# Rhythmic modulation of transcriptional burst frequency in circadian gene promoters

THÈSE N$^O$ 8016 (2017)

PRÉSENTÉE LE 3 NOVEMBRE 2017
À LA FACULTÉ DES SCIENCES DE LA VIE
UNITÉ DU PROF. NAEF
PROGRAMME DOCTORAL EN APPROCHES MOLÉCULAIRES DU VIVANT

## ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

## Damien Lionel NICOLAS

acceptée sur proposition du jury:

Prof. J. Lingner, président du jury
Prof. F. Naef, Prof. D. M. Suter, directeurs de thèse
Prof. S. Pelet, rapporteur
Prof. J. Chao, rapporteur
Prof. B. Fierz, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2017

# Contents

# Abstract

Isogenic cells sharing a common environment present a large degree of heterogeneity in gene expression, and stochasticity inherent to transcription substantially participates in this cell-to-cell variability. Notably, a majority of mammalian genes are transcribed during short periods termed transcription bursts followed by longer periods of transcriptional inactivity. Interestingly, genes display variability in the frequencies and sizes of their bursts, implying that different regulatory mechanisms actively participate in shaping gene-specific bursting signatures. However, the nature of these molecular mechanisms and their precise contribution to transcriptional bursting remains elusive.

In this work, we used a short-lived luciferase reporter under the control of a *Bmal1* circadian promoter stably inserted into the genome of NIH-3T3 cultured fibroblasts to quantify its transcriptional bursting along the circadian cycle and at three different reporter integration sites. By recording dynamic variations of the luminescence signal at the single-cell level and counting individual transcripts using smRNA-FISH at specific time-points, we could infer the transcriptional bursting parameters characteristic of these conditions using a telegraph (on-off) model of gene expression. We observed that while the integration site-specific differences in expression levels mainly arose from burst size dissimilarities, the burst frequency predominantly modulated the temporal variations in expression of *Bmal1* over the circadian cycle. Thus, both parameters are uncoupled and can be independently modulated to regulate expression levels.

By focusing on the molecular origins of bursting, we found that the rhythmic circadian modulation of burst frequencies depended on the presence of ROR responsive elements (ROREs) on *Bmal1* promoter. These DNA motifs recruit the REV-ERBs repressors involved in the rhythmic regulation of the histone acetylation state at target promoters. Indeed, the H3K27ac profile in the *Bmal1* promoter corresponded to that of its burst frequency. More generally, higher histone acetylation levels were observed during *Bmal1* circadian peaks of expression, while H3K27ac signal did not vary between clones harboring different reporter integration sites.

Similar properties were observed on other rhythmically expressed genes: despite variability in promoter motifs and expression phases, endogenous circadian genes displaying rhythmic variations in their promoter acetylation state also modulated their burst frequencies over the circadian period.

By inferring the transcriptional bursting parameters of non-circadian genes using smRNA-FISH datasets, we also observed significant correlations between histone acetylation signal around promoters and the burst frequency.

In conclusion, this study identified an association between the burst frequency and the histone acetylation state of promoters. While the molecular mechanisms behind this association remain elusive, it could be related to the facilitated binding of transcription regulators upon histone

acetylation-mediated chromatin loosening. In this thesis we clarified how transcription of circadian genes is rhythmically modulated, and we further elucidated the link between molecular events and transcriptional bursting, with particular focus on histone acetylation.

## Keywords

Gene expression noise, transcriptional bursting, histone acetylation, *Bmal1*, circadian, bio-luminescence, smRNA-FISH

# Résumé

Un important degré d'hétérogénéité est présent entre cellules isogéniques au sein d'un même environnement, et la stochasticité liée à la transcription participe de façon substantielle à ces variations entre cellules. Notamment, chez les mammifères, une majorité de gènes sont transcrits durant de courtes périodes appelées « rafales» de transcription (transcription bursts) et suivies par une période plus longue d'inactivité transcriptionelle. La grande variabilité présente entre les gènes en terme de fréquence et de taille de ces rafales suggère l'existence de différents mécanismes de régulation impliqués dans la formation de rafales spécifiques. Cependant, la nature des ces mécanismes moléculaires et leurs contributions exactes restent mal connues.

Dans cette thèse, j'ai utilisé un reporteur exprimant une luciférase éphémère sous le contrôle du gène circadien *Bmal1* intégré de manière stable dans le génome de fibroblastes de culture NIH-3T3 afin de mesurer les rafales transcriptionelles de ce gène au cours du cycle circadien et à trois sites d'intégration distincts. La quantification des variations dynamiques du signal luminescent dans des cellules individuelles ainsi que le comptage du nombre de transcrits par cellules à des phases précises permet de déduire les caractéristiques des rafales transcriptionelles dans ces différentes conditions en utilisant un modèle télégraphique (on et off) d'expression de gènes. Cette approche révéla que les variations d'expressions liées au site d'intégration du reporteur provenaient essentiellement de différences de taille des rafales, alors que leurs fréquences changeaient principalement au cours du cycle circadien. Ces deux paramètres sont donc découplés et peuvent être adaptés indépendamment afin de réguler le niveau d'expression.

En cherchant les origines moléculaires des rafales, nous avons observé que les variations de fréquences des rafales au cours du cycle circadien dépendaient de la présence de deux éléments de réponse à ROR (ROREs) sur le promoteur de *Bmal1*. Ces motifs d'ADN sont impliqués dans le recrutement des répresseurs REV-ERBs qui régulent la rythmicité de l'acétylation des histones. En effet, le taux d'acétylation des histones sur le promoteur de *Bmal1* correspondait à la fréquence de ses rafales, puisque les taux élevés d'acétylations étaient élevés durant les sommets d'expression, alors que ces mêmes taux d'acétylation ne variaient pas entre les clones dont le promoteur était intégré à différents endroits.

Des propriétés similaires furent relevées sur d'autres gènes circadiens : malgré des différences de motifs d'ADN aux promoteur ou de phases d'expression, chez les gènes circadiens endogènes dont les niveaux d'acétylation d'histones oscillaient, la fréquence des rafales variait également au cours de la période circadienne.

Le calcul des paramètres de rafales transcriptionelles correspondant à d'autres gènes non-circadiens par smRNA-FISH montra également des corrélations significatives entre les taux d'acétylation d'histones aux promoteurs et la fréquence des rafales.

Pour conclure, cette étude identifia un lien entre la fréquence des rafales et le taux d'acétylation du promoteur. Alors que l'implication de mécanismes moléculaires précis n'est pas encore établie, ce lien est potentiellement lié à l'attachement facilité des facteurs de régulation de la transcription sur l'ADN assoupli par la présence d'histones acétylées. Ce travail de thèse permet de mieux comprendre la manière dont la transcription des gènes circadiens est régulée rythmiquement, et les liens entre les évènements moléculaires (et tout particulièrement l'acétylation des histones) et les rafales de transcription.

## Mots-clés

Stochasticité d'expression génique, rafales (bursts) de transcription, acétylation d'histones, *Bmal1*, circadien, bioluminescence, smRNA-FISH

# Acknowledgements

I first would like to thank my thesis director Prof. Felix Naef for giving me the opportunity to work in his laboratory. Working under his supervision was a very stimulating experience that allowed me to experience state of the art technical approaches, and interact with plenty of brilliant minds. I greatly appreciated his calmness, his availability, his lucidity, and the help he provided at each steps of this thesis. Our interactions greatly enriched my scientific vision and my way of thinking.

I also would like to thank my thesis co-director Prof. David Suter. I could benefit from his experience and technical skills throughout my entire thesis. I also thank him for his investment in this project. The numerous ideas he came up with and directions he suggested largely contributed in shaping this work.

A warm thank you also goes to the present and past members of the Naef lab. It has always been a pleasure to work in such an enriching environment. I greatly appreciated socially interacting with all of you during lunch and coffee breaks, lab outings, aperos and many other events. It was also a great pleasure to share scientific experiences and point of view with talented people from different backgrounds and with various profiles. In particular, I would like to thank Dr. Benjamin Zoller, the computational master who helped me with the analysis of most of my data. His help was extremely precious, and mixing our two backgrounds resulted in a very enriching collaboration. Dr. Nick Phillips also substantially participated in the development of this project, and was always present when I needed explanations on mathematical concepts.

A very special thank you also goes to my co-workers of the wet lab Dr. Rosamaria Cannavo and Dr. Jérôme Mermet. You guys were a huge source of motivation, and turned the three years we shared in the lab into an unforgivable experience.

I also thank the members of the Suter lab for the great times spent together and for their help.

J'aimerais aussi remercier ma famille, qui m'a soutenu tout au long de mes études. Ils ont su faire preuve de beaucoup de compréhension et ont toujours été présents à mes côtés, notamment lors des périodes les plus éprouvantes de ma thèse.

Finalement, je remercie aussi infiniment Clémence Hurni pour la confiance qu'elle m'accorde, pour sa gentillesse, sa bonne humeur constante, l'immense aide qu'elle m'a apporté durant les derniers mois de ma thèse et la compréhension dont elle a su faire preuve lors des moments difficiles. Becs!

# Contributions to this work

In the project presented in this manuscript, I performed all bio-molecular experiments. However, several collaborators contributed in the development of the project, notably for the analysis of the data. This work could not have been completed without their precious collaborations.

First, Dr. Benjamin Zoller (currently in Thomas Gregor lab at Princeton University, New Jersey) analyzed all single-cell traces to infer the bursting parameters corresponding to the different experimental conditions. He also calculated the transcript and protein half-lives from luminescence decay measurements.

Dr. Nick E Phillips (Naef group) analyzed the smRNA-FISH transcripts distributions and inferred the likelihood for the corresponding transcriptional bursting parameters.

Finally, Jake Yeung (Naef lab) mapped the ChIP-seq reads on the mouse genome. The same data were also mapped to the transgene sequence with the help of Felix Naef, Colas Droin (Naef lab) and Elias Friman (Suter lab), while Dr. Eric Paquet assisted me with the post-mapping analysis of the data.

# Chapter 1    Introduction

Transcription is a multifaceted and tightly regulated process involving a great variety of cellular factors, complex biochemical reactions and a number sophisticated biophysical phenomena. To better understand my motivations to study, in depth, the expression pattern of a circadian gene at the transcriptional bursting level, I will first summarize general mechanistic concepts of transcription in eukaryotes and mention the key molecular factors involved. I will also explain the concept of noise in gene expression, notably by focusing on the contribution of transcriptional bursting. Finally, I will justify the choice of studying transcription in the circadian clock model system.

## 1.1    Transcription in eukaryotes

In virtually every living organism, the heredity material and the information required for the synthesis of most functional components of the cell are encoded on the DNA. Indeed, the genome contains sparse fragments of various length referred as genes that are further transcribed into RNA. RNA will then serve as template for protein synthesis, but also participate in the modification of various cellular molecules or regulate gene expression (Breaker and Joyce 2014). In eukaryotic systems, the complex task of transcription is carried out by the RNA polymerase (Pol) family of enzymatic complexes. Eukaryotic RNA polymerases comprises notably RNA PolI, specialized in the transcription of rRNA precursors (Russell and Zomerdijk 2006), and PolIII that produces short non-coding RNAs such as tRNAs or 5S ribosomal RNAs (Schramm and Hernandez 2002).

However, the most studied RNA polymerase is certainly PolII. Indeed, it is responsible for the transcription of virtually all protein coding genes as well as a majority of small non-coding RNAs and microRNAs. PolII-mediated transcription is a highly sequential process that undergoes several rate-limiting steps before leading to the release of a transcript (Fuda et al. 2009).

### 1.1.1    Transcription initiation

Enabling the access of PolII to the promoter is a complex task that starts with the combinatorial interaction of multiple transcription factors with DNA sequences and other transcriptional regulators. Indeed, genes are surrounded by regulatory DNA sequences located in close proximity to the promoter or at distal regions, and that serve as binding site for specific transcription factors called activators or repressors (Juven-Gershon et al. 2008).

The identity of these specific factors diverges considerably between different genes. Also, their molecular functions are multiple and include direct regulation of the transcriptional machinery or, more commonly, interactions with co-regulators involved in nucleosome reorganization, chromatin modification or the recruitment or regulation of other transcriptional complexes (Fuda et al. 2009; Voss and Hager 2014). In all cases, these specific transcription factors actively

participate in the recruitment of the pre-initiation complex (PIC) to the promoter (or prevent its binding in case of repressors).

Transcription regulation can also involve enhancers. These DNA sequences, typically located tens of kb away from the gene (Sanyal et al. 2012), form chromatin loops to physically contact the promoter (Shlyueva et al. 2014). These long-range interactions are transient (Fukaya et al. 2016; Chen et al. 2017; Siersbæk et al. 2017) and, as for proximal regulatory elements, enhancer-mediated expression regulation involves transcription factor recruitment, chromatin landscape modification or PolII delivery to the promoter (Beagrie and Pombo 2016). Thus, proximal and distal DNA sequences involved in the recruitment of specific transcription factors as well as the availability of the latter will greatly contribute to determining the expression level of the gene.

Consequently to the recruitment of activators, the PIC can assemble on the core promoter. In addition to PolII, this complex consists in several general transcription factors named TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH that cooperate for the recruitment and positioning of PolII on the promoter, the stabilization of the complex, the initiation of the transcription and the escape of PolII from the proximal promoter pausing (**Figure 1.1**). A typical assembly of the pre-initiation complex first consists in the recruitment of a TFIIF-PolII complex to a pre-formed TFIID-TFIIA-TFIIB complex binding the promoter (Shandilya and Roberts 2012; Sainsbury et al. 2015). TFIID, through its TATA-binding protein (TBP) subunit, plays an important role in the recognition and binding of the core promoter. This core PIC then recruits TFIIE and TFIIH, whose role is essential for transcription initiation as they initiate the promoter DNA opening. Simultaneously to the PIC assembly, activators recruit Mediator to the promoter. This large coactivator complex participate in the stabilization and assembly of the PIC, and favors PolII initiation (Allen and Taatjes 2015).
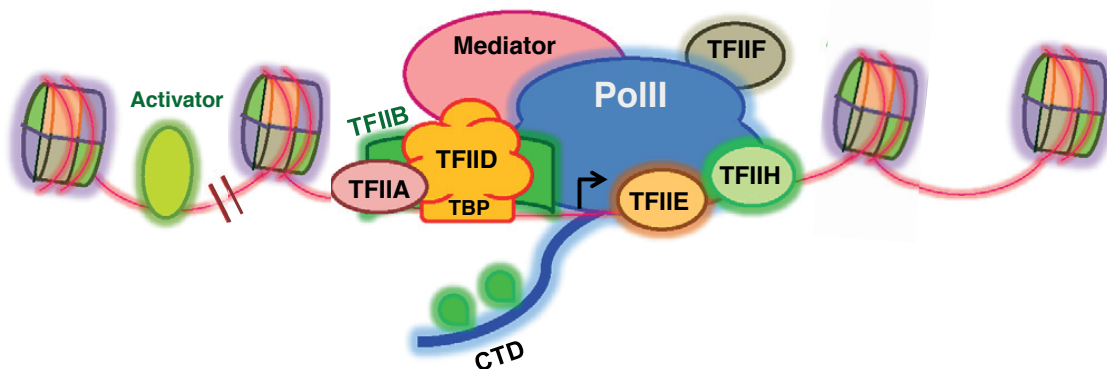


**Figure 1.1 Assembly of the pre-initiation complex (PIC)**
Schematic representation of the PIC assembled on the core promoter. The binding of an activator to a specific DNA target sequence favors the recruitment of the TFIID-TFIIA-TFIIB complex to the core promoter via the TBP sub-unit, followed by PolII and TFIIF, Mediator, TFIIE and TFIIH binding. Following the phosphorylation of its CTD at Serine 5 by TFIIH, PolII initiate transcription. The transcription start site is represented with an arrow. Modified from (Shandilya and Roberts 2012).

## 1.1.2 Transcription elongation and termination

Following the assembly of the PIC together with Mediator on the core promoter, DNA is unwound and PolII can initiate transcription. This process also involves the phosphorylation of the PolII

carboxy terminal domain (CTD) via a kinase subunit of TFIIH. The CTD, which consists in 52 tandem repeats of 8 amino acids, actively participates in transcription regulation. It is sequentially modified during key steps of the process, which coordinates the temporal recruitment of specific transcription-related complexes to PolII (Phatnani and Greenleaf 2006). Typically, shortly after transcription initiation, phosphorylated CTD Ser5 is recognized by capping enzymes to add the 5' cap to the nascent RNA.

If PIC assembly is a key step to control transcription, it is also tightly regulated after PolII transcription initiation. In a large fraction of mammalian genes, PolII transiently stops after transcribing 30-60 nucleotides (Core et al. 2008; Jonkers et al. 2014) and accumulates shortly downstream of the promoter (Quinodoz et al. 2014). This PolII proximal pausing is regulated by pause inducing factors (notably DSIF and NELF) that associate with PolII to block its progression, and the P-TEFb kinase that phosphorylates the pausing complex to resume transcription (Adelman and Lis 2012). The gene-specific pause duration, which notably depends on the promoter composition, typically ranges from 5 to 20 minutes in most genes but can also least up to one hour in specific cases (Shao and Zeitlinger 2017). This phenomenon likely participates in mediating quick and synchronized expression activation, notably in signal-response genes.

Upon proximal pausing release, PolII continues productive synthesis. The elongation rate is influenced by multiple factors such as the histone context, the GC content or gene features, and varies considerably along the gene body (Lenstra et al. 2016). Notably, splice sites slow the PolII transcription rate, leading to its accumulation in exonic regions (Jonkers et al. 2014). Thus, the recruitment of factors during elongation to perform co-transcriptional splicing may influence PolII velocity. Transcription elongation rate not only varies within but also between genes. In most tested genes, it is comprised between 2 and 5 kb per minute (Lenstra et al. 2016).

After completing the RNA synthesis, transcription terminates. This process often involves the presence of a polyA sequence in the DNA. Once PolII transcribes this motif, the complementary sequence in the nascent RNA, together with specific modifications of PolII CTD, recruit protein complexes involved in cleaving the ribonucleic chain and adding a poly(A) tail to its 3' end (Shandilya and Roberts 2012; Lenstra et al. 2016). This process destabilizes PolII, which eventually falls off the gene with the help of termination factors or exonucleases degrading the uncapped RNA strand synthetized from the polyA sequence (Rosonina et al. 2006). After its release from the DNA template, PolII CTD is dephosphorylated to recover its original form and be rapidly recycled to another or the same promoter. Indeed, complexes of general transcription factors such as TFIID-TFIIA-TFIIB can remain associated with the core promoter after PolII dissociation from the PIC to favor rapid reinitiation.

### 1.1.3   The role of histone modifications in transcription

An additional layer of complexity participating in transcription is the state of the chromatin around and within genes. Indeed, eukaryotic DNA is wrapped around histone octamers called nucleosomes, which are positioned approximately every 200bp along the genome (147pb directly wrapped around the nucleosomes and a linker of variable length). In parallel to their role in compacting DNA, these nucleosomes play an active role in transcription regulation through

histone modifications (Kouzarides 2007; Li et al. 2007). Indeed, nucleosomes are typically composed of 2 copies of histones H2A, H2B, H3 and H4. While the core of the histones around which DNA is wrapped is compact, their N-terminal domains, referred as histone tails, are unstructured. Several residues composing these tails, notably on histone H3, can be modified post-translationally. These modifications actively participate in determining the density of the chromatin environment and its permissiveness to transcription.

Histone post-translational marks typically consist in acetylation, methylation and phosphorylation although other modifications also exist. So far, around 550 possible modifications have been identified (Andrews et al. 2016). These marks are dynamic, and a large number of enzyme families are involved in their addition, removal or modification (Kouzarides 2007). Together, these histone marks coexist to form a code that contains information regarding specific structural and/or functional outcome.

The role of the histone code is diverse. Although some histone marks directly impact the chromatin structure by disrupting contacts within nucleosomes or between nucleosomes and DNA, most of the histone marks function through the recruitment of other non-histone proteins. These proteins typically contain domains involved in the targeting of specific histone modifications (or the lack of modifications) (Patel and Wang 2013). Thus, the histone code will encourage the binding of a set of proteins to specific nucleosomes. Once these proteins are recruited on chromatin, their enzymatic activities participate in diverse processes such as transcription, DNA repair, replication or condensation.

During transcription, the chromatin state and histone code play a decisive role at every possible step. Indeed, the DNA organization into chromatin presents an obstacle for the binding of several factors involved in transcription and for PolII elongation along the gene body. Experiments such as ChIP-seq greatly contributed to improve our understanding on the histone code function during transcription (Barski et al. 2007; O'Geen et al. 2011; Ho et al. 2014). Indeed, the presence of specific histone marks around gene features can efficiently predict its transcriptional output (**Figure 1.2**).

Among all possible histone modifications, the role of acetylation is particularly well understood. Indeed, this histone mark actively participates in chromatin destabilization by neutralizing the positive charge of the lysine. This modification facilitates nucleosomes disassembly and eviction, therefore favoring gene expression (Simon et al. 2011; Di Cerbo et al. 2014). Consequently, histone acetylation marks are virtually always associated with active transcription. Notably, H3K27ac is often used as a marker of active enhancers, and is also enriched in the promoter region of active genes (Tie et al. 2009; Creyghton et al. 2010). Similarly, H3K9ac is also tightly correlated with actively transcribing promoters and is associated with reduced nucleosome density (Nishida et al. 2006). Acetylation on the H4 histone also plays an essential role in regulating chromatin structure since it disrupts the interactions between adjacent nucleosome required for the formation of a condensed state (Shogren-Knaak 2006).
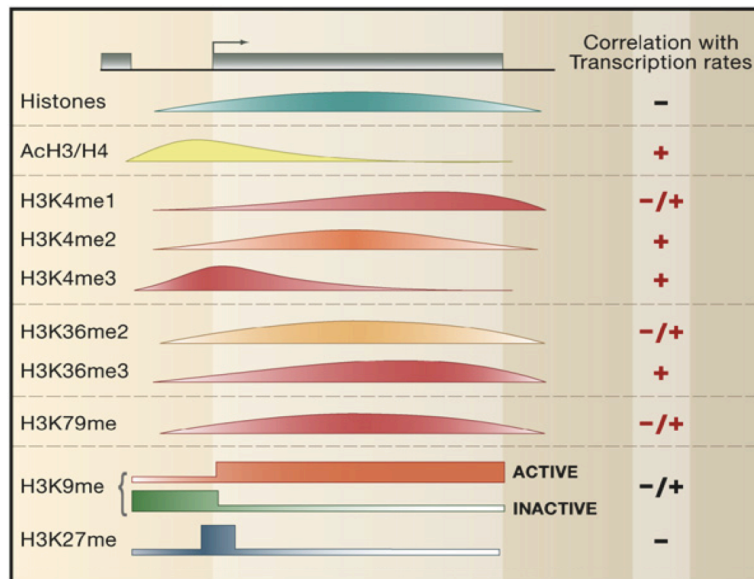
**Figure 1.2 Post-translational modifications of H3 lysine residues and their impact on transcription**
Non-exhaustive list of possible post-translational modifications observed on lysine residues of the N-terminal tail of histone H3. Marked histones can be detected in different regions of the gene such as the enhancers, the proximal promoter or the gene body. The various histone marks are associated with specific permissiveness to transcription: reduced transcription (-), improved transcription (+) or both cases observed/neutral (+/-). Modified from (Li et al. 2007).

