Foletto, S. Quake, and N. Elvassore. High Temporal Resolution Detection of Patient-Specific Glucose Uptake from Human ex Vivo Adipose Tissue On-Chip. *Analytical Chemistry*, 87(13):6535–6543, 2015.

R. Zhang, N. F. Lahens, H. I. Ballance, M. E. Hughes, and J. B. Hogenesch. A circadian gene expression atlas in mammals: implications for biology and medicine. *Proceedings of the National Academy of Sciences of the United States of America*, 111(45):16219–24, 2014.