Methylation of the lysine residues present on histone H3 tail is on the other hand more complex and can be both associated with transcription activation or repression. Two of these, H3K27me3 and H3K9me3 have a predominant role in repression and are commonly associated with heterochromatin. The methylation of H3K27 is indeed catalyzed by the Polycomb complex involved in DNA compaction, notably in developmental genes (Francis et al. 2004; Di Croce and Helin 2013). H3K9me3, on the other hand, recruits heterochromatin protein 1 (HP1) to the chromatin (Azzaz et al. 2014; Hiragami-Hamada et al. 2016). HP1 then associates with adjacent nucleosomes to condensate chromatin. But histone methylation can also be associated with active transcription. For example, both H3K36me3 and H3K4me3 are found in the coding region of transcribed genes. H3K36me3, which is observed along the entire transcribed region, follows elongating PolII to avoid accidental transcription initiation in this permissive chromatin region (Carrozza et al. 2005). H3K4me3 however is predominantly observed in the 5' end and promoter region of actively transcribed genes (Bernstein et al. 2005; Heintzman et al. 2007). The monomethylated form of H3K4 is also interesting as it is often found in enhancer regions (Heintzman et al. 2007).

This complex histone code largely participates in organizing chromatin in domains with specific histone marks combinations, chromatin compaction levels and precise functions, notably in transcription (Filion et al. 2010; Ernst et al. 2011). Often, binding sites for specific transcription factors are positioned within accessible regions, whether it is in nucleosome-free or in exposed regions at the surface of the nucleosomes (Yuan et al. 2005; Albert et al. 2007). However, once

bound to DNA, they typically tether a cascade of co-activators including histone modification enzymes and chromatin remodeling complexes (Voss and Hager 2014). These co-activators establish a favorable environment to facilitate binding of additional co-activators or general transcription factors, allowing the proper loading of the PIC on the core promoter and PolII elongation trough a permissive chromatin environment.

## 1.2    Stochasticity in transcription

After mentioning general concepts relative to transcription and the multiple factors involved, I will focus on noise in gene expression and notably the role of transcriptional bursting. Indeed, transcription is a highly stochastic process and a major cause of variability between identical cells. This biophysical phenomenon could only be monitored using technical approaches developed during the past 15 years. Thus, crucial aspects regarding the origins of transcriptional bursting, its functions, implications and consequences on the cell remain to be determined. In this section, I will explain the concept of transcriptional bursting in greater details. Then, I will explain the technical approaches often chosen to monitor and measure this cellular phenomenon. Finally, I will summarize the knowledge accumulated over the past decade on how molecular mechanisms participate in defining transcriptional bursting parameters. This whole section is largely inspired from (Nicolas et al. 2017), a review that I recently wrote on the role of various molecular mechanisms in modulating transcription by influencing the burst size and/or the burst frequency.

### 1.2.1    Transcriptional bursting and the telegraph model

In most biological systems, genetically identical cells in a common environment display great variability in the expression levels of their gene products (Raj and van Oudenaarden 2008; Sanchez and Golding 2013) (**Figure 1.3A**), and transcription is a source of this gene expression noise (Raj and van Oudenaarden 2008; Eldar and Elowitz 2010). Indeed, as previously mentioned, this cellular process consists in complex and tightly ordered sequences of biochemical reactions (Coulon et al. 2013; Voss and Hager 2014), and phenomena such as low molecular concentrations, diffusion or transcription factor dynamics naturally bestow randomness on it (Elowitz et al. 2002; Paulsson 2004; Pedraza and Paulsson 2008; Schoech and Zabet 2014). Part of this variability can be explained by factors extrinsic to the gene itself, such as the cell size, the cell-cycle state, or the concentration of certain factors involved in transcription (Raser and O'Shea 2004; Rosenfeld et al. 2006; Rinott et al. 2011; Zopf et al. 2013). However, an additional important source of transcriptional noise arises from genes being transcribed irregularly, with RNA production greatly fluctuating over time (**Figure 1.3B**). Indeed, notably in mammalian systems, RNA synthesis is often subject to a pulsatile pattern and occurs mainly during short, often intense periods referred as transcriptional bursts followed by longer periods of transcription inactivity (Raj and van Oudenaarden 2008; Larson 2011). Consequently, transcript distributions of many genes are too widely spread and cannot be explained by Poissonian distributions implying mRNA production at a constant rate and mRNA degradation proportional to the number of mRNA molecules (**Figure 1.3C**, grey distribution). Thus, a widely used model to account for transcriptional bursting

proposes that the promoter stochastically switches between active and inactive transcriptional states (**Figure 1.3D**) (Golding et al. 2005; Chubb et al. 2006; Raj et al. 2006; Paré et al. 2009). Mathematically this can be represented with a two-states "telegraph" model of gene expression, which assumes that the promoters can be in two different states: a transcriptionally active "on" state, or a silent "off" state characterized by a lack of transcriptional activity (Peccoud and Ycart 1995)**.** This model can produce mRNA distributions with a variety of shapes and typically more variance than Poisson distributions (**Figure 1.3C**, blue distribution) (Shahrezaei and Swain 2008; Mugler et al. 2009; Munsky et al. 2012). The telegraph model has one parameter describing the rate of mRNA degradation $\gamma_m$, and three parameters describing the rate of mRNA production: the rate of switching from an "off" to an "on" state and vice versa ($k_{on}$ and $k_{off}$, respectively), and $k_m$ the average rate of mRNA transcription while in the "on" state (**Figure 1.4**).
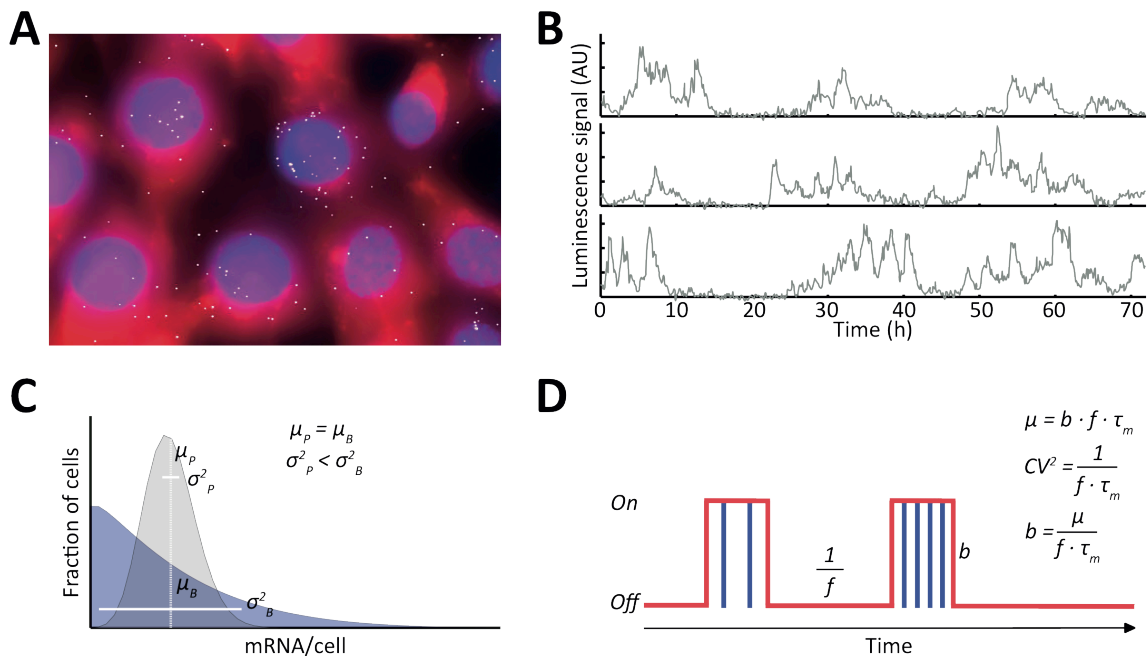


**Figure 1.3 Transcriptional bursting as a source of cell-to-cell variability**

(**A**) Illustration of heterogeneity in gene expression in fixed cells. Isogenic NIH-3T3 cells display large variability in their cellular number of transcripts (white dots), as exemplified here with smRNA-FISH labeling *Bmal1* mRNA. Cells are labeled in red (HCS CellMask), and nuclei in blue (DAPI). (**B**) Real-time monitoring of discontinuous transcription using a *Bmal1* destabilized bioluminescence reporter (Blanchoud et al. 2015). Each trace represents a single-cell tracked over three days. (**C**) Comparison of transcripts per cell in a population between a discrete probability distribution (Poisson, grey) and a stochastic regime (Bursting, blue). When the mean number of transcript per cell is the same ($\mu_P = \mu_B$), the variance is larger in the bursting condition ($\sigma_P < \sigma_B$). (**D**) Schematization of the promoter activity of a bursting gene. The promoter switches between active (*On*) or silent (*Off*) transcriptional states. RNAs production (blue bars) only occurs during the active periods and defines the burst size *b*. In a bursting regime where the "on" states are considerably shorter than the "off" states and produce an important amount of transcripts, the average number of mRNA per cell $\mu$ is the product of the burst site *b*, the burst frequency *f* and the mRNA half-life $\tau_m$. The burst frequency is inversely proportional to the length of the silent period, and can be expressed as the inverse of the coefficient of variation $CV^2$. The burst size is expressed as the mean number of transcripts divided by the frequency. Figure from (Nicolas et al. 2017).

Transcriptional bursting, together with additional probabilistic processes underlying gene expression such as post-transcriptional regulation (Battich et al. 2015) or translation (Albayrak et al. 2016), will actively contribute to generate diversity in isogenic cells (Symmons and Raj 2016). Interdisciplinary approaches have integrated quantitative measurements of gene expression with mathematical models in order to understand the origins and consequences of transcriptional bursting. Measurements of gene expression can be broadly categorized as either static or time resolved (often also termed "live"), and the mathematical approach will depend on the type of data used as input. Experimental and mathematical approaches that allow the obtaining of mechanistic insight from gene expression data will be describe in the following section.



**Figure 1.4 Graphical representation of the telegraph model**
The promoter can switch between in a transcriptionally repressed state (red) to an active state (green) with a $k_{on}$ rate. During the active state, mRNA is produced with a transcription rate of $k_m$. mRNA also undergoes degradation with a $\gamma_m$ rate. The return to a repressed promoter state occurs with a rate of $k_{off}$. Modified from (Teles et al. 2013).

## 1.2.2 Methods and models to monitor bursting

Over the past years, several emerging experimental approaches have allowed the monitoring of transcriptional bursting (Raj and van Oudenaarden 2009; Larson et al. 2009; Lionnet and Singer 2012). Notably, transcription can be "directly" monitored in real-time using the MS2-GFP approach. After introduction of specific stem loops in the transcripts to recruit fluorescently tagged viral capsid proteins, nascent RNAs at the transcription site (TS) appear as a fluorescent dots whose intensity fluctuates with the promoter activity (Chubb et al. 2006; Larson et al. 2011). Although this approach greatly improved over the recent years (Ochiai et al. 2014; Tantale et al. 2016), its application to mammalian systems remains delicate and comprises technical challenges such as the generation of a stable cell line carrying an appropriate form of the reporter (Boireau et al. 2007; Lionnet et al. 2011; Yunger et al. 2013), or the maximization of the signal to noise ratio in order to reliably detect the transcription spots (Yunger et al. 2010; Suter et al. 2011a).
An alternative to MS2-GFP for real-time monitoring of transcriptional bursting consists in the use of destabilized reporters. Although the detected signal arises from the protein instead of the transcript, short-lived luminescent (Harper et al. 2011; Suter et al. 2011a; Molina et al. 2013) or fluorescent reporters (Harper et al. 2011; Dar et al. 2012) were both shown to be a powerful tool for the study of transcription in living cells. Indeed, upon promoter activation, sporadic signal can be detected with limited decay (**Figure 1.5**). The transcriptional bursting parameters can be inferred by calculating the likelihood of moving between successive time points in the time series given the telegraph parameters (Bronstein et al. 2015). Live-cell imaging methods have elucidated transcriptional mechanisms that are difficult to quantify in fixed cells, such as the existence of the refractory period in the "off" state preventing promoters from reactivation shortly after the

preceding burst (Harper et al. 2011; Suter et al. 2011a; Zoller et al. 2015). In addition to the two-state telegraph model, the analytical methods have also been extended to consider multiple intermediate inactive states before expression reactivation (Zoller et al. 2015) and discontinuous transitions between multiple different levels of transcriptional activity (Innocentini et al. 2013; Hey et al. 2015; Featherstone et al. 2016; Corrigan et al. 2016).
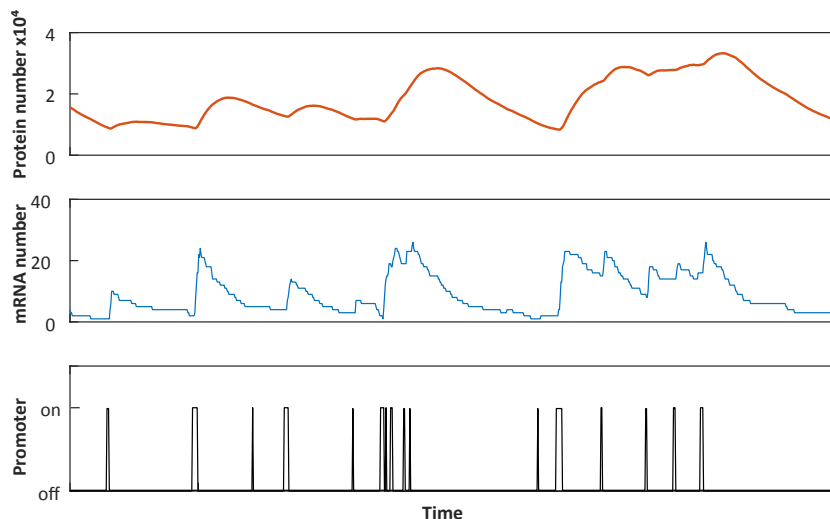


**Figure 1.5 Concept of the short-lived reporter**

In a short-lived reporter system, the gene switches between the active "on" and inactive "off" state (bottom panel). mRNA production only occurs during the "on" state (middle panel), shortly followed by the protein (upper panel). Because the reporter is destabilized both at the mRNA and protein level, the state of the gene at each time-point can be inferred from the quantified pulses of reporter protein. Simulation by Dr. Nick E Phillips.

Alternatively, the telegraph model can also be used to estimate parameters by fitting measurements of mRNA copy number at a single-cell level, predominantly using single-molecule RNA FISH (smRNA-FISH) (Raj et al. 2006; Senecal et al. 2014; Bahar Halpern et al. 2015b; Padovan-Merhar et al. 2015; Skinner et al. 2016). smRNA-FISH consists in the *in situ* labeling of single RNA molecules by fluorescently labeled nucleic acid probes whose sequences are complementary to the target. Originally developed with a limited number of probes harboring multiple fluorophores (Femino et al. 1998), the technique nowadays usually uses a higher number of shorter probes recognizing different portions of the transcript (typically 20 to 50 probes of 18-20 nucleotides per transcript), and each harboring a single fluorophore to limit the detection of false positives (Raj et al. 2008). The co-localization of several sequence-specific probes onto the target transcript enables its detection as a diffraction-limited spot using conventional widefield fluorescence microscopes (**Figure 1.6**). In addition to the count of transcripts in a cell, this approach also provides the cellular location of the RNA molecules. smRNA-FISH was successfully applied to culture cells (Raj et al. 2006, 2008; Singer et al. 2014), tissue slices (Itzkovitz et al. 2011; Bahar Halpern et al. 2015b) and thin organisms (Lécuyer et al. 2008; Raj et al. 2008; Ji et al. 2013). Although several strategies were further developed using DNA adaptors to increase the number of binding probes per target molecule, they suffer reduced penetrance into some cellular compartments including the nucleus due to the larger size of the adaptors (Player et al. 2001;

Battich et al. 2013). Although this imaging strategy only applies to fixed cells and thus loses the dynamic aspects of the previously mentioned approaches, it can still be used to infer transcriptional bursting parameters. Indeed, by only assuming that on-states are brief on the scale of the transcript life-time, the steady state distribution becomes a negative binomial distribution (Raj et al. 2006), whose parameters can be readily estimated using maximum likelihood. Similarly, the telegraph model can also be fitted using single-cell RNA-seq data (Kim and Marioni 2013).
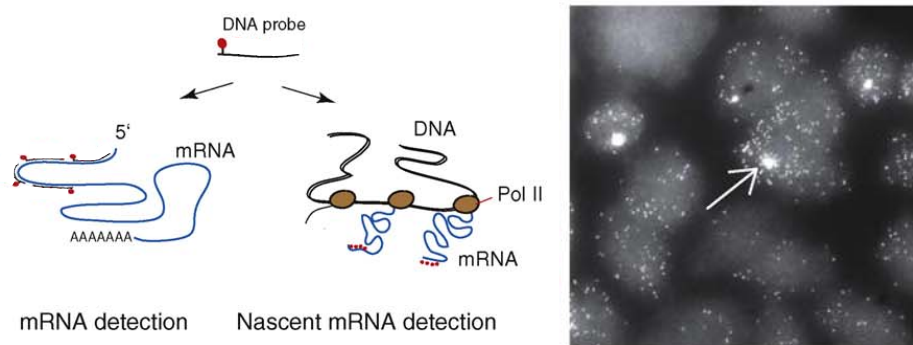


**Figure 1.6 Concept of single-molecule RNA FISH (smRNA-FISH)**
Fluorescently labeled DNA probes specifically anneal to different region of the target mRNA, resulting in diffraction limited fluorescent dot. Nascent mRNA is also labeled upon synthesis of the target region. Because of the large amount of transcripts simultaneously synthetized during bursts, active transcription site appear as large. Modified from (Larson et al. 2009).

### 1.2.3 Molecular mechanisms involved in shaping transcriptional bursting

These different approaches highlighted the frequent character of transcriptional bursting, which was observed in virtually every organism, from prokaryotes to yeasts and higher eukaryotes. While in bacteria pulsatile transcription seems to be a phenomenon of rare occurrence (Elowitz et al. 2002; Taniguchi et al. 2010; So et al. 2011) that possibly directly arises from the formation of positive supercoiled DNA following the passaging of RNA Polymerase (Chong et al. 2014; Sevier et al. 2016), it seems to be more widespread in systems with higher levels of chromatin organization. Notably, in higher eukaryotes transcriptional bursting appears to be a quasi-universal phenomenon governing the expression of a majority of genes observed in cultured cells (Norris et al. 2003; Raj et al. 2006; Suter et al. 2011a), tissues (Bahar Halpern et al. 2015b) and small organisms (Muramoto et al. 2012; Little et al. 2013; Ji et al. 2013). These studies highlighted that different genes can display dramatically different transcriptional bursting kinetics (Suter et al. 2011a; Muramoto et al. 2012; Singer et al. 2014; Bahar Halpern et al. 2015b; Skinner et al. 2016). A convenient way to describe the transcriptional bursting behavior of a gene is to refer to its burst frequency (i.e. the number of bursts in time units) and its burst size (i.e. the mean number of transcripts produced per burst episode). Typically, across mammalian systems, the burst frequency of expressed genes ranges from a burst every 30 minutes to up to 10 hours, and the burst size from one to several hundreds of transcripts (Lionnet and Singer 2012). This diversity in transcriptional bursting profiles likely reflects the complexity of gene regulation and the variety of molecular mechanisms involved in tuning gene expression at the transcriptional level. However,

how molecular aspects of transcription directly influence the bursting specificities of the genes has long remained elusive (Suter et al. 2011b; Lenstra et al. 2016). Over the past years, considerable efforts were made to modulate transcriptional bursting and identify its molecular mechanisms in higher eukaryotic systems (**Figure 1.7**) (**Table 1.1**).
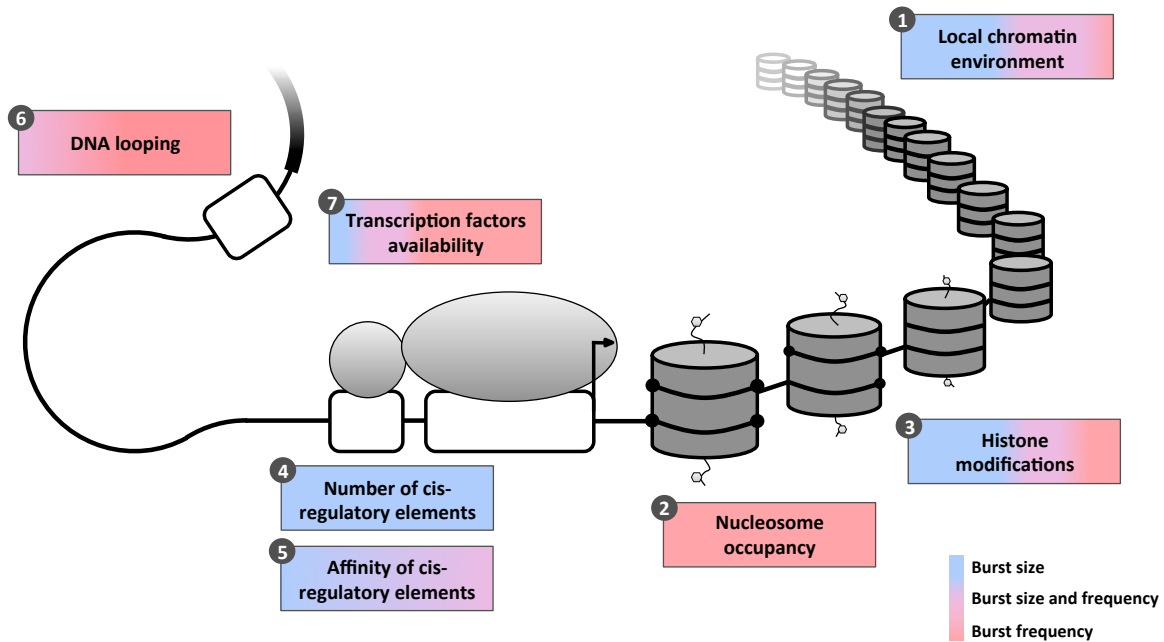


**Figure 1.7 Molecular mechanisms regulating transcriptional bursting**
Schematic representation of a transcribed gene, including DNA (black line), a core promoter (large white box), cis-regulatory elements (small white boxes), the TSS (arrow), a specific transcription factor (small grey sphere), the pre-initiation complex (large grey oval) and nucleosomes (grey cylinders). Text boxes highlight specific molecular mechanisms participating in transcriptional bursting modulation. Their coloration represents their relative role on tuning the burst size (blue), the burst frequency (red) or both (purple). Colors proportions refer to the number of studies referenced in Table 1.1. Figure taken from (Nicolas et al. 2017)

Despite the variety of experimental systems probed, a majority of studies assessing changes in transcription dynamics upon stimulation demonstrated an important effect on burst frequency (Singh et al. 2010; Larson et al. 2013; Senecal et al. 2014; Bahar Halpern et al. 2015b). These experimental approaches typically activate cellular pathways that eventually lead to changes in the availability of transcription factors involved in initiating early steps of transcription. The binding of these transcription factors to DNA can occur close to the TSS or at distal region, leading to the formation of DNA loops, also identified as regulators of burst frequency (Bartman et al. 2016; Fukaya et al. 2016; Chen et al. 2017). Thus, the burst frequency may be directly proportional to the concentration of transcription activators. This would be consistent with observations performed on the presence of transcription factors on gene promoters at steady state in the *Drosophila* embryo: fitting *hunchback* expression levels with the nuclear concentration of its activator BICOID indeed revealed that the regulation could be achieved by only affecting burst frequency (Xu et al. 2015)

| Promoter | Cell type | Monitoring approach | Experimental condition | Consequence on bursting | Reference |
|---|---|---|---|---|---|
| **1) Local chromatin environment** | | | | | |
| HIV-1 LTR | Jurkat | Flow cytometry | Random integration | Burst size | (Singh et al. 2010) |
| HIV-1 LTR | Jurkat | Flow cytometry | Random integration | Burst size | (Skupsky et al. 2010) |
| Synthetic GRE | U2OS | MS2-GFP | Random integration | Burst size | (Larson et al. 2013) |
| Synthetic CCAAT-box | NIH-3T3 | Short-lived protein reporter | Random integration | Burst size | (Zoller et al. 2015) |
| HIV-1 LTR | Jurkat | Short-lived protein reporter | Random integration | Burst size and frequency | (Dar et al. 2012) |
| HIV-1 LTR | Jurkat | smRNA-FISH, Flow cytometry | Random integration | Burst size and frequency | (Dey et al. 2015) |
| CMV | 6C2 | Flow cytometry | Random integration | Burst frequency | (Viñuelas et al. 2013) |
| **2) Nucleosome occupancy** | | | | | |
| HIV-1 LTR | Jurkat | smRNA-FISH | DNase hypersensitivity assays | Burst frequency | (Dey et al. 2015) |
| **3) Histone modifications** | | | | | |
| Synthetic CCAAT-box | NIH-3T3 | Short-lived protein reporter | TSA treatment | Burst size | (Suter et al. 2011a) |
| prl2c2 | NIH-3T3 | Short-lived protein reporter | TSA treatment | Burst size | (Suter et al. 2011a) |
| hPRL | GH3 | Short-lived protein reporter | TSA treatment | Burst size | (Harper et al. 2011) |
| HIV-1 LTR | Jurkat | Short-lived protein reporter | TSA treatment | Burst size | (Dar et al. 2012) |
| Genome-wide | hESC | Single-cell RNA-seq | H3K4me2 ChIP-seq correlation | Burst size | (Wu et al. 2017) |
| act5 | *Dictyostelium* | MS2-GFP | H3K4 methyltransferase mutants | Burst size and frequency | (Muramoto et al. 2010) |
| arntl | NIH-3T3 | Short-lived protein reporter | TSA treatment | Burst frequency | (Suter et al. 2011a) |
| CMV | 6C2 | Flow cytometry | TSA treatment | Burst frequency | (Viñuelas et al. 2013) |
| Genome-wide | hESC | Single-cell RNA-seq | H3K36me3 ChIP-seq correlation | Burst frequency | (Wu et al. 2017) |
| Genome-wide | hESC | Single-cell RNA-seq | H3K79me2 ChIP-seq correlation | Burst frequency | (Wu et al. 2017) |
| Genome-wide | hESC | Single-cell RNA-seq | H4K20me1 ChIP-seq correlation | Burst frequency | (Wu et al. 2017) |
| **4) Number of cis-regulatory elements** | | | | | |
| Synthetic tetO | CHO | smRNA-FISH | 1 or 7 tTA binding sites | Burst size | (Raj et al. 2006) |
| Synthetic CCAAT-box | NIH-3T3 | Short-lived protein reporter | 1 or 2 NF-Y binding sites | Burst size | (Suter et al. 2011a) |
| c-Fos | U2OS | smRNA-FISH | 1 or 4 TALE binding sites | Burst size | (Senecal et al. 2014) |
| **5) Affinity of cis-regulatory elements** | | | | | |
| act5 | *Dictyostelium* | MS2-GFP | TATA-box mutant | Burst size | (Corrigan et al. 2016) |
| Synthetic CCAAT-box | NIH-3T3 | Short-lived protein reporter | CCAAT-box mutants | Burst size and frequency | (Suter et al. 2011a) |
| **6) DNA looping** | | | | | |
| β-globin | Hemato precursor | smRNA-FISH | Looping in erythroid maturation | Burst size and frequency | (Bartman et al. 2016) |
| β-globin | G1E-ER4 | smRNA-FISH | Forced promoter-enhancer looping | Burst frequency | (Bartman et al. 2016) |
| sna | *Drosophila* embryo | MS2-GFP | Substitution of distal enhancers | Burst frequency | (Fukaya et al. 2016) |
| eve | *Drosophila* embryo | MS2-GFP | Forced promoter-enhancer looping | Burst frequency | (Chen et al. 2017) |
| **7) Transcription factors availability** | | | | | |
| Synthetic tetO | CHO | smRNA-FISH | Doxycycline-modul. tTA availability | Burst size | (Raj et al. 2006) |
| ctgf | NIH-3T3 | Short-lived protein reporter | TGF-β stimul. and serum induction | Burst size | (Suter et al. 2011a) |
| HIV-1 LTR | Jurkat | Short-lived protein reporter | TNF-α stimulation | Burst size and frequency | (Dar et al. 2012) |
| nanog | Bruce4 mESCs | MS2-GFP | Induction in 2i medium | Burst size and frequency | (Ochiai et al. 2014) |
| β-actin | MEFs | MS2-GFP | Serum induction | Burst size and frequency | (Kalo et al. 2015) |
| pck1 and g6pc | Mouse liver | smRNA-FISH | Fasting | Burst size and frequency | (Bahar Halpern et al. 2015b) |
| cyclinD1 | HEK293 | MS2-GFP | Wnt3a stimulation | Burst size and frequency | (Kafri et al. 2016) |
| HIV-1 LTR | Jurkat | Flow cytometry | TNF-α stimulation | Burst frequency | (Singh et al. 2010) |
| Synthetic GRE | U2OS | MS2-GFP | Steroid induction | Burst frequency | (Larson et al. 2013) |
| nanog | E14 mESCs | smRNA-FISH, protein reporter | Induction in 2i medium | Burst frequency | (Singer et al. 2014) |
| c-Fos | U2OS | smRNA-FISH | Zinc or serum induction | Burst frequency | (Senecal et al. 2014) |
| hunchback | *Drosophila* embryo | smRNA-FISH | Immunofluorescence of Bcd | Burst frequency | (Xu et al. 2015) |
| HIV-1 LTR | Jurkat | Short-lived protein reporter | TNF-α stimulation | Burst frequency | (Dar et al. 2016) |
| Erg, Gfi1b, Hhex, Mpl | HPC7 | Single-cell qPCR | Gfi1 overexpression | Burst frequency | (Ezer et al. 2016) |
| Procr and Gfi1b | HPC7 | Single-cell qPCR | Gata2 knock-down | Burst frequency | (Ezer et al. 2016) |

**Table 1.1: Experimental modulation of transcriptional bursting in higher eukaryotes genes**
List of genes with modulated bursting kinetics between experimental conditions. The list is divided into seven types of molecular mechanisms influencing transcriptional bursting and corresponding to the ones described in Figure 1.7. Each entry comprises a specific promoter, the transcriptional bursting monitoring approach, the cellular system used, the type of experimental conditions tested, the aspect of transcriptional bursting predominantly modulated between the conditions, and the reference of the original study. Modified from (Nicolas et al. 2017).

Binding of transcription activators to specific regions on the promoter may also be linked to the nucleosome clearance at the TSS. Indeed, the low nucleosome occupancy found to be linked with high burst frequency in both yeast (Brown et al. 2013; Dadiani et al. 2013) and mammalian cells (Dey et al. 2015) could result from the cascade of sequential events following transcription factor

binding occurring at the promoter to initiate transcription. In addition to modulating the process of bursting at the frequency level, transcription factors can also, in some cases, participate in tuning bursting intensity by additionally affecting the burst size (Dar et al. 2012; Ochiai et al. 2014; Kalo et al. 2015). This possibility could be specific to some transcription factors, or reflect a more general complementary mode of transcription regulation when burst frequency reaches an upper limit and higher expression levels can only be achieved by modulating alternative bursting parameters (Dar et al. 2012). Molecularly, this phenomenon could arise from saturating concentrations of transcription factor around the gene, provoking quasi-immediate reformation of the transcription initiation complex following its detachment after a transcription event.

In contrast to the factors involved in transcription activation and mainly influencing burst frequency, others, notably DNA regulatory elements of the promoter, predominantly modulate the size of the bursts (Raj et al. 2006; Suter et al. 2011a; Senecal et al. 2014; Corrigan et al. 2016). Indeed, the sequence of DNA regulatory elements influences transcription factor residency time rather than availability. Transcription factor residency time on DNA will stabilize transcription initiation complexes, thus allowing the production of a higher number of transcripts before it detaches from the promoter and provokes the switch back into the inactive state.

Aside from transcription factors, the crucial role of nucleosomes in shaping transcriptional bursting was unambiguously demonstrated by noticing absence of bursting when a gene was expressed from a plasmid with no chromatin context (Larson et al. 2013). The general chromatin state of the gene, mainly assessed by random insertion of the same reporter at different genomic locations, primarily influences the burst size in both yeasts (Batenchuk et al. 2011) and mammalian cells (Singh et al. 2010; Skupsky et al. 2010; Dar et al. 2012; Dey et al. 2015). Rather than participating in transcription initiation, the local chromatin environment is more likely involved in facilitating transcription efficiency once it is already initiated, therefore increasing the transcription yield in every on-phase. This is compatible with the recurrent pausing of PolII in proximity of nucleosomes observed in yeasts (Churchman and Weissman 2011).

However, it is not yet possible to link specific histone modifications with a particular pattern of transcriptional bursting. Indeed, similar histone marks were shown to display various bursting responses that seem to be gene-specific. Notably, histone deacetylases inhibitors could increase global expression levels by modulating both the burst size or the burst frequency (Harper et al. 2011; Suter et al. 2011a; Dar et al. 2012), while acetylation marks in a genome wide study were not found to specifically correlate with either of these bursting parameters (Wu et al. 2017).

Although the molecular mechanisms listed here have been the most investigated ones in the context of bursting regulation, the participation of other phenomenon is not excluded. This could notably embrace transcriptional pausing or additional processes related to PolII elongation (Suter et al. 2011b; Lionnet and Singer 2012; Lenstra et al. 2016). Also, while size and frequency are widely used parameters to describe transcriptional bursting, these concepts are approximations that satisfyingly permit a quantitative description of the stochastic transcription process, and the impact of the molecular mechanisms listed here could possibly be better captured by alternative descriptions of the transcription process.

## 1.3    The mammalian circadian clock system

So far, studies assessing transcriptional bursting in real-time in higher eukaryotes have mainly focused on isolated genes (Paré et al. 2009; Harper et al. 2011; Ochiai et al. 2014; Kafri et al. 2016), or groups of uncorrelated genes (Suter et al. 2011a). The mammalian circadian clock offers the opportunity to study gene expression in an endogenously dynamic system composed of various genes with different modes of regulation, and has thus been widely studied to improve our understanding of this crucial cellular process at every possible regulatory step (Mermet et al. 2017). The circadian clock is indeed a model system for the study of transcription factor binding (Rey et al. 2011; Sobel et al. 2017), transcription (Le Martelot et al. 2012; Koike et al. 2012), histone modifications (Ripperger and Schibler 2006; Feng et al. 2011), post-transcriptional regulation (Morf et al. 2012; Du et al. 2014) or translation (Atger et al. 2015; Janich et al. 2015). The massive amount of knowledge accumulated greatly facilitates the understanding of less understood phenomenon (such as transcriptional bursting) studied in the same system.

In the following chapter, I will briefly explain the concept of circadian clock and the molecules involved in its generation and maintain. I will then further focus on the knowledge accumulated over the past years on the transcriptional regulation of this system, and I will finally detail the transcriptional regulation of *Bmal1*, a core clock gene whose transcriptional bursting behavior is assessed in detail in this study.

### 1.3.1    The core clock

From Cyanobacteria to animals, plants and fungi, life on earth evolved with the 24 hour rotation of the planet. To anticipate daily changes in their environment such as light/dark cycle, temperature variation or food availability, organisms developed an internal timing system called circadian clock (Hastings et al. 2003; Bell-Pedersen et al. 2005). In mammals, the circadian rhythm regulates several behavioral and physiological aspects. Among those are the sleep/wake cycle, hormonal secretions, the body temperature and numerous metabolic processes (Panda 2016).

Synchronization of the circadian clock mainly occurs through environmental cues. Indeed, although the organism endogenously generates its own rhythm of around 24 hours, it is constantly entrained by external signals called Zeitbegers that adjust its phase and period. The main Zeitgeber is light. The 20'000 neurons in the anterior part of the hypothalamus composing the suprachiasmatic nucleus (SCN) integrate the light signal from the photoreceptors of the retina trough the retinohypothalamic tract to synchronize their own clocks (Dibner et al. 2010; Welsh et al. 2010). The SCN will in turn transmit the information to peripheral clocks through systemic cues such as hormones (notably glucocorticoids), metabolites, body temperature or the sympathetic nervous system (Hastings et al. 2003; Mohawk et al. 2012). Indeed, most cells of the organism contain their own cell-autonomous and self-sustained oscillator (Yoo et al. 2004; Nagoshi et al. 2004). In addition to the synchronization through the SCN central pacemaker, peripheral clocks can also be sensitive to additional cues that participate in regulating their circadian clocks. For example, rhythmic expression in the liver is largely influenced by feeding (Damiola et al. 2000; Stokkan et al. 2001; Vollmers et al. 2009).

In individual cells, the circadian clock consist in a complex interplay between transcriptional-translational feedback loops oscillating with a periodicity of ~24 hours (Partch et al. 2014). The canonical feedback loop is composed of the two basic helix-loop-helix (bHLH) activators BMAL1 and CLOCK. These proteins form a heterodimer and activate the expression of downstream genes by binding E-boxes in their promoter regions (Gekakis et al. 1998; Ripperger and Schibler 2006; Rey et al. 2011) (**Figure 1.8**). Among CLOCK-BMAL1 targets, Cryptochromes (CRY1 and CRY2) and Periods (PER1, PER2 and PER3) act as repressors for the same loop: upon translation, CRYs and
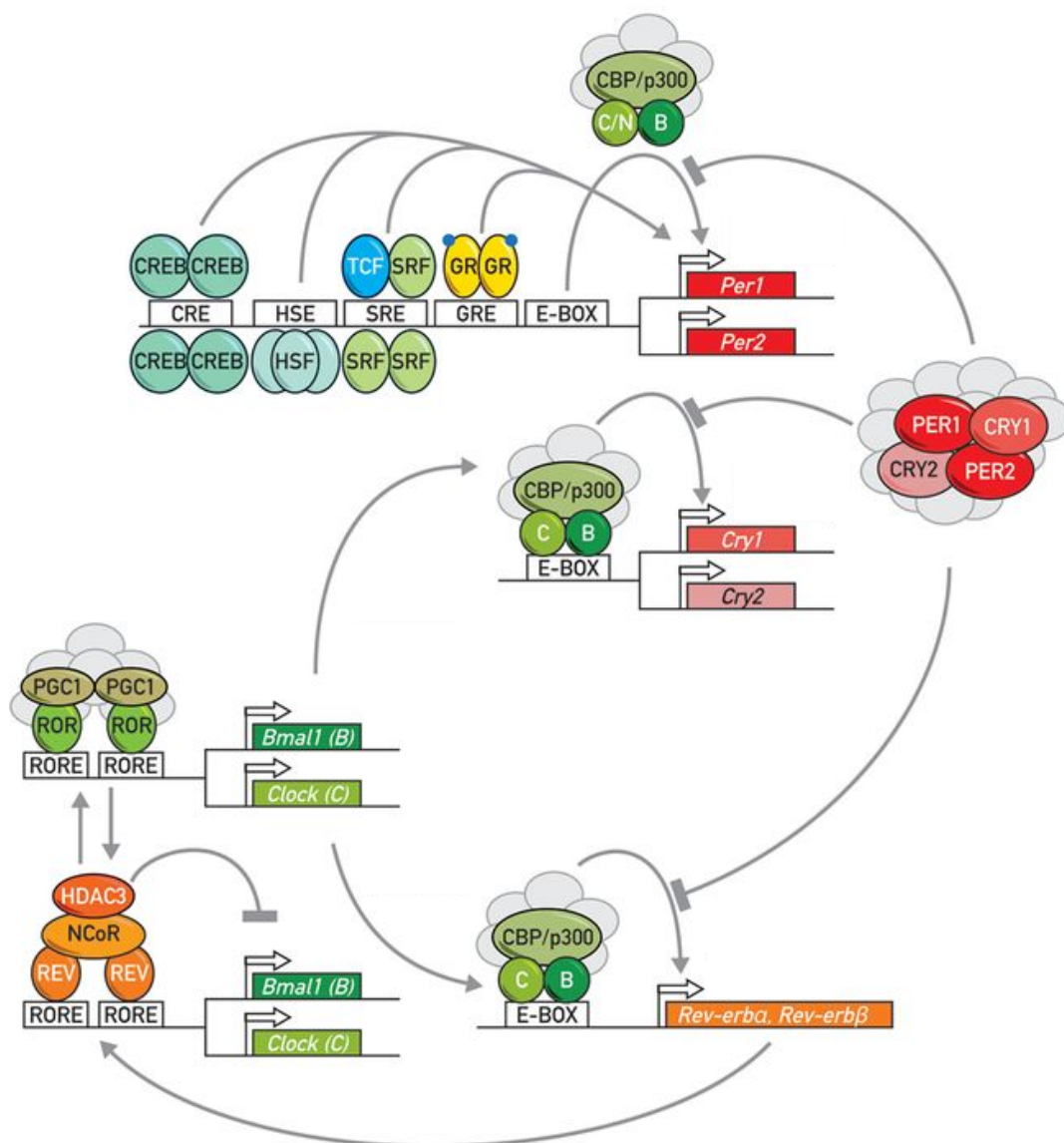


**Figure 1.8 The molecular core clock**
The molecular core clock is composed of two interlocked transcriptional-translational feedback loops. The canonical loop consists in the transcriptional activators BMAL1 and CLOCK (green) that regulate the expression PERs and CRYs (red). PERS and CRYs form a complex that will inhibit the activity of BMAL1 and CLOCK, thus repressing their own activity. PERs are further regulated by multiple systemic factors acting though CRE, HSE, SRE and GRE elements in their promoters. Expression of *Bmal1* and *Clock* is regulated by another loop involving the ROR activators (green) and the REV-ERB inhibitors (orange). Modified from (Schibler et al. 2015).

PERs associate in the cytosol to form a complex that translocates into the nucleus to inhibit CLOCK and BMAL1 activity (van der Horst et al. 1999; Kume et al. 1999; Sato et al. 2006; St. John et al. 2014). Therefore, CRYs and PERs indirectly repress their own expression. This negative feedback loop is tightly regulated at the post-transcriptional translational level, since the nuclear translocation and degradation of the complex is regulated by the activity of enzymes notably of the Casein kinase 1 (CKI) family (Vielhaber et al. 2000; Eide et al. 2005; Reischl et al. 2007). Interestingly, peripheral clocks are mainly synchronized through modulation of *Per1* and *Per2* expression since their promoters contain several elements such as CRE, GRE and SRR that directly respond to systemic cues (Travnickova-Bendova et al. 2002; Gerber et al. 2013; Cheon et al. 2013). Other CLOCK-BMAL1 targets participate in accessory regulatory loops. It is the case of a family of nuclear receptors composed of both activators (RORa, RORb and RORc) and repressors (REV-ERBα and REV-ERBβ), which are directly involved in controlling BMAL1 and CLOCK expression through the RORE motifs in their promoter regions (Guillaumond et al. 2005; Cho et al. 2012). Finally, DBP/TEF/HLF and E4BP4 are activators and repressor, respectively, that additionally participate in the regulation of other clock genes such as PERs and RORs in a D-box-mediated way (Mitsui et al. 2001; Ueda et al. 2005).

In addition to generating and maintaining the 24 hours rhythmicity, the core clock regulates the expression of downstream targets. Thousand of transcripts are dynamically expressed with a period of 24 hours in various tissues. However, only a small fraction of these rhythmically expressed genes overlap between the different tissues (Storch et al. 2002; Zhang et al. 2014). Therefore, the circadian clock is crucial to regulate the tissue-specific temporal expression of genes involved in various cellular processes such as metabolism (Panda et al. 2002; Eckel-Mahan et al. 2012), proliferation (Miller et al. 2007) or signaling (Ueda et al. 2002; Zhang et al. 2014).

### 1.3.2 The transcriptional regulation of circadian genes

As implied above, the rhythmic expression of the core clock and other circadian genes is largely influenced by the motifs present in their promoters. Indeed, promoter region of most circadian genes is composed of regulatory elements of three types that recruit distinct rhythmic transcription factors and define the gene expression phase (Ukai-Tadenuma et al. 2008, 2011) (**Table 1.2**). E-boxes are involved in the expression of morning genes while D-boxes triggers daytime expression and RORE nighttime expression phases (Ueda et al. 2005). According to the phase vector model, the expression phase of circadian genes is determined by the combination of phase-specific regulatory elements of their promoters and the rhythmic activity of their corresponding activators and repressors (Yamamoto et al. 2004; Ukai-Tadenuma et al. 2011; Jolley et al. 2014).

The phase-specific binding of transcription factors induces additional transcriptional regulatory changes. Indeed, the BMAL1-CLOCK complex is thought to have pioneering properties that allow its binding to DNA wrapped around nucleosomes to promote rhythmic nucleosome removal (Menet et al. 2014). Additionally, histone modifications such as acetylation or methylation were shown to oscillate with a 24-hours periodicity at the promoter of circadian genes (Etchegaray et al. 2003; Ripperger and Schibler 2006; Koike et al. 2012; Vollmers et al. 2012; Fang et al. 2014).

| Gene name | Clock function | Promoter motifs | References |
|---|---|---|---|
| *Arntl*<br>*(Bmal1)* | Act (E) | R (2) | (Preitner et al. 2002; Ueda et al. 2005; Akashi and Takumi 2005; Guillaumond et al. 2005; Takeda et al. 2012)2) |
| *Clock* | Act (E) | R | (Ueda et al. 2005; Takeda et al. 2012)2) |
| *Npas2*<br>*(Bhlhe9)* | Act (E) | R (2) | (Ueda et al. 2005; Crumbley et al. 2010; Takeda et al. 2011)1) |
| *Per1* | Rep (E) | E(5), D, G(2), C | (Hida et al. 2000; Mitsui et al. 2001; Travnickova-Bendova et al. 2002; Honma et al. 2002; Ueda et al. 2005; Nakashima et al. 2008; So et al. 2009; Yamajuku et al. 2010; Reddy et al. 2012)2) |
| *Per2* | Rep (E) | E(2), D(2), G(2), C, S | (Travnickova-Bendova et al. 2002; Ueda et al. 2005; Ohno et al. 2006; So et al. 2009; Yamajuku et al. 2010; Gerber et al. 2013; Cheon et al. 2013)3) |
| *Per3* | Rep (E) | E, D(2) | (Jin et al. 1999; Ueda et al. 2005; Nakahata et al. 2008)8) |
| *Cry1* | Rep (E) | E(2), D, R(2) | (Etchegaray et al. 2003; Ueda et al. 2005; Fustin et al. 2009; Ukai-Tadenuma et al. 2011; Takeda et al. 2012)2) |
| *Cry2* | Rep (E) | NA | - |
| *Dbp* | Act (D) | E(3) | (Ripperger et al. 2000; Ueda et al. 2005; Nakashima et al. 2008; Stratmann et al. 2010; Yamajuku et al. 2010; Ueshima et al. 2012)2) |
| *Hlf* | Act (D) | NA | - |
| *Tef* | Act (D) | E(2) | (Nakahata et al. 2008)8) |
| *Nfil3*<br>*(E4bp4)* | Rep (D) | R(2) | (Ueda et al. 2005; Takeda et al. 2012)2) |
| *Rorα* | Act (R) | D(3) | (Ueda et al. 2005)5) |
| *Rorβ* | Act (R) | D(1) | (Ueda et al. 2005)5) |
| *Rorγ* | Act (R) | E(2), R | (Ueda et al. 2005; Takeda et al. 2012)2) |
| *Nr1d1*<br>*(Rev-Erbα)* | Rep (R) | E(5), R(2) | 2)(Adelmant et al. 1996; Raspè et al. 2002; Ueda et al. 2005; Nakashima et al. 2008; Stratmann et al. 2010; Ueshima et al. 2012) |
| *Nr1d2*<br>*(Rev-Erbβ)* | Rep (R) | E(2), D | (Ueda et al. 2005; Yang et al. 2013)3) |

**Table 1.2 Promoter composition of circadian genes**
Non-exhaustive list of the cis-regulatory elements regulating the core clock genes and some of their principal downstream targets. The name of the gene (and putative alternative names) is indicated, as well as the function in the clock: activator (Act) or repressor (Rep) or E-boxes (E), D-boxes (D) or RORE (R). The motifs found in the promoters of these genes can comprise E-boxes (E), D-boxes (D) or RORE (R). The number of each identified motifs is indicated in brackets. In addition to phase-specific elements, Per1 and Per2 are regulated by motifs involved in the clock synchronization through systemic cues: Glucocorticoid receptor (G), cAMP response (C) and Serum response (S) elements. NA stands for not assessed

Indeed, several core clock complexes were shown to contain histone-modifying enzymes. For example, on E-boxes, the BMAL1-CLOCK complex favors transcription by recruiting the p300 acetyl-transferase, the MLL1 methyl-transferase and the JARID1a demethylase (Etchegaray et al. 2003; Katada and Sassone-Corsi 2010; DiTacchio et al. 2011). Interestingly, CLOCK itself was proposed to functions as a histone acetyl-transferase (Doi et al. 2006). In contrast, during transcriptional repression, H3K27 residues in E-box-containing promoters are di- and tri-methylated by the Polycomb repressive complex (Etchegaray et al. 2006). Additionally, BMAL1-CLOCK-mediated histone acetylation is reversed by the recruitment of the SIN3A and NuRD histone deacetylases to the promoter via the PER-CRY repressive complex (Duong et al. 2011; Kim et al. 2014). These temporal variations in histone marks and nucleosome occupancy likely have a

predominant role in the phase-specific loading of PolII onto circadian promoters (Koike et al. 2012; Sobel et al. 2017).

Finally, transcription regulation of circadian genes also likely involves global reorganization of chromosomes within the nucleus. Although the exact impact on transcriptional output remains uncertain, the genomic loci of *Clock* and *Per2* were found to get physically closer to each other with a 24-hour periodicity, the closest distance corresponding to *Clock* peak of expression (Chen et al. 2015a). Similarly, *Dbp* promoter was shown to form long-range rhythmic and clock-dependent contacts with other genomic regions (Aguilar-Arnal et al. 2013). In addition to the dynamic interactions between rhythmically expressed loci, some genes change sub-nuclear compartments over the circadian period. It is the case of the rhythmically expressed *Pard3* gene, which locus translocates to the lamina in a CTCF and PARP1-mediated way during its trough of expression (Zhao et al. 2015).

Thus, although all core clock genes belong to the same system and are involved in the generation and maintain of the 24 hour rhythmicity in cells, their transcriptional regulation differs depending on the circadian regulatory elements composing their promoters and recruitment of specific transcription factors and co-regulators.

### 1.3.3   *Bmal1* promoter and its regulation

Among core clock genes, *Bmal1* is particularly interesting. First, its role in the core clock is very central: in addition to being a key regulator of the positive arm of the canonical feedback loop, it is the only core clock gene without paralogues and consequently the only simple knockout to confer arrhythmicity (Bunger et al. 2000). Additionally, the *Bmal1* is among the simplest core clock promoters since its rhythmicity is driven by two RORE motifs located downstream of the TSS (+36/+47 and +72/+83) (Preitner et al. 2002). These ROREs are rhythmically bound by two groups of transcription factors of the nuclear receptors family (Forman et al. 1994).

The activators of *Bmal1* expressions are the RORs nuclear receptors (Guillaumond et al. 2005; Takeda et al. 2012). While RORα and RORγ are expressed in most tissues, RORβ expression pattern is largely limited to the brain (Hirose et al. 1994; André et al. 1998). RORs bind to DNA as monomers (Giguère et al. 1995; Harding et al. 1997; Sato et al. 2004), and further recruit co-activators that combine several mechanisms to trigger the expression of downstream genes (Atkins et al. 1999). Notably, RORs are suspected to recruit histone acetyl-transferases (HAT) to *Bmal1* promoter since histone acetylation levels on *Bmal1* are rhythmic and temporally coincide with the recruitment of RORs (Liu et al. 2007; Sun et al. 2011; Papazyan et al. 2016). This characteristic may be link to the RORs co-activator PGC-1a, which forms a complex with several HATs (Liu et al. 2007). Additionally, oscillations in *Bmal1* expression require H3K4me3 triggered by the rhythmically expressed histone methyltransferase MLL3 (Valekunja et al. 2013). This process is likely mediated by RORs, since they can direly interact with both MLL3 and MLL4 (Kim et al. 2015). Finally, the RORs-mediated recruitment of members of the SWI/SNF nucleosomes remodeling complex substantially participates in the amplitude of *Bmal1* rhythmic expression (Zhu et al. 2015). Interestingly, RORs natural ligands, cholesterol and other sterol metabolites, modulate their activity (Kallen et al. 2002; Wang et al. 2010a). However, ligand binding reduced the affinity of

RORs for their co-regulators (Wang et al. 2010b). Thus, RORs ligands act as inverse agonists, and their presence diminishes *Bmal1* expression levels (Wang et al. 2010a).

The transcriptional repression of Bmal1 during its low expression phase is ensured by the two paralogues REV-ERBα and REV-ERBβ (Harding and Lazar 1995; Preitner et al. 2002; Guillaumond et al. 2005). Two mechanisms of action permit the transcriptionally repressive action of REV-ERBs (Harding and Lazar 1995). First, trough competition, monomers can bind to individual ROREs and thus restrict their access for ROR activators. However, for longer-term repression of Bmal1, two REV-ERB proteins are required. By binding the two adjacent ROREs on the promoter, they recruit the NCoR co-repressor (Zamir et al. 1997) and histone deacetylase HDAC3 to modify the epigenome surrounding *Bmal1* gene (Yin and Lazar 2005; Zhang et al. 2015). This mechanism establishes a repressive chromatin state oscillating with a 24-hour periodicity to down-regulate *Bmal1* expression (Feng et al. 2011). Functionally, REV-ERBα and REV-ERBβ use heme metabolite as natural ligand (Raghuram et al. 2007; Yin et al. 2007). However, in contrast to RORs ligands, it is required for REV-ERBs activity and reduction of intracellular heme concentration decreases REV-ERB-mediated gene repression.

In addition to the ROREs, other binding sites such as CCAAT-boxes and GC cluster located closely upstream of the TSS may additionally regulate *Bmal1* basal expression (Hirota et al. 2010a; Xiao et al. 2013; Shostak et al. 2016). Also, SAF-A is thought to bind the nucleosome depleted region downstream of *Bmal1* TSS (Onishi et al. 2008). However, their exact role is poorly understood, and they do not seem to affect the expression rhythmicity.

Besides histone modification and transcription factor binding, DNA methylation can be considered as a regulatory mechanism affecting the expression of CpG-rich promoters including *Bmal1* (Deaton and Bird 2011). Around promoters, CpG methylation often corresponds to stable silencing of transcription. If *Bmal1* promoter remains unmethylated in normal conditions (Onishi et al. 2008; Lin et al. 2012), it was found to be hypermethylated in some types of cancer leading to its downregulation and the loss or rhythmicity in its expression (Taniguchi et al. 2009; Satou et al. 2013).

## 1.4    Hypothesis and aims of the study

Transcription is a very complicated process. The recent discovery of transcriptional bursting can be considered as an opportunity to better understand gene regulation. Indeed, recent studies highlighted the gene-specific character of transcriptional bursting (Suter et al. 2011a; Singer et al. 2014; Bahar Halpern et al. 2015b). The important variations in burst size and frequency observed between genes likely contain substantial information on their regulation and transcription mode. Indeed, genes with similar transcriptional bursting behaviors probably share regulatory properties. However, how molecular mechanisms influence the burst size and frequency remains unclear (Nicolas et al. 2017).

Only very recently, transcriptional bursting started to be studied in dynamic systems to improve the understanding of the underlying mechanisms (Molina et al. 2013; Senecal et al. 2014; Bahar Halpern et al. 2015b). For such approach, the mammalian circadian clock is ideal. Indeed,

transcription of circadian genes is endogenously dynamic, with their promoters spontaneously switching between active and repressed phases in a 24 hours period. The length of this period is optimal since it is sufficiently short to monitor several cycles simultaneously, but still largely longer than the bursting time-scale. Also, the switch between the active and inactive circadian phase is smooth, which also allows the study of the transition states. Additionally, the circadian regulatory network is simple and well understood, with promoters composed of three types of regulatory elements and known regulatory mechanisms. Finally, the circadian system can be studied in cultured cells (Nagoshi et al. 2005), which greatly facilitates the monitoring of transcriptional bursting and the panel of experimental possibilities.

Because of its central role in the circadian clock, the simplicity of its promoter and its well-understood regulatory mechanisms, *Bmal1* is particularly interesting. Thus, in this project, we used the previously published short-lived luciferase system (Suter et al. 2011a) to monitor in real-time the variations in transcriptional bursting properties of the *Bmal1* core clock gene over several circadian cycles. The technical approach was highly modular and permitted variations in the experimental conditions such as the integration of the promoter at different genomic sites, or the modulation of *Bmal1* expression using specific drugs or promoter point mutations. Using this approach, we could deduce which molecular changes at the promoter level correlated with specific bursting variations (and thus possibly causally participate in their establishment). In addition to *Bmal1*, the circadian system offered the opportunity to expand the discoveries to other rhythmically expressed promoters, but at different phases and involving different regulatory mechanisms.

In this work, we addressed to following questions:

- How does the transcriptional bursting behavior of *Bmal1* and other rhythmically expressed genes oscillate within the circadian period?

- How does the integration site of a promoter influences its transcriptional bursting characteristics, and does it affect the same bursting properties than the circadian clock?

- What molecular mechanisms occurring at the promoter level temporally coincide with variations of bursting

- Do other circadian genes regulated by various factors display the same periodic variations in their transcriptional bursting behavior than *Bmal1*?

- Are correlations between a molecular mechanism and a bursting property observable genome-wide?

# Chapter 2     Results

## 2.1     Design and validation of the short-lived luciferase expression vector

Although transcriptional bursting parameters can be measured using both static and live approaches (Raj and van Oudenaarden 2009; Larson et al. 2009; Lionnet and Singer 2012), real-time monitoring typically provides more robust estimations since it directly relies on measured transcriptional dynamics. However, the real-time monitoring of transcriptional bursting is a challenging and delicate experimental task. The experimental approach should be sufficiently precise to trustfully detect transcription products, and ideally allow their quantification at the single-molecule resolution. Among the available approaches, destabilized protein reporters are commonly used to infer the transcriptional bursting parameters of selected promoters (Harper et al. 2011; Suter et al. 2011a; Dar et al. 2012; Molina et al. 2013). Although the detected signal arises from the protein rather than the transcript, the reduced protein and mRNA half-lives permit signal detection with limited delay from the transcriptional event. The transcriptional bursting parameters can be inferred with a telegraph model whose parameters are fitted to the single-cell time traces. In this study, we opted for a destabilized luciferase reporter approach. In addition to reliably estimate the transcriptional bursting properties of a given promoter, this approach is used in our laboratory and thus required little optimization at the experimental setup and computational analysis level (Suter et al. 2011a; Molina et al. 2013; Zoller et al. 2015).

The following chapter will describe the design and experimental validation of a new version of the short-lived firefly luciferase reporter vector specifically designed for this work.

### 2.1.1    Design of a short-lived luciferase expression vector

The design of the short-lived luciferase vector specifically developed for the present study was based on an existing short-lived luciferase reporter mainly used to assess bursting signatures of endogenous genes using a gene-trapping strategy (Suter et al. 2011a). As mentioned in chapter 1.4, the work presented in this manuscript first aimed at characterizing the transcriptional bursting properties of the *Bmal1* promoter, notably during the circadian period. However, the reporter had to be extensively moldable to allow bursting properties comparison between modified versions of the promoter and with additional circadian promoters with different expression phases and regulatory mechanisms. Also integration site-driven variations in transcriptional bursting should be assessed. Thus, the experimental design required the following additions to the existing reporter vector:

a) A facilitated cloning approach to integrate any promoter of choice or promoter variants to drive the expression of the short-lived luciferase

b) A system permitting the integration of a single copy of the reporter per cell, either at the same genomic location to compare conditions or at different loci to assess the role of the integration environment on transcriptional bursting.

To account for these needs, and despite a previously existing *Bmal1* short-lived luminescence lentiviral reporter (Suter et al. 2011a), several features of the original vector were replaced. It is notably the case of the former promoter, all lentiviral elements and the Blasticidin resistance gene used as a selection marker that were removed in the new version. Instead, we opted for a Gateway cloning cassette upstream of the luciferase coding sequence to facilitate the integration of any type of promoter into the expression vector. We also implemented the Flp/FRT system to stably integrate a single copy of the vector in a given genomic location (Wirth and Hauser 2004) (**Figure 2.1**).
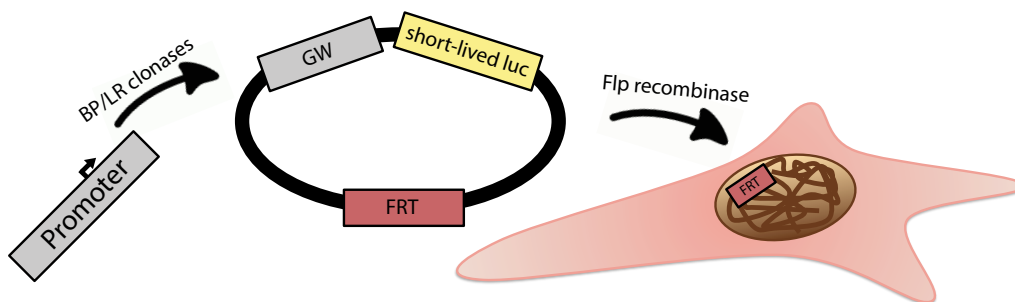


**Figure. 2.1 Generation of cells stably expressing the short-lived luciferase reporter**
The flexibility of the system was made possible by a Gateway cassette, which allowed the integration of any type of promoter to drive the expression of the short-lived luciferase, and the Flp/FRT system that permited the recombination of the plasmid always at the same location into FRT-compatible NIH-3T3 cells.

Concretely, the short-lived luciferase expression vector designed for this study was composed of the following elements (5' to 3') (**Figure 2.2**):

1) A Gateway cloning cassette to facilitates the integration of any promoter of choice (or variants of the same promoter) to drive the expression of the short-lived luciferase without using a restriction enzyme-based cloning strategy.

2) A Kozak consensus sequence and an F2A peptide to handle promoters with various properties. While the TSS is expected to be present on the promoter region integrated using the Gateway system, endogenous start codon is not required. The Kozak sequence was added downstream of the promoter region to compensate for its possible absence in the integrated promoter region (which is notably the case for *Bmal1* promoter). For promoters already containing their own endogenous Kozak sequence, the F2A peptide allowed cleavage by ribosome skipping between the N-terminal region encoded by the

cloned promoter region and the short-lived luciferase (Ryan and Drew 1994). This strategy avoided putative lengthening of the reporter half-life resulting from the protein fusion.

3) The short-lived luciferase coding sequence, whose expression is driven by the cloned promoter. This luciferase is identical to the previous version of the reporter (Suter et al. 2011a). Its N-terminal region consists in the *Luc2* firefly luciferase gene (Promega), while the C-terminal region contains a nuclear localization signal (NLS) to both concentrates the signal in the nucleus and reduce the half-life of the protein (Suter et al. 2011a). The C-terminal region also contains a PEST sequence and an AU-rich elements to destabilize the protein and the transcript. Indeed, the PEST sequence acts as a signal for proteasome degradation (Rogers et al. 1986; García-Alai et al. 2006), while the AU-rich element targets mRNAs for rapid degradation through the binding of multiple cellular factors (Shaw and Kamen 1986; Barreau 2005).

4) An Flp recognition target (FRT) cassette to recombine the entire short-lived luciferase expression plasmid into an FRT cassette in the genome of compatible cells. The cassette comprises an FRT site and the 3' region of a Hygromycin resistance gene. Upon transfection of the FRT cassette-containing short-lived luciferase expression vector in FRT-compatible cells, co-transfected Flippase (Flp) recombinase drives the plasmid recombination into a genomic FRT site (Schlake and Bode 1994; Zhu and Sadowski 1995). The recombination event reconstitutes the full-length Hygromycin resistance gene, allowing for the selection of cells with properly integrated reporter.
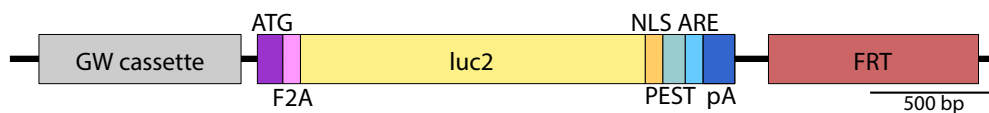


**Figure 2.2 Graphical representation of the short-lived luciferase expression vector**
At scale representation of the different elements composing the short-lived expression vector, with the Gateway cloning cassette (grey), the Kozak consensus sequence (ATG, purple), the F2A peptide (pink), the firefly coding sequence (yellow), the nuclear localization signal (NLS, orange), a PEST sequence (turquoise), the AU-rich elements (ARE, light blue) the polyadenylation site (pA, dark blue) and the FRT cassette (red) composed of an FRT site and the C-terminal region of an Hygromycin resistance gene.

2.1.2   Validation of the short-lived luciferase expression vector in transient transfections

After designing and generating the backbone of the short-lived luciferase expression vector, its functionality was examined in depth. To confirm the versatility of the system, various types of promoters were cloned into the expression vector. The short half-life of the expression products was then tested in parallel for each promoter. In total, four promoters and a negative control were selected and integrated into the short-lived luciferase expression vector using the Gateway cassette:

1) *Bmal1* -970/+157 region (**Figure 2.3**). The selected promoter region of this essential core clock gene comprised ~1kb of sequence upstream of the TSS, and spread until the end of the first non-translated exon. This region is sufficient to reproduce the rhythmic expression

pattern of *Bmal1* (Nagoshi et al. 2004; Stratmann et al. 2012). Also, it contained all transcription factors binding sites known to take part in *Bmal1* expression regulation (two ROREs and two CCAAT-boxes combined with GC cluster, each located around the TSS) (Preitner et al. 2002; Xiao et al. 2013). Additionally, it encompassed the majority of the DHS signal around *Bmal1* promoter in NIH-3T3 cells (Dunham et al. 2012), implying that most regulatory factors involved in the proximal regulation of *Bmal1* bind within the selected fragment. This -970/+157 *Bmal1* promoter region did not contain an endogenous start codon, and it thus required the Kozak sequence contained in the expression vector to properly express the reporter.
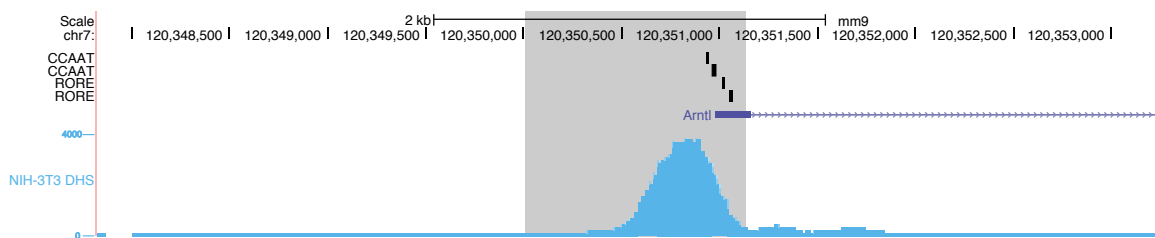


**Figure 2.3 Region of the *Bmal1* promoter selected to drive the expression of the short-lived luciferase**
Genome browser view of the 5' region of the *Arntl* (*Bmal1*) gene. The -970/+157 region selected to drive the expression of the short-lived luciferase is highlighted in grey. It includes the two ROREs and the two CCAAT-boxes combined with GC cluster known to participate in *Bmal1* expression (black boxes), and most of the DHS signal around the TSS region (blue, from ENCODE NIH-3T3 track).

2) *SV40* early promoter (later abbreviated *SV40*). This strong viral promoter was expected to drive elevated though arrhythmic expression levels (Qin et al. 2010). As for *Bmal1*, it did not contain its own start codon and required the Kozak sequence contained on the vector.

3) *Period3* (*Per3*) -536/+1982 region. The expression of this circadian gene is anti-phasic with *Bmal1* (Ueda et al. 2002; Ramanathan et al. 2014). The selected *Per3* promoter region contained the two D-boxes and the E1-E2 box located upstream of the TSS (Ueda et al. 2005; Nakahata et al. 2008), as well as most of the DHS signal located in its intronic regions and likely participating in its regulation (Sobel et al. 2017). To incorporate these putative regulatory elements, the three first coding introns of *Per3* were included in the cloned promoter region. Consequently, the N-terminal region of the PER3 protein separates from the short-lived luciferase protein upon translation of the F2A peptide.

4) H1 CCAAT-box synthetic promoter. This promoter composed of a single CCAAT-box and a minimal TATA-box promoter was designed for a previous study (Suter et al. 2011a). Although its expression level is fairly weak, its bursting signature had already been assessed and could thus be compared with that of the new version of the short-lived luciferase expression vector.

5) A negative control consisting in the +1689/+2196 intronic region of the *Car11* gene (Neg). This region is free of any identified regulatory elements and DHS signal, and does not contain a TSS. It was thus expected to display little or no luciferase expression, and was used to estimate the background expression of the vector

First, the global expression properties of each clone were tested. After integration of the five promoters using the Gateway cassette, the expression vectors were transiently transfected into NIH-3T3 mouse fibroblasts. The expression levels recorded over a period of three days corresponded well to the predicted patterns (**Figure 2.4**). *Bmal1* and *Per3*, the two circadian clock promoters, were the only examples displaying clear circadian rhythmicity in their expression pattern, with *Per3* being anti-phasic to *Bmal1* (**Figure 2.4C**). The expression level of the *SV40* viral promoter was higher than any other promoter tested, and the H1 promoter displayed among the weakest signals. Fortunately, the expression level of the negative control was close to the detection limit, suggesting a quasi-absence of expression leakage from the vector. Thus, independently of the type of integrated promoter and the presence or absence of endogenous start codon, luciferase was efficiently expressed and the expression patterns were consistent with previous reports.
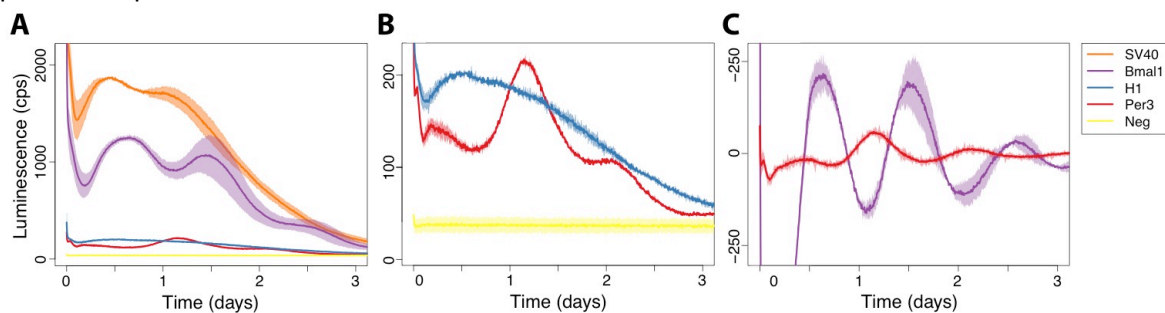


**Figure 2.4 Short-lived luciferase expression driven by various prompters**

Real-time luminescence levels (in counts per second) corresponding to various promoters driving the expression of the short-lived luciferase in populations of transiently transfected cells over three days. The solid line corresponds to the mean and transparent fill to the standard deviation over three replicates. (**A**) Global view of all five constructs. (**B**) Rescaled Y-axis version of (A) to facilitate the visualization of the three weakest constructs (**C**) 24-hours running average baseline subtracted visualization of the two circadian promoters.

Next, both the mRNA and the protein half-lives of the luciferase reporter were tested. This verification was done on all four promoters. Indeed, since the endogenous TSS was included in the cloned promoter region, all clones expressed a specific 5' region of the transcript encoded by the promoter, followed by a common region comprising the short-lived luciferase coding sequence and the 3' end. Although most of the regulatory mechanisms involved in mRNA stability operate in the 3' UTR region common to all transcripts (Mignone et al. 2002; Matoulkova et al. 2012), the 5' promoter-specific region and the transcript length could also participate in defining the transcript half-life (Sharova et al. 2009; Duan et al. 2013; Mauer et al. 2017). Consequently, mRNAs half-lives were likely to differ between the clones. At the protein level however, the presence of the F2A peptide should generate identical proteins products with consistent half-lives in all clones.

Half-lives were measured by recording the luminescence decay after blocking translation and transcription with Cycloheximide and Actinomycin D respectively (**Figure 2.5A**). Cycloheximide treatment provoked a rapid exponential decay of the signal. For all clones, the protein half-life directly measured from the slope of the decay was consistent between the different constructs

and corresponded to 22 minutes (**Table 2.1**) (**Figure 2.5B**). Thus, even upon expression of an N-terminal protein region upstream of the F2A peptide encoded by the cloned promoter region, the half-life of the protein remained consistently short.
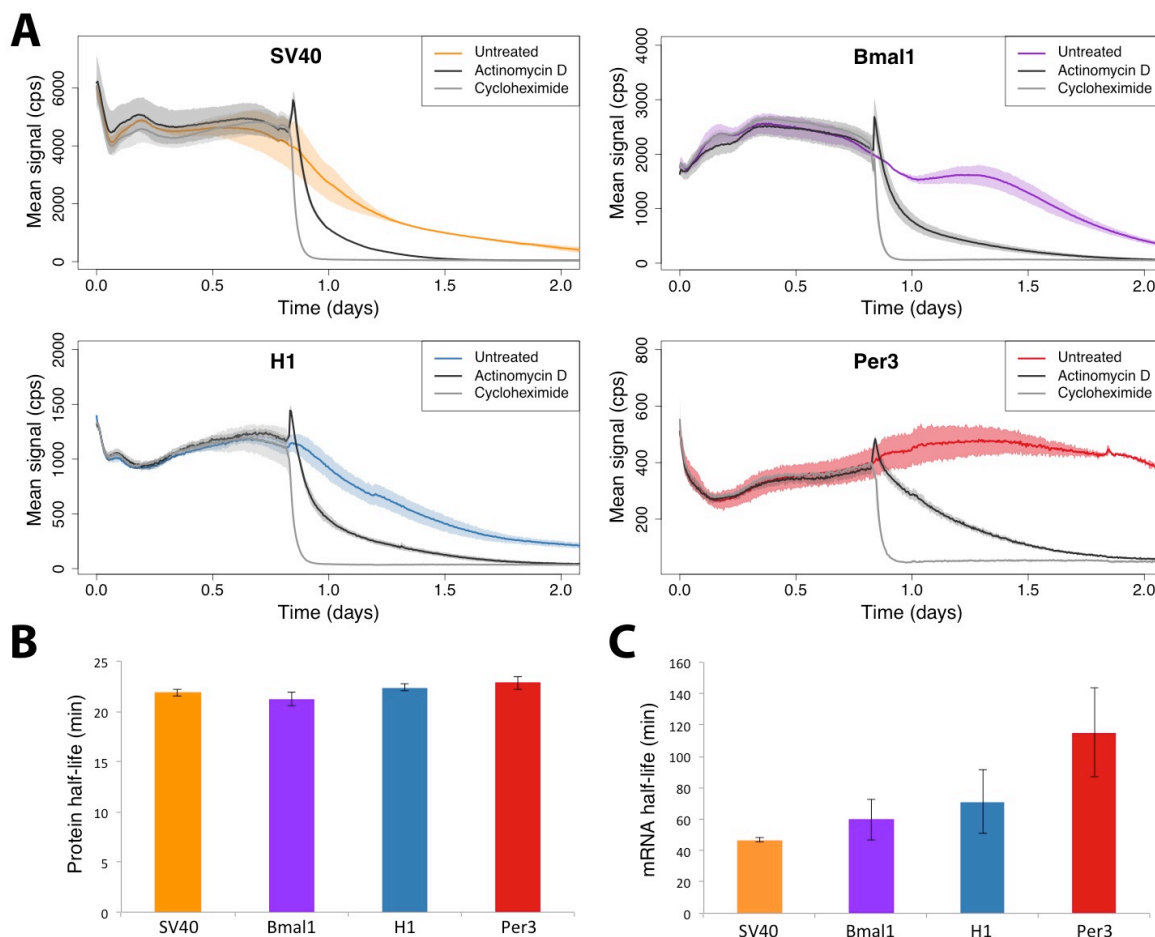


**Figure 2.5 Estimation of luciferase mRNA and Protein half-lives**
Estimation of the short-lived luciferase protein and mRNA half-lives from luminescence decay at the population level for the *SV40* (orange), *Bmal1* (purple), H1 (blue) and *Per3* (red) promoters. (**A**) Real-time measurement in counts per second of luminescence decay after treating the cells with 25 µg/ml of Cycloheximide (grey line) or 5 µg/ml of Actinomycin D (black line). Treatments were applied 22h after starting the recording. The solid line corresponds to the mean and transparent fill to the standard deviation over three replicates. (**B**) Protein half-life estimation from the luminescence decays. (**C**) Transcript half-life estimation from the luminescence decays.

The luciferase expression decay observed after transcription inhibition was slower than for translation inhibition (**Figure 2.5A**). Since the luminescence decay directly reflected the degradation of the protein rather than that of the transcript, half-lives of proteins already present in the cell upon Actinomycin D treatment and of proteins produced after the treatment were considered while inferring the transcript lifespan from the translation inhibition luminescence decay (**Equation 4.5**). Due to the presence of clone-specific 5' transcripts regions encoded on the cloned promoter region, mRNA half-lives displayed more variability than protein half-lives (**Figure 2.5C**). Between the four clones tested, mRNA half-lives typically ranged from 45 minutes to 2

hours (**Table 2.1**). These values are highly consistent with the mRNA and protein half-lives measured with the previous version of the short-lived luciferase reporter (Suter et al. 2011a).

|  | mRNA | Protein |
|---|---|---|
| SV40 | 46.74 ± 1.16 | 21.88 ± 0.36 |
| Bmal1 | 59.75 ± 13 | 21.20 ± 0.65 |
| H1 | 71.23 ± 20.2 | 22.37 ± 0.32 |
| Per3 | 115.32 ± 28.3 | 22.83 ± 0.67 |

**Table 2.1 Estimated short-lived luciferase mRNA and protein half-lives**
Estimation of protein and transcript half-lives from population luminescence decay for four promoters driving the expression of the short-lived luciferase. The mean and standard deviation over three experimental replicates are displayed.

Together, these results demonstrated that the short-lived luciferase expression vector could be used to assess the expression levels driven by various types of promoters, and that the composition of the cloned promoter (notably the presence or absence of an endogenous start codon) did not impact the stability of the protein product. Among all four selected promoters, the short-lived luciferase displayed reduced stability at both the mRNA and protein level. The protein degradation rate was very similar between all clones, and the resulting half-life of 22 minutes was largely shorter than that of the unmodified firefly luciferase protein (2 hours) (Ignowski and Schaffer 2004). The mRNA degradation rate was more variable between the clones, suggesting a role of the 5' region of the transcripts in degradation regulation. For most promoters, the measured transcript half-lives were shorter than the 1.5 hours of the unmodified version of the firefly luciferase (Wilsbacher et al. 2002).

## 2.1.3   Generation of NIH-3T3 FRT cells

Monitoring transcriptional bursting with a short-lived reporter necessitated the presence of a single copy of the expression vector per cell. Indeed, the intrinsic noise inherent to the presence of additional copies would provoke unsynchronized promoter firing that would greatly complicate the gene state inference for each copy. The Flp/FRT system introduced in the expression vector permitted the controlled integration of a single copy of the reporter. Also, using this approach, the construct insertion always occurred at the same genomic location, enabling the comparison of expression properties between different promoter or variants of the same promoter without suffering any undesirable contribution of the genomic environment at the integration site. However, this integration strategy required a specific type of cultured cells compatible with the Flp/FRT system. Indeed, the cells to be used in an FRT system should contain a single FRT site in their genome, where the integration of the FRT-compatible plasmid will occurs upon Flp recombination (**Figure 2.1**). To characterize transcriptional bursting properties of circadian genes, a cell line with a strong endogenous circadian rhythmicity is also essential. NIH-3T3 mouse fibroblasts is a widely used cell line in chronobiology (Nagoshi et al. 2005), and a version compatible with the FRT stable insertion system was commercially available (Invitrogen). To test the rhythmicity of this NIH-3T3-FRT cell line, cells were both transiently transfected with the

*Bmal1* promoter short-lived luciferase (later referred as *Bmal1-sLuc2*) expression vector, or used to stably integrate the same construct into their genomic FRT site.

Unfortunately, while NIH-3T3 cells lacking the FRT genomic integration site (WT NIH-3T3) displayed robust rhythmicity in real-time bioluminescence assay upon transient transfection of the circadian reporter, the global expression level and the amplitude of the circadian oscillations were poorer in the commercial NIH-3T3-FRT cell line (**Figure 2.6A**). This is not due to lower transfection efficiency since co-transfection with 10% of GFP reporter vector did not reveal obvious transfection defect (data not shown). Additionally, stable integration of the short-lived luminescence reporter into the FRT site of the NIH-3T3-FRT Invitrogen cells displayed limited circadian rhythmicity and dampened luminescence signal after a few days of recording (**Figure 2.6A**). This signal dampening was likely caused by cell death in confluent culture environment.
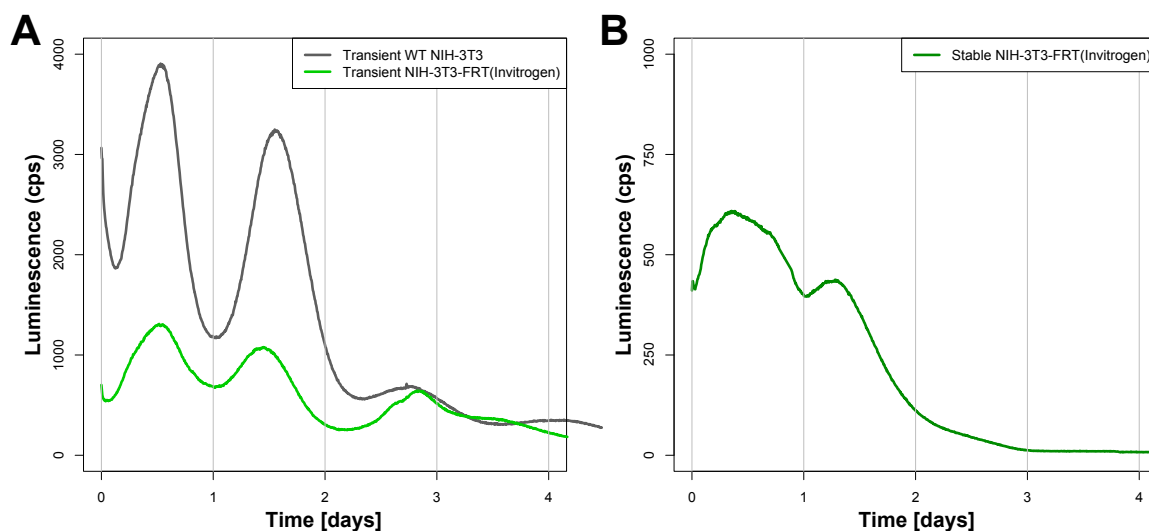


**Figure 2.6 Circadian rhythmicity properties of NIH-3T3-FRT (Invitrogen) cells** The circadian rhythmicity of NIH-3T3-FRT (Invitrogen) cells was tested at the population level in real-time luminescence. Representative traces from individual experiments are displayed in counts per seconds (**A**) Comparison between WT NIH-3T3 (grey) and NIH-3T3-FRT (Invitrogen) (light green), both transiently transfected with *Bmal1-sLuc2* expression vector (**B**) NIH-3T3-FRT (Invitrogen) with a stably integrated *Bmal1-sLuc2* expression vector into the genomic FRT site (dark green).

Because of these poor circadian properties and survival issues in culture condition comparable to the ones used in single-cell luminescence recording (Suter et al. 2011a; Molina et al. 2013), NIH-3T3-FRT cells from Invitrogen could not be used for the monitoring of *Bmal1* transcriptional bursting. Instead, homemade NIH-3T3-FRT cells were generated. This process simply consisted in introducing a pFRT-Neo plasmid (kindly provided by Dr. Jürgen Rippberger, University of Fribourg) containing the genomic FRT cassette into cells, and select for successful integration using the appropriate antibiotic. After selection of individual clones, isogenic populations were grown and tested for the presence of a unique FRT cassette integration. This verification was done by Southern blot on digested genomic DNA (gDNA) harvested from 18 distinct isogenic clones, using a radiolabelled probe targeting a 450bp region of the genomic FRT cassette. 12 clones contained a single band, indicating that a unique FRT cassette successfully integrated in the genomic DNA of a

majority of clones (**Figure 2.7**). Also, the presence of bands at different molecular sizes suggested that most of the 12 positive clones arose from distinct integration events, and consequently carried FRT cassettes at different genomic locations. Only two clones had multiple FRT cassettes, whereas four clones were false positive that likely survived the antibiotic selection even in absence integrated FRT cassettes.
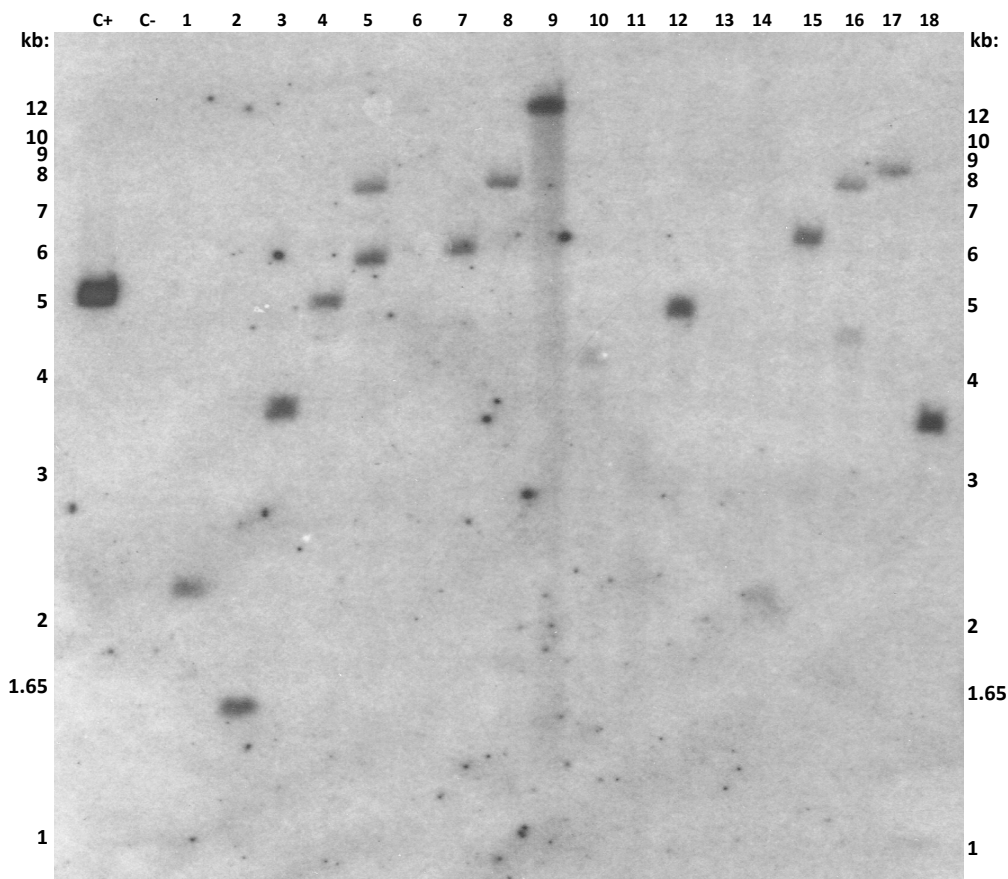


**Figure 2.7 Southern blot verification for unique integration of FRT cassettes in NIH-3T3-FRT clones**
Southern blot on HindIII digested gDNA from 18 individual isogenic clones selected for the stable integration of the FRT cassette, and labeled with a 450 bp FRT cassette probe. 12 clones contained a single copy of the FRT cassette (clones 1-4, 7-9, 12, 14-15 and 17-18), 4 did not integrate the FRT cassette at all (clones 6, 10-11 and 13) and 2 clones contained more than one cassette (clones 5 and 16). The DNA ladder sizes are indicated on the sides in kb. The positive control (C+) consist in HindIII digested pFRT-Neo vector diluted into WT NIH-3T3 gDNA, and negative control (C-) corresponds to HindIII digested WT NIH-3T3 gDNA.

Finally, the positive homemade NIH-3T3-FRT cells were co-transfected with the *Bmal1-sLuc2* expression vector and a Flippase expression vector to stably integrate the reporter into their respective genomic FRT cassettes. This step aimed at verifying the proper and sufficiently high luminescence expression at each integration site and confirming that all selected NIH-3T3-FRT displayed descent circadian rhythmicity. Out of the 12 FRT clones tested, proper integration of the *Bmal1* short-lived luminescence reporter succeeded in 8 cases. Their respective luminescence levels were then assessed at the population level upon circadian synchronization with

Dexamethasone (DEX) (**Figure 2.8A**). Each cell line displayed detectable luminescence levels, with nevertheless important variations in global expression levels between the clones (**Figure 2.8B**). Indeed, up to 10-fold differences in expression levels could be observed between a group of clones displaying the weakest expression levels (clones 1, 7, 8 and 15), and the clone displaying the highest luminescence level (clone 12). Since the same parental cell line was used to generate all FRT clones, these differences in expression levels probably arose from integration site-specific expression variations, with some FRT loci enabling higher expression levels than others. Despite these expression level dissimilarities, rhythmic expression of the *Bmal1* circadian reporter was systematically observed in each cell line. Additionally, similar amplitude fold-changes were measured between the clones, with a 2-3-fold difference in signal between circadian peaks and troughs). Small differences of unknown origin were observed in the expression phase of *Bmal1* between the clones, likely reflecting heterogeneity in the expression of the clock in the population of cells used to generate the FRT clones (**Figure 2.8B**).
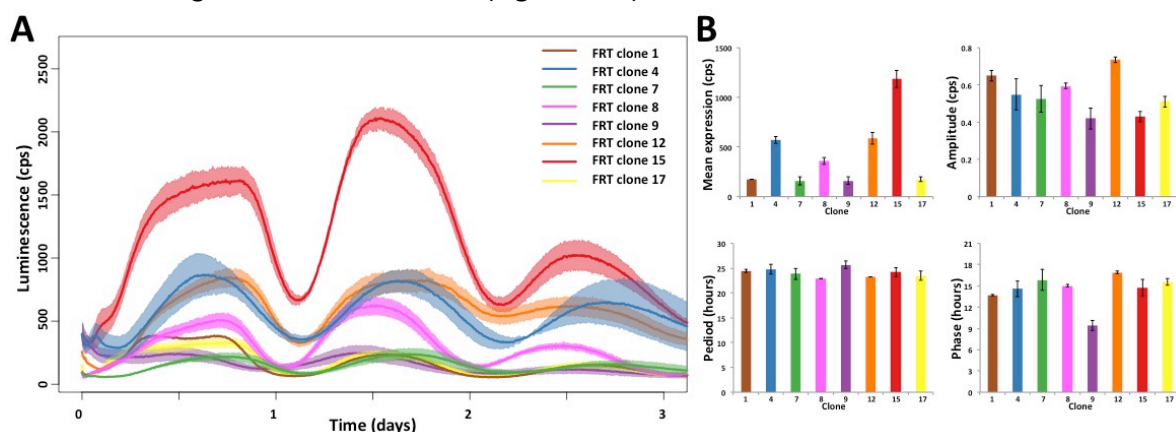


**Figure 2.8 Population luminescence traces of the NIH-3T3-FRT clones after stable integration of the *Bmal1-sLuc2* expression vector**

(**A**) Real-time luminescence expressed from *Bmal1-sLuc2* expression vector stably integrated into the FRT site of 8 homemade NIH-3T3-FRT clones. Cells were recorded at the population level after DEX synchronization. Luminescence levels in counts per seconds are displayed as mean signal (solid line) with standard deviation (transparent fill) over three replicates. (**B**) Expression parameters of each clone: mean expression level (in luminescence counts per second, upper left panel), amplitude (in luminescence counts per second, upper right panel), period (in hours, lower left panel) and peak phase (in hours after DEX, lower right panel) were inferred from the population luminescence data using the ChronoStar software (Spörl et al. 2011).

### 2.1.4   Summary

The design of an appropriate vector for driving the expression of the short-lived luciferase was an essential step for monitoring transcriptional bursting at the single cell-level using the previously published strategy of the destabilized luciferase (Suter et al. 2011a; Molina et al. 2013). We opted for a versatile system that could be used with any promoter of interest and facilitated their stable integration into cellular DNA at a single, always identical, genomic location. Testing several promoters with various expression patterns demonstrated that the short-lived luciferase expression could easily be adapted to any type of promoters. Additionally, the stability of the luciferase transcripts and proteins was greatly reduced in our version of luciferase with a protein

half-life of 22 minutes and an mRNA half-live of ~1 hour for the *Bmal1* reporter. Degradation was thus 5 times more efficient for the protein and 1.5 times for the mRNA compared to the original firefly enzyme (Wilsbacher et al. 2002; Ignowski and Schaffer 2004). The system could not be used as originally planned with commercially available NIH-3T3-FRT (Invitrogen) cells. Instead, homemade NIH-3T3 cells compatible with the FRT recombination system were generated. In addition to their capacity to survive culture condition at high cellular densities for several days, the homemade clones displayed satisfying circadian rhythmicity upon stable integration of the *Bmal1-sLuc2* reporter. Also, the different integration sites displayed various expression levels that could be further used to evaluate the impact of the integration site on transcriptional bursting. Consequently, the 8 NIH-3T3-FRT clones expressing *Bmal1-sLuc2* fulfilled all the criteria to be used for transcriptional bursting analysis, and could be further monitored at the single-cell level.

## 2.2    Bursting monitoring using real-time single-cell luminescence recording

Cell lines stably expressing the short-lived luciferase from a *Bmal1* promoter were generated to be eventually individually monitored in a luminescence microscope. The resulting single-cell luminescence traces would then serve as raw data for the inference of *Bmal1* transcriptional bursting parameters. In this section we describe the single-cell luminescence recording process, the experimental acquisition of additional parameters for the calibration of the telegraph model, and the resulting estimated transcriptional bursting parameters corresponding to the *Bmal1* promoter.

### 2.2.1   Recording luminescence at the single-cell level

To record luminescence at the single-cell level, cells stably expressing short-lived luciferase were diluted into non-luminescent cells and seeded at high confluence. This strategy permitted optimal signal delimitation of individual cell expressing luciferase by avoiding spatial overlap, and also prevented cellular division and migration during the recording. Indeed, the short half-life of the reporter causes flickering signal and moving cells in silent "dark" expression periods can easily be lost during the tracking (Blanchoud et al. 2015). Also, confluent culture environment causes contact inhibition and thus maintains most cells in the G0 cell cycle state. This resting phase reduces the prevalence of extrinsic noise largely caused by variation of cell cycle states within cellular populations (Rosenfeld et al. 2005; Gut et al. 2015; Battich et al. 2015), and likely improves the circadian rhythmicity of the cells. Indeed, in NIH-3T3 cells, the circadian clock and the cell cycle are tightly linked, and dividing populations are likely to unsynchronize faster than cells in cell cycle arrest (Nagoshi et al. 2004; Bieler et al. 2014). Consequently, cultures at high cellular densities typically display robust circadian oscillations (O'Neill and Hastings 2008; Noguchi et al. 2013).

Single-cell luminescent traces were recorded for seven of the eight FTR clones stably expressing *Bmal1-sLuc2*. To ensure consistent circadian phase within the population and avoid circadian phase determination from noisy single-cell recordings, cells were synchronized with dexamethasone (DEX) prior to their recording in a luminescence microscope. Luminescence was

recorded for 48 hours, and single-cell luminescence traces were extracted from the resulting movies using the CAST software (Blanchoud et al. 2015) (**Figure 2.9**). In addition to being optimized for luminescence data, this platform excelled at segmenting cells at low signal-to-noise ratios (**Figure 2.9B**) and could track the cells even in absence of signal for prolonged periods (**Figure 2.9C**).
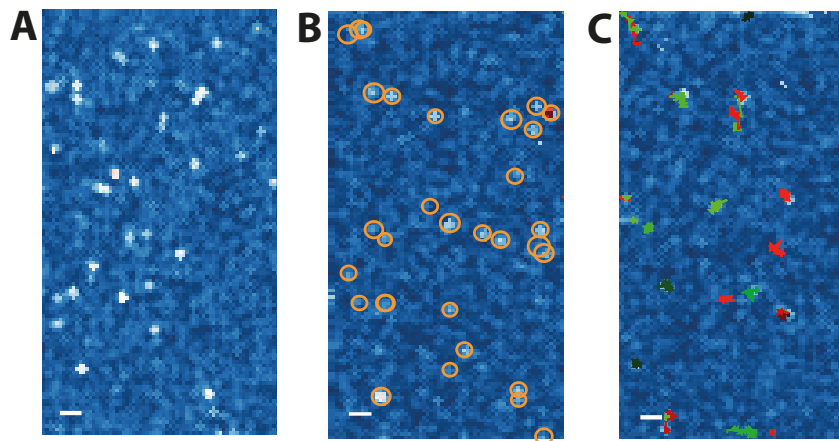


**Figure 2.9 Luminescence recording analysis using CAST software**
To obtain single-cell luminescence traces, signal from 2-days recording in a luminescence microspore were analyzed using the CAST software (Blanchoud et al. 2015). From the raw recordings (**A**), cells were segmented independently in each frame (orange circle) (**B**) and the full 24-hours tracks corresponding to individual cells throughout the recording were identified by computing the cell trajectories and interpolating the signal for cells transiently disappearing objects (random red-green color code) (**C**). The presence of a marked track in absence of luminescence signal indicates the presence of a transcriptionally silent cell in the presented frame. Note that panels A-C do not illustrate the same frame. The white scale represents 10 microns. Figure modified from (Blanchoud et al. 2015).

The resulting single-cell traces displayed substantial cell-to-cell variability within the same clones despite their identical genetic background as well as noticeable expression level differences between the various clones (**Figure 2.10**). The fluctuations in the signal reflected well the stochasticity inherent to *Bmal1* expression. However, in most traces, the circadian rhythmicity over the two days of recording was still distinguishable.

Estimations of the population expression behaviors were obtained by averaging all single-cell traces for a given FRT clone (**Figure 2.11**). This representation highlighted the different expression levels observed between the different FRT integration sites, and recapitulated well the expression features of population luminescence monitoring (**Figure 2.8**). Despite modest nevertheless noticeable variability in expression phase and period, all clones displayed clear circadian rhythmicity. From this population luminescence estimation, the seven clones harboring different FRT integration sites could be separated into 3 groups based on their luminescence expression levels. FRT clone 15 was the only clone displaying high expression levels. Two other clones formed a group with intermediate expression levels (FRT clones 4 and 12), with luminescence signal globally 40% lower than that of the highly expressing clone. Finally, the remaining four clones all expressed similar weak levels of luciferase, about 60% lower than the clone with the highest expression level.
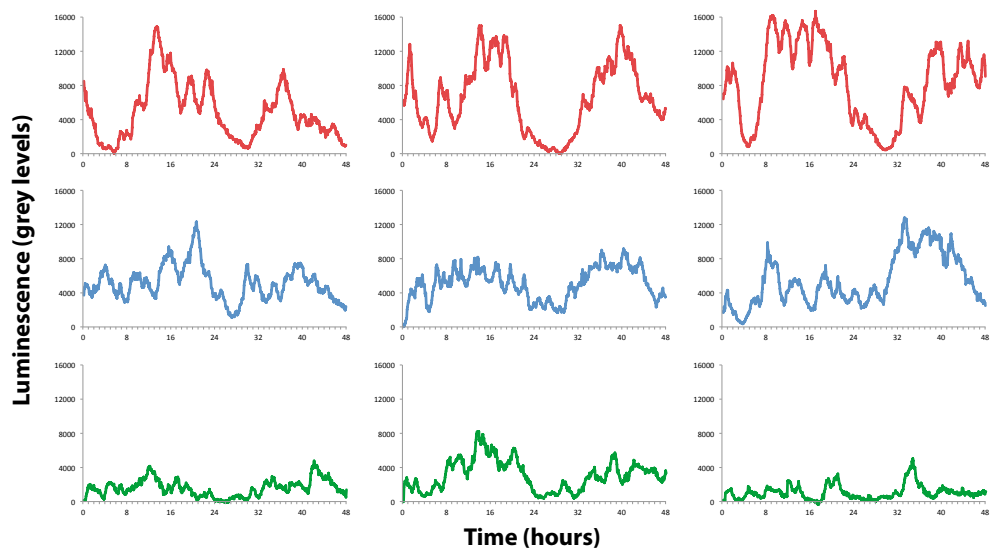
**Figure 2.10 Single-cell luminescence traces**

Nine representative examples of single-cell traces corresponding to the expression of *Bmal1*-driven luciferase over the two days of recording and for three different clones. Red traces correspond to a clone with high expression level (H, clone 15), blue traces to a clone with intermediate expression level (M, clone 4) and the green traces to a clone with low expression level (L, clone 7).
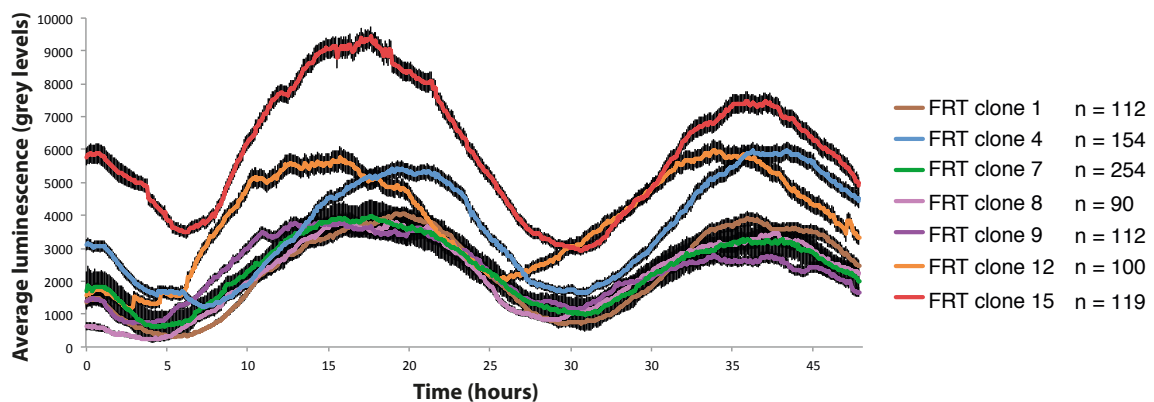


**Figure 2.11 Average luminescent trace from single-cell measurements**

Averaged luminescence trace from individually monitored single-cells of 7 different FRT clones. The colored line represent the average luminescence signal, and the black periphery the standard error (the signal variability within clones being too important to graphically represent the standard deviation). The number of individual luminescence traces recorded for each clone is indicated in the legend.

Both single-cell traces (**Figure 2.10**) and population average traces (**Figure 2.11**) indicated that luminescence microscopy successfully captured essential expression features of *Bmal1-sLuc2* clones. Notably, these recording highlighted the stochastic expression pattern of the highly cell-specific single-cell traces, the circadian rhythmicity observable at the population level as well as in a majority of individual traces, and the difference of global expression levels observed between

the FRT clones. Due to the satisfactory quality of these single-cell traces, they were further used to infer the transcriptional busting properties of the *Bmal1* promoter.

## 2.2.2   Monitoring transcriptional bursting using the short-lived luciferase approach

Single-cell traces collected from the 7 FRT clones monitored in luminescence microscopy were used to compute the transcriptional bursting properties of the *Bmal1* promoter. The likelihood of the promoter to be in the "on" or "off" state was calculated in the successive time points of individual luminescence traces (**Equation 4.5**). The fitted telegraph parameters were then determined for each condition.

To parameterize the model and thus infer the bursting kinetics, six rates were required: the $k_{on}$ and $k_{off}$ switching rates between the transcriptionally inactive to active state and active to inactive states respectively, the transcription rate $k_m$, the translation rate $k_p$, and finally the mRNA and protein degradation rates $Y_m$ and $Y_p$ respectively (**Figure 2.12**). Although potentially subject to fluctuations over the circadian period, these rates were kept constant to simplicity reasons.
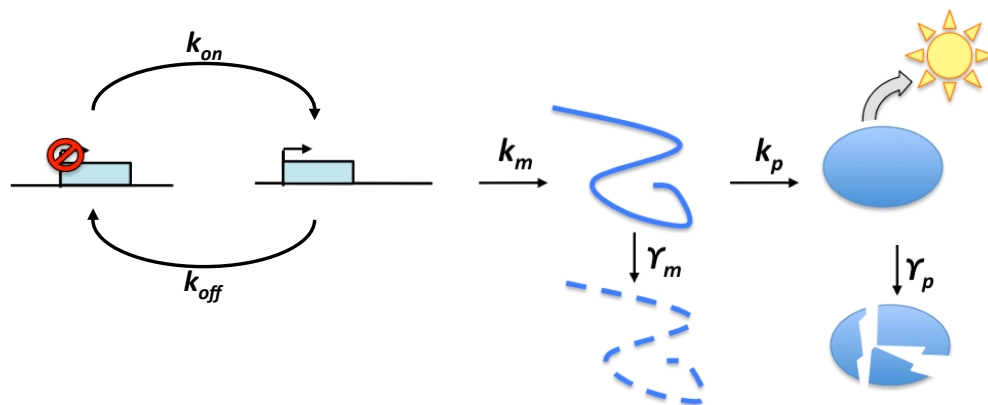


**Figure 2.12 The short-lived luciferase system and telegraph model**
Schematic representation of the two-state model of gene expression (telegraph model) used to infer transcriptional bursting parameters from short-lived luciferase activity. The promoter can stochastically switch between a transcriptionally active ("on") or inactive ("off") state with the $k_{on}$ and $k_{off}$ transition rates. The "on" state allows the production of mRNA at a transcription rate $k_m$, and the mRNA is translated into luciferase proteins at a translation rate of $k_p$. The number of proteins in the system directly defines the luminescence intensity. Both mRNA and protein degradation are modeled as Poissonian processes, with rates of $Y_m$ and $Y_p$ respectively.

While the switching rates between the two promoter states as well as the transcription rate needed to be computationally inferred, some parameters of the model could be measured experimentally. It is the case for mRNA and protein half-lives as well as the translation rate. Indeed, the translation rate could be directly calculated from the experimentally determined mean number of luciferase proteins and transcripts per cell, considering both of their degradation rates. Also, the correlation between a given luminescence signal and the corresponding number of luciferase proteins had to be determined prior to analyzing the luminescence time-traces using the telegraph model. These experimentally measurable parameters were determined in the following section.

### 2.2.3  Experimental quantification of the model parameters

Among the four experimentally measurable constants ($k_p$, $\Upsilon_m$ and $\Upsilon_p$), the mRNA and protein degradation rates were already calculated in transiently transfected cells in section 2.1.2. Since the obtained values (22 minutes for protein half-live and 60 minutes for the transcript half-life) were highly similar to the previously calculated ones for an alternative version of *Bmal1* promoter (22 minutes and 69 minutes for protein and mRNA half-lives respectively) (Suter et al. 2011a), the estimated protein and mRNA half-lives were directly integrate into the model.

The translation rate $k_p$ was then quantified. Calculation of this parameter required the comparison between the mean absolute number of luciferase proteins and luciferase transcript per cell. The mean number of luciferase protein per cell could directly be estimated from the luminescence signal knowing the linear relationship between photon emission and the amount of luciferase proteins. Typically, this calibration process involves the comparative quantification of the luminescence signal between known amounts of a recombinant luciferase resuspended in luciferase-negative cells (and displaying comparable luminescence levels as the short-lived version of the enzyme), and a fixed amount of cells stably expressing the short-lived luciferase (Suter et al. 2011a).

Alternatively, microscope calibration can be achieved by comparing luminescence intensities obtained in our luminescence microscope with those of a previously calibrated system using similar reporters (Suter et al. 2011a). More specifically, five cell lines expressing the short-lived luciferase reporter from various promoters and for which the luminescence intensities/number of luciferase protein ratios was known were monitored in our uncalibrated microscope. The mean luminescence values obtained after 6 hours of recording for each clones were then compared to the analogous values previously obtained in a calibrated microscope (**Figure 2.13**). The comparison revealed an excellent fit between the two studies, with the microscope used in this study being ~1.5 times more sensitive than the previously calibrated one (Suter et al. 2011a). From this comparison, individual luciferase enzymes were found to produce an average signal of ~11 grey levels in our luminescence microscope and imaging conditions.

Finally, the mean number of transcripts per cell was quantified by qPCR. Total RNA was harvested from a fixed amount of cells in three clones expressing the short-lived luciferase from the *Bmal1* promoter at various expression levels. In parallel, short-lived luciferase mRNA was generated *in vitro*, and spiked at known concentrations into non-luminescent cells. After reverse transcription, the absolute amount of luciferase transcripts per cell was deduced from the standard of *in vitro* transcribed RNA by qPCR (**Figure 2.14**). The obtained values corresponding to the mean number of mRNA molecules in unsynchronized cells were compared to the average amount of luciferase proteins per cells deduced from microscope calibration for a period of 24 hours. The average translation rate $k_p$ was estimated separately for all three clones, but the same average value over the three clones, 1.70 ± 0.02 transcripts per minute, was used to parameterize the two-states telegraph model.
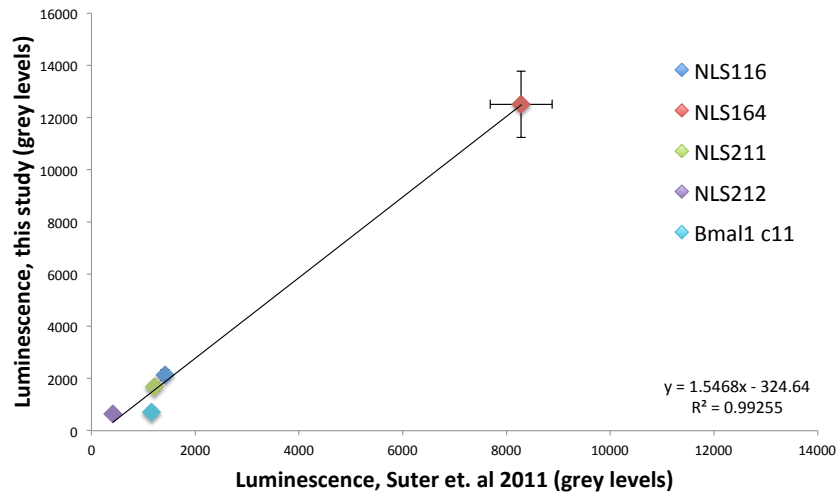
**Figure 2.13 Luminescence microscope calibration**

The luminescence microscope was calibrated to the one previously used in (Suter et al. 2011a) by comparing the mean luminescence intensities per cell (in grey levels) for 5 distinct clones displaying various luminescence expression levels after six hours of recording. The trend line was then used to estimate the mean luminescence signal per luciferase protein, knowing the corresponding value in (Suter et al. 2011a). The trend line equation and its $R^2$ are indicated in the lower right corner. The values correspond to the mean ± SE for n > 50 cells.

Thus, by estimating the mean number of molecules per cell and the correspondence between luminescence intensities and luciferase proteins, we could measure the short-lived luciferase translation rate. $k_p$, together with the luciferase mRNA and protein degradation rates arising from the half-lives measured in section 2.1.2, were the three parameters of the two-states model of gene expression that could be measured experimentally. They were further used in the telegraph model to infer the transcriptional bursting parameters of *Bmal1*.
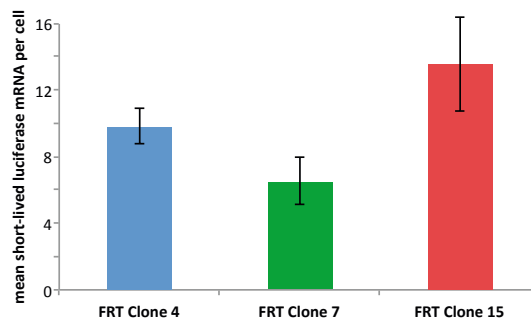


**Figure 2.14 Mean copies of short-lived luciferase transcripts per cell**

Estimated mean copy number of *Bmal1-sLuc2* transcripts in 3 unsynchronized FRT clones displaying various luciferase expression levels. The values correspond to qPCR quantifications of luciferase transcript normalized using a standard of fixed *in vitro* transcribed short-lived luciferase mRNA, and are shown as mean ± SD for 3 replicates.

## 2.2.4 Inferring the transcriptional bursting pattern of *Bmal1* from luminescence traces

Using the experimentally measured translation and degradation rates, single-cell traces were analyzed to infer transcriptional bursting properties of the *Bmal1* promoter. In particular, the analysis aimed at revealing which aspects of transcriptional bursting differed between the FRT

clones and between the circadian phases, since these conditions displayed expression levels variations in averaged single-cell luminescence traces (Figure 2.11). Notably, we aimed at determining whether these two independent ways of modulating *Bmal1* expression affected the burst size, the burst frequency or both simultaneously. All computational analysis based on the telegraph model were carried on by Dr. Benjamin Zoller (Naef lab), using similar approaches as descried earlier (Zoller et al. 2015).

As a first approach, the entire time-traces representing 48 hours of recording were analyzed as a single input without taking into account circadian variations in transcription (**Figure 2.15**). This approach provided information regarding the average transcriptional bursting behavior of each FRT clone, but ignored the circadian phase-specific variations in transcriptional bursting parameters. This strategy was used to determine which conditions should be studied in depth with more reliable approaches accounting for the differentiation between the circadian phases. Global analyzes of the 48-hours of recording highlighted the low variability between transcriptional bursting parameters of clones belonging to the low expression level group (FRT clones 1, 7, 8 and 9) (**Figure 2.15**). Also, the major transcriptional bursting disparities between the high expression level clone (FRT clone 15) and the clones displaying lower expression levels arose from burst size variations. Indeed, the burst frequency is described by the invert of the combined "on" and "off" time ($f = (\tau_a+T)^{-1}$), while the burst size corresponds to the product of the transcription rate and the time spent in the "on" state ($b = k_m\tau_a$). Yet, while neither $\tau_a$ nor $T$ significantly varied between the expression levels, $k_m$ is 2-times higher in the high expression clone than in the weaker ones. It will consequently influence the burst size between the two groups of clones (**Figure 2.15**).
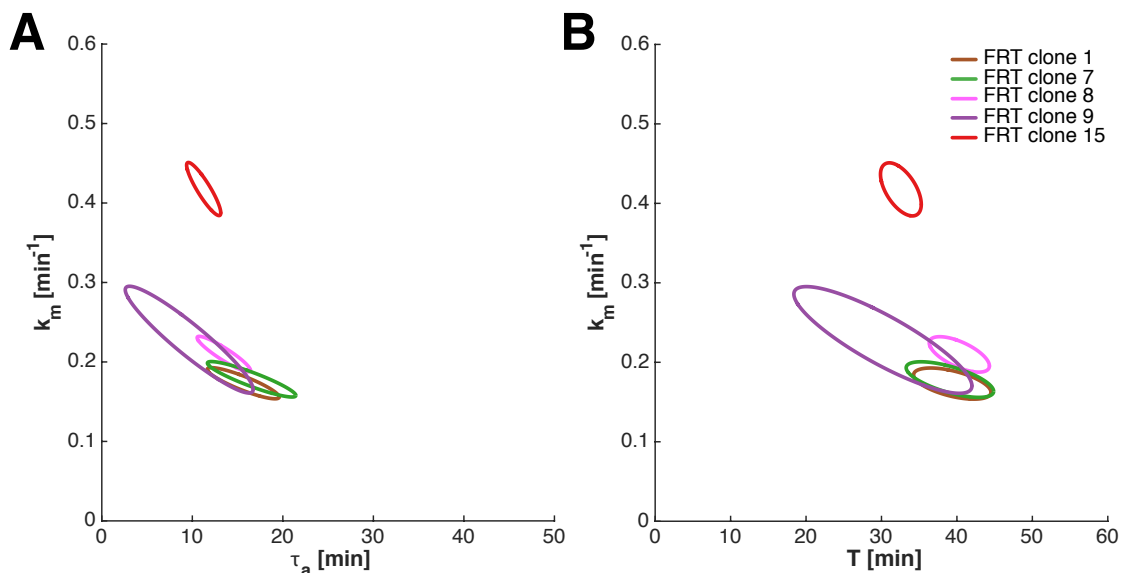


**Figure 2.15 Transcriptional bursting parameters of the *Bmal1* promoter inferred from entire 48-hour traces**
The bursting parameters of the *Bmal1* promoters inferred from the full-length 48-hours traces recorded from a clone with high expression level (FRT clone 15, red) and 4 clones with low expression levels (FRT clones 1, 7, 8 and 9, brown, green, pink and purple respectively). The clones displaying intermediate expression levels (FRT clones 4 and 12) are not presented in this figure. (**A**) Transcription rate $k_m$ against average minutes spent in "on" state $\tau_a$ (**B**) Transcription rate $k_m$ against average minutes spent in "off" state $T$. Traces corresponding to a minimum of 90 individual cells were used for each clone.

Based on this pilot analysis of full-length 48 hours traces, we considered that the luminescence level similarities observed between clones of the same expression group (high, intermediate or low expression levels) embodied comparable transcriptional bursting kinetics. Thus, following analysis were performed on a reduced number of three FRT clones representative of the three expression levels. FRT clone 15, as only representative of the high expression level group was renamed clone H (for high). FRT clone 4 was selected to illustrate intermediate expression levels condition and was renamed M (for medium), while clone 7 was chosen to represent the low expression level group and renamed L (for Low).

To identify aspects of transcriptional bursting fluctuating over the circadian period, the three selected clones were re-analyzed by separating the 48 hours traces in shorter circadian phases. More specifically, a sliding-window approach was applied to independently infer the bursting kinetics from fractions of traces of eight hours, sliding every four hours (**Figure 2.16**).
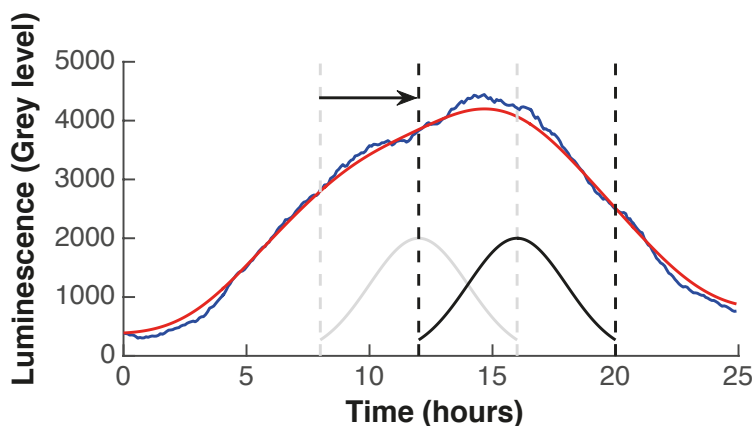


**Figure 2.16 Sliding window strategy for time-resolved analysis of *Bmal1* bursting features**
Schematic representation of the sliding window strategy used to temporally dissect the transcriptional bursting signature of *Bmal1* in the H, M and L clones. The window contains an 8-hours fraction of the luminescence traces, and is sliding every 4 hours until complete analysis of the 48 hours (sliding represented as arrow between grey an black windows). A Gaussian weight with σ=2h was applied to reinforce the parameters inferred from the central fraction of the window. The sliding window is exemplified on the first 25 hours of the average single-cell trace corresponding to the M clone (blue). Circadian rhythmicity is highlighted by a cosine curve fit (red).

Time-resolved analysis of the three selected FRT clones revealed inconsistencies between the inferred bursting kinetics arising from the two circadian cycles during the 2 days of recording. Indeed, while the bursting kinetics were expected to repeat as a 24-hours loop if the two circadian cycles were sharing similar luciferase expression properties, we observed a sharp decrease in burst size between the first and second circadian cycles (**Figure 2.17**). Consequently, although the parameters best describing the bursting properties of the *Bmal1* did cycle over a 24-hours period, the two circadian rounds were markedly shifted with the second cycle displaying a largely lower burst size than the first one. This phenomenon was considered to arise from cellular changes caused by extreme recording conditions. Indeed, while the two circadian cycles are unlikely to behave differently in optimal culture conditions, recorded cells are maintained in a highly

confluent environment for prolonged periods, which could eventually alter their behavior. Additionally, the maintain of uniform culture condition at the temperature, humidity and $CO_2$ level is more challenging in the luminescence than in a conventional culture incubator. Thus, the refined analysis of *Bmal1* promoter bursting kinetics over the circadian cycle and in the three selected differentially expressing integration sites was only examined for the first circadian period of the single-cell traces corresponding to CT6 to 30 (in hours after DEX synchronization).
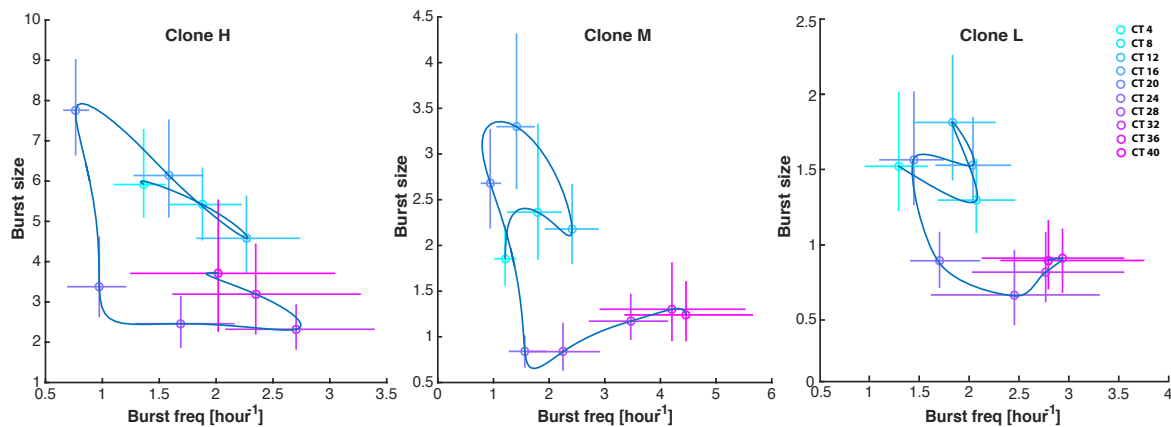


**Figure 2.17 Time-resolved analysis of *Bmal1* promoter bursting properties over two circadian cycles**
Transcriptional busting parameters (burst size and burst frequency) inferred over 48-hours of recording using the sliding window time-trace partitioning strategy. The different time-windows analyzed are indicated in a blue-to-pink gradient in circadian time (CT, corresponding to hours after DEX synchronization). Each point is represented by the average value (circle) and the 5th and 95th percentiles of the estimate (bars) over all MCMC iterations. The time-points are linked with a polynomial trend line (blue). Parameters corresponding to clones H (left panel), clone M (middle panel) and clone L (right panel) are displayed.

The time-resolved analysis of the first circadian period of time-traces corresponding to the H, M and L clones nicely recapitulated the rhythmic accumulation of *Bmal1* transcripts (**Figure 2.18A**). While expression noise (expressed in $CV^2$) also displays circadian rhythmicity, it oscillated in complete anti-phase to the mRNA accumulation (**Figure 2.18B**). This was consistent with previous studies stating that gene expression noise and more specifically intrinsic noise, are anti-correlated with expression level (Elowitz et al. 2002; Becskei et al. 2005; Taniguchi et al. 2010; Stewart-Ornstein et al. 2012). When *Bmal1* expression levels were partitioned into bursting parameters, the two-state model clearly revealed rhythmicity in burst frequency at a similar phase than mRNA accumulation (**Figure 2.18C**). However, the burst frequencies were comparable between the clones, indicating that the integration site did not impact this bursting property. On the other hand, burst size was found to largely vary between the clones coherently with their global expression levels (**Figure 2.18D**). Thus, the different genomic location of the three selected clones likely affected the number of transcripts produced per burst rather than their frequencies. However, no obvious circadian rhythmicity was detected in burst size. Indeed, while the L clone displays relatively flat burst size over the 24-hours of recording, the temporal pattern of clones H and M, although less regular, did not correspond to circadian oscillations. Thus, from these data,

transcriptional bursting variations over the circadian period were found to mainly arise from changes in burst frequency. These phase-specific variations in burst frequency mainly arose from longer promoter "off" states during the circadian trough of expression (**Figure 2.18F**). While the H and M clones also display increased "on" times around CT22 (**Figure 2.18E**), they were compensated by a slight decrease in translation rate at the same phase (**Figure 2.18G**).
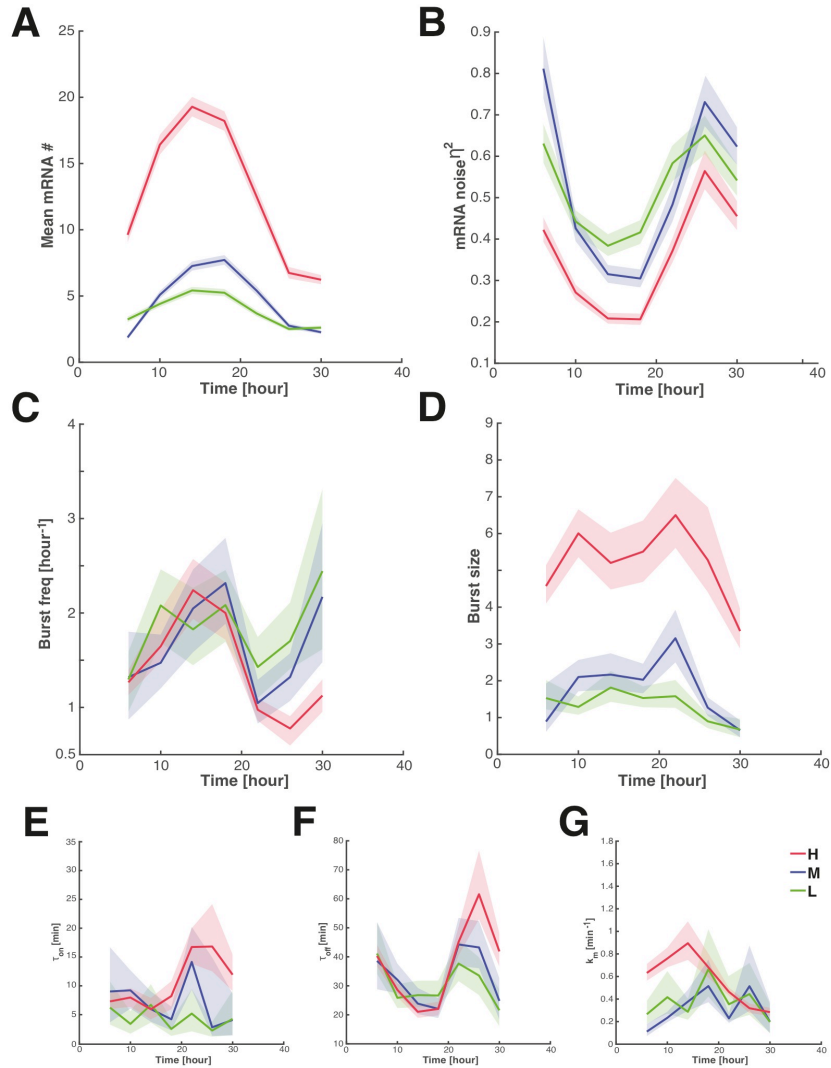


**Figure 2.18 Time-resolved bursting parameters of the *Bmal1* promoter over 24-hours in three clones**
The two-states model of gene expression was used to extract bursting parameters from 24-hours single-cell luminescence *Bmal1* time-traces corresponding to clones with high (H, red), intermediate (M, blue) or low (L, green) expression levels. The four hours sliding window strategy was used to infer the temporal variations of the different parameters over the circadian period. The aspects of bursting shown over the different time-points are (**A**) the mean number of transcripts per cell, (**B**) the expression noise $\eta^2$ (in squared coefficient of variation, corresponding to variance over squared mean number of transcripts), (**C**) the burst frequency, (**D**) the burst size, (**E**) the mean time spent in the "on" state ($\tau_{on}$), (F) the mean time spent in the "off" state ($\tau_{off}$) and (**G**) the transcription rate during the "on" state ($k_m$). In each panel, the solid line corresponds to the average and the transparent fill to the 5th (lower boundary) and 95th (upper boundary) percentiles of the estimate over all MCMC iterations. Time in X-axis is displayed in hours after DEX synchronization. Bursting parameters were inferred from 119 (H clone), 154 (M clone) or 254 (L clone) single-cell luminescence traces.

Since in some genes a refractory period prevents promoter reactivation shortly after a preceding burst (Harper et al. 2011; Suter et al. 2011a; Zoller et al. 2015), the luminescence traces were analyzed with an extension of the two state telegraph model of gene expression that included a promoter cycle composed of *N* inactive states representing the rate limiting events necessary for promoter reactivation (Zoller et al. 2015). For all three clones, the most likely promoter version only contained a single "off" state (**Figure 2.19**). Thus, *Bmal1-sLuc2* did not have such refractory period.
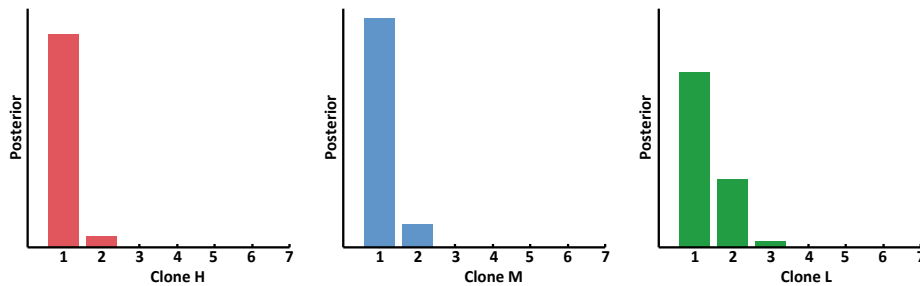


**Figure 2.19 Estimated number of *N* steps composing the "off" state**
Posterior distribution for *N*, the number of rate limiting steps composing the "off" state inferred from the single-cell luminescence traces corresponding to clones H (red), M (blue) and L (green).

Overall, inferring the bursting parameters corresponding to *Bmal1* promoter using a two-states model of gene expression on sliding window-partitioned time-traces highlighted the central role of burst frequency variation in determining the circadian oscillations in *Bmal1* transcription. More specifically, the transcriptionally inactive "off" state of the promoter was prolonged during the circadian phase with reduced expression. On the other side, the bust size remained largely invariant over the circadian period, but varied considerably between the clones. These results indicate that while the circadian phase modulates the burst frequency, the integration site controls the burst size to regulate expression levels.

2.2.5   Summary

The short-lived luminescence approach developed in section 2.1 allowed the recording of luminescence traces at the single-cell level over several circadian cycles. These time-traces, combined with the experimentally measured luciferase translation rate and protein and mRNA degradation rates, served to infer the transcriptional bursting properties of the promoter using a two-state model of gene expression. Since clones with similar global expression level displayed comparable transcriptional bursting kinetics (**Figure 2.15**), three representative clones were selected for deeper analyzes: a clone with high expression level (H, red), medium expression level (M, blue) and low expression level (L, green). The phase-specific transcriptional bursting parameters were estimated over the first circadian cycle with a 4-hours resolution using a sliding window. Interestingly, the burst frequency was mostly fluctuating over the circadian time, with peaks and troughs corresponding to the rhythmic profile of *Bmal1-sLuc2* transcripts accumulation (**Figure 2.18C**). More specifically, we observed prolonged "off" states during circadian expression trough (**Figure 2.18F**). Interestingly, this phase also displayed the highest noise in gene expression (**Figure 2.18B**). However, the burst frequency only marginally varied between the three clones

displaying different global expression levels (**Figure 2.18C**). Integration site-specific variation in expression could however be explained by clear burst size dissimilarities between the clones (**Figure 2.18D**). Indeed, the number of transcripts produced per burst episode remained relatively constant during the circadian period, but varied with the integration site in the different clones. Overall, these results indicate that burst size and burst frequency can be individually regulated to modulate expression levels. During the circadian period, the expression modulation mainly occurs through burst frequency variations, while expression levels driven by the integration site of the reporter mostly reflect changes in burst size.

To validate these observations, the following section will focus on an alternative strategy to infer bursting kinetics of the *Bmal1* promoter. This comparison using an alternative technical approach is expected to consolidate the results obtained using the short-lived luciferase strategy and confirms their meaningfulness.

## 2.3     Extracting the bursting parameters from mRNA distributions

Beside live reporters, transcriptional bursting parameters can also be inferred from static distributions of expression products (Raj and van Oudenaarden 2009; Larson et al. 2009). Indeed, the telegraph model can be fitted to a distribution of transcripts per cell (or less ideally proteins per cell) to infer the bursting parameters most likely to generate such distribution (Raj et al. 2006; So et al. 2011; Senecal et al. 2014; Kalo et al. 2015). Thus, we thought to use single molecule RNA-FISH (smRNA-FISH) to detect luciferase transcripts expressed from the *Bmal1* reporter generated in section 2.1 in individual cells. This imaging strategy only applies to fixed cells and thus looses the dynamic properties of the short-lived reporter approach. However, with smRNA-FISH, the bursting properties of the gene are directly measured at the transcript level instead of the protein. In addition to clarifying and validating the results previously obtained using the short-lived reporter, smRNA-FISH also allows direct quantification of the number of transcripts per cell. In section 2.2, such distributions were inferred from the individual traces. Thus, comparison between the inferred values and the experimentally obtained smRNA-FISH distributions should be informative regarding the accuracy of our two-state expression model. In addition, smRNA-FISH imaging of large sets of cells can be done rapidly and thus increases the throughput compared to the live reporter.

### 2.3.1   Imaging single molecules of mRNA using smRNA-FISH

To perform smRNA-FISH, cells are typically crosslinked in 3.7% formaldehyde. After permeabilization of the membranes and dehydration of the sample with 70% ethanol, the probes are hybridized to their target transcripts. This reaction is performed in presence of ribonuclease inhibitors and unspecific tRNAs used as blocking agents. During this step, the stringency can be achieved either by varying the concentration of formamide, or by changing the incubation temperature (Raj and Tyagi 2010).

The average number of *Bmal1-sLuc2* transcripts per cell was estimated to vary between ~4 to 20 transcripts, depending on the clone and circadian phase (**Figure 2.14**). Because of this low number

of mRNA per cell, the smRNA-FISH procedure was experimentally validated using more abundant transcripts. The approach was first tested using probes targeting *Neat1*. This long non-coding RNA (lncRNA) is involved in the paraspeckles nuclear sub-structure formation and stability (Sunwoo et al. 2008; Sasaki et al. 2009; Mao et al. 2011). This highly specific localization of the transcript greatly facilitates its detection since the large amounts of *Neat1* present in paraspeckles enhances the fluorescent signal. Indeed, in NIH-3T3 cells, *Neat1* smRNA-FISH probes revealed the presence of multiple bright and localized dots within the nucleus, which likely corresponded to paraspeckles (**Figure 2.20A**). The approach was further validated with *Gapdh*, an abundant transcript present in
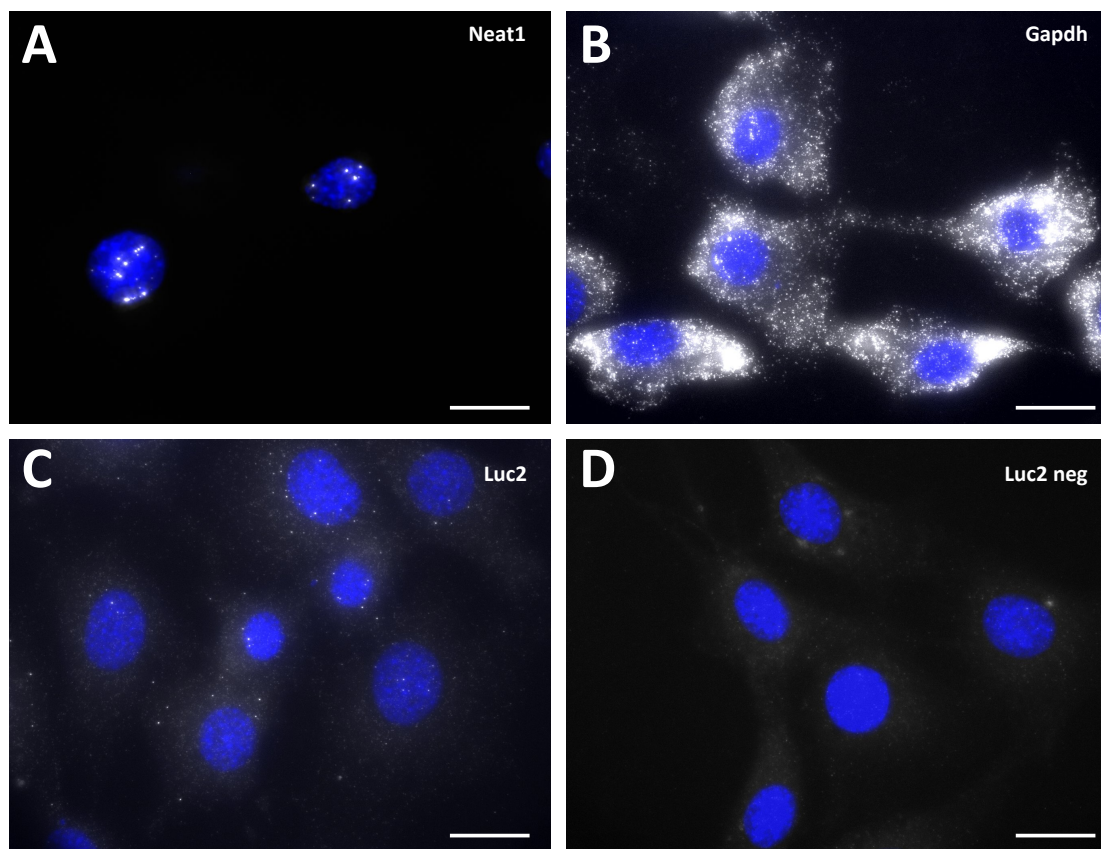


**Figure 2.20 Individual transcript detection using various smRNA-FISH probes**

Fluorescence microscopy images of NIH-3T3 cells hybridized with far-red labeled smRNA-FISH probes targeting transcripts with various properties: (**A**) *Neat1* was selected for its localized signal in paraspeckles and (**B**) *Gapdh* for the abundance of transcripts present per cell. *Luc2* targeting probes, which specifically detect luciferase transcripts including the short-lived version, were tested in the M clone expressing the transcript (**C**) or in WT NIH-3T3 cells lacking the short-lived luciferase reporter (**D**). All images correspond to Z-stack projections of at least 30 stacks (separated by 0.3 µm). Nuclei (DAPI) are shown in blue, and smRNA-FISH signal (far-red) in white. The white scale bar corresponds to 20 µm.

virtually any cell-type (Piechaczyk et al. 1984; Said et al. 2007). When cells were hybridized with smRNA-FISH probes targeting this housekeeping gene transcript, signal was abundantly detected in the entire cellular area to the point that precise quantification of number of transcripts per cell was compromised (**Figure 2.20B**). Interestingly, signal was predominantly located in the

cytoplasmic region, suggesting rapid transport of *Gapdh* mRNAs out of the nucleus. Finally, we tested luciferase (*Luc2*) probes to target the short-lived luciferase transcripts. In cells containing the *Bmal1-sLuc2* reporter, luciferase transcripts are efficiently detected (**Figure 2.20C**). However, in WT NIH-3T3 cells lacking the reporter, no transcripts were observed suggesting that the *luc2* probe is specific to luciferase transcript (**Figure 2.20D**).

Thus, smRNA-FISH provided an efficient approach to label individual molecules of specific transcripts. To reliably quantify the exact number of transcripts per cell, the experimental procedure still had to be improved. Notably, it was crucial to set up a satisfying cell segmentation technique, and to optimize culture condition.

## 2.3.2 Cells segmentation and culture optimization for smRNA-FISH

To estimate the transcriptional bursting parameters of the *Bmal1* promoter from distributions of short-lived luciferase transcript per cell, proper estimations of the cellular boundaries to accurately assign smRNA-FISH signal to a specific cell was crucial. Therefore, several cytoplasmic or cellular membrane markers supposed to efficiently delimitate the cells shapes were tested.

Wheat Germ Agglutinin Conjugates (WGA) combined with Alexa 488 fluorophore stains the plasma membrane. These lectins specifically bind gylco-residues notably present at the cellular membrane (Raikhel et al. 1984). Unfortunately the cellular permeabilization inherent to the smRNA-FISH experimental conditions allows partial penetrance of the compound into the cytoplasm, resulting in the undesirable labeling of intracellular residues (**Figure 2.21A**). Anti-Cadherin antibodies were also tested as an alternative strategy to label cellular membrane. These antibodies target conserved epitopes of central proteins for cell adhesion and are often used simultaneously to smRNA-FISH (Lyubimova et al. 2013; Bahar Halpern et al. 2015b). Although membranes were efficiently stained, the rough rendering in our experimental system complicated the identification of cellular boundaries in automated cell segmentation plateforms (**Figure 2.21B**). Dyes labeling the entire cell were also tested to segment the cell area. Carboxyfluorescein succinimidyl ester (CFSE) combined with green fluorescent dyes were notably tested for their capacity to covalently bind intracellular lysine residues (Parish 1999). Although the permeable properties of the compound and its repartition into the intra-cellular space were appreciable, the dye highlighted sub-cellular structures, thus biasing the cell segmentation tools (**Figure 2.21C**). Finally, we tested HCS CellMask staining, a cell delineation tool labeling both the cytoplasm and the nucleus. This time, the HCS CellMask staining provided satisfying results, since darker inter-cellular spaces delimitated neighboring cells while intra-cellular regions were uniformly stained (**Figure 2.21D**). Thus, HCS CellMask staining was included to the smRNA-FISH protocol to facilitate cell segmentation. It should be noted that autofluorescence from the smRNA-FISH probe channel can also be used to delimitate cellular boundaries (Raj et al. 2006, 2008), with however limited success at densities allowing physical contacts between cells.
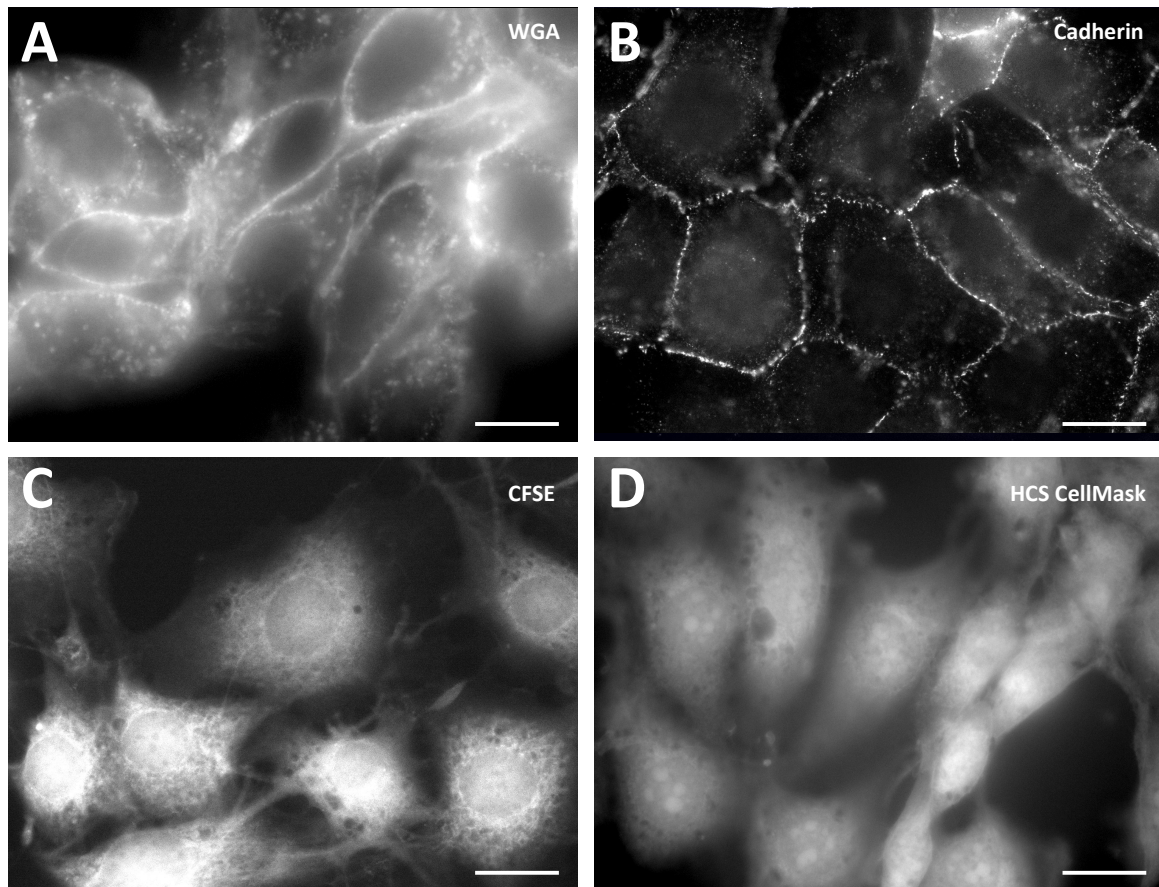
**Figure 2.21 Membrane or cytoplasmic markers for cell segmentation**
Several approaches were tested to delimitate the cellular boundaries in smRNA-FISH samples. Wheat Germ Agglutinin Conjugates (WGA) combined with Alexa 488 fluorophore was used to stains plasma membranes (**A**). Simultaneous Immunofluorescence (IF) and smRNA-FISH protocol was used to stain Cadherins covering the cellular membranes. The anti-Pan Cadherin primary antibody was then labeled using anti-rabbit Alexa Fluor 488 secondary antibodies (**B**). Carboxyfluorescein succinimidyl ester (CFSE) green fluorescent dyes were used to label intra-cellular proteins (**C**). HCS Cell Mask was also used with the objective of labeling intra-cellular space (**D**). To be compatibility with the smRNA-FISH approaches that already used far-red (probes) and blue (DAPI) fluorescent filters, all markers tested emit in the green fluorescent spectrum. Images correspond to Z-stack projections of at least 30 stacks (separated by 0.3 μm). The white scale corresponds to 20 μm.

A major drawback of HCS CellMask staining approach is its incompatibility with high cellular density. Indeed, cells overlapping or in tight physical contact cannot be differentiate using this intra-cellular dye. Yet, high cellular confluence was preconized to eliminate potential biases caused by cell growth and division during single-cell time-laps luminescence recording. Avoiding cell cycle is also crucial in smRNA-FISH experiments, since the number of transcripts per cell tightly correlates with the size of the cell (Schmidt and Schibler 1995; Marguerat and Baehler 2012; Kempe et al. 2015; Padovan-Merhar et al. 2015) (**Figure 2.22A**). Consequently, cells fixed during later time-points will overall display lower absolute number of transcripts since the higher density resulting from cell division between time-points inexorably reduces the space at disposal of each cell. Indeed, cells plated at the same time, but fixed 24 hours apart (thus during the same circadian

phase) displayed significantly reduced cell size (**Figure 2.22B**) and consistently lower absolute number of transcripts per cell (**Figure 2.22C**) in the later time-point.
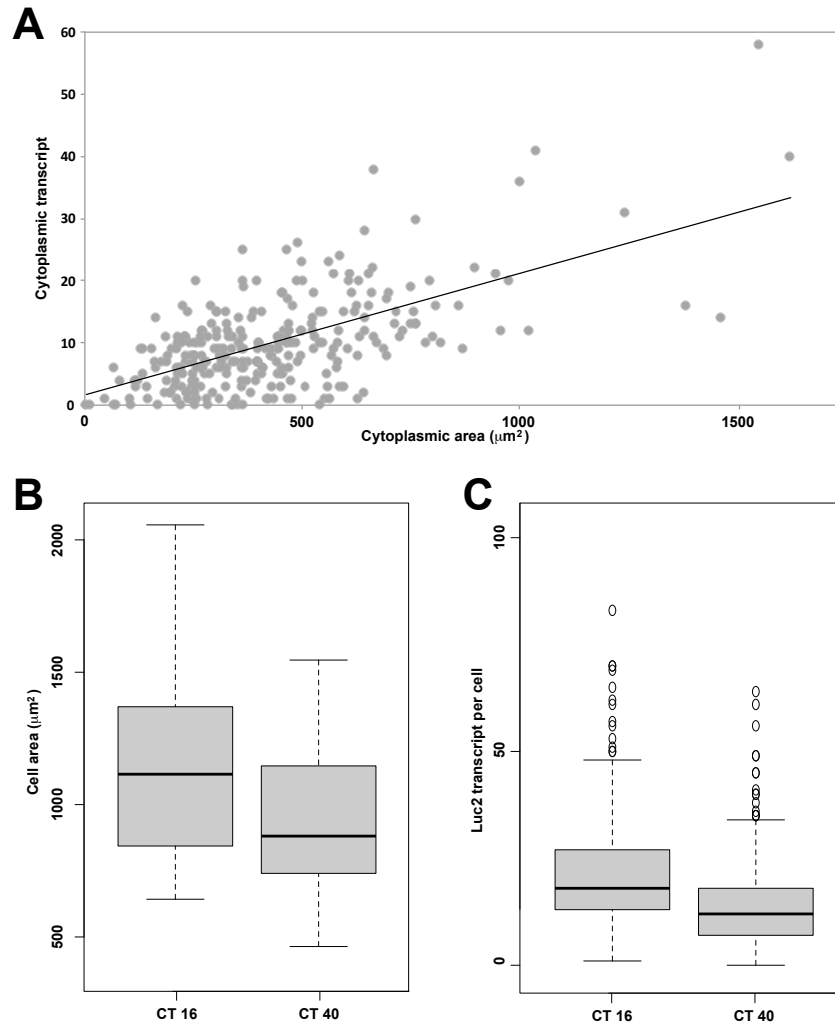


**Figure 2.22 Absolute number of transcripts per cell scales with cell area**

(**A**) Correlation between cell size (measured in cytoplasmic area) and number of *Luc2* transcript in the corresponding area (absolute number of cytoplasmic transcripts) in 290 cells (grey dots) corresponding to the L clone fixed at CT 40. The trend line is shown in black and the R-squared value of the correlation is 0.38031. Decrease in cell size (**B**) and in absolute number of *Luc2* transcript (**C**) observed between time-points fixed 24 hours apart (CT 16 and 40). N > 200 cells per condition. For all panels, surfaces were estimated from HCS mask staining and number of transcript by smRNA-FISH using *Luc2* probes.

To keep the size of the cells constant during time-course experiments while avoiding extremely confluent culture environment, we thought of performing smRNA-FISH experiments in serum-starved cells. Although cells typically cannot be indefinitely cultured in such condition and undergo apoptosis after a few days, serum starvation is known to efficiently block the cell-cycle in a dividing population (Caro-Maldonado and Muoz-Pinedo 2011; Longo and Mattson 2014), and is regularly used on NIH-3T3 cells in circadian experiments (Balsalobre et al. 1998). To identify the optimal conditions, cells were cultured in presence of different concentrations of serum. During

three days of time-course, cells were harvested every day and cell growth and viability were estimated. As expected, low serum concentrations efficiently reduced cell growth (**Figure 2.23A**). However, small serum concentrations were still sufficient to promote cell division, and as little as 0.5% of serum in the culture medium increased the cell concentration by almost 3 fold after 48 hours of culture (**Figure 2.23B**). In cells cultured in absence of serum, few division events were still observed (~140 % increase in cells between days 0 and 2 time-points). Although likely toxic in longer-term experiments, serum starved cells did not suffer excessive cell death during the three days of the experiment (**Figure 2.23C**).
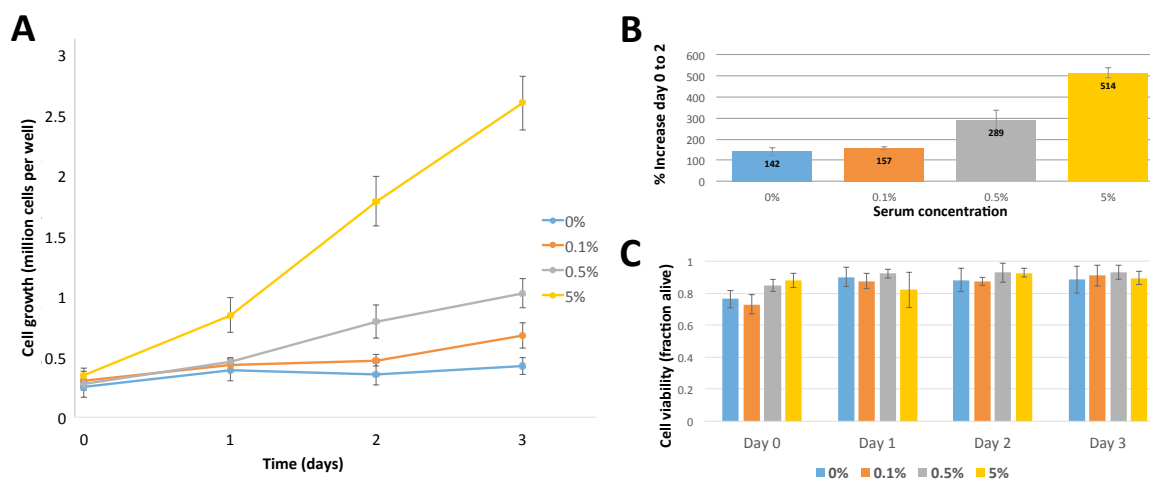


**Figure 2.23 Cell growth and survival in serum starved conditions**
Estimation of the cell growth in absolute number of cells per well (**A**), increased cell number between the day 0 and day 2 time-points (**B**) and cell viability (**C**) during a 72 hours time-course. Day 0 correspond to cells harvested 1 hour after seeding 0.5 million cells per well of a 6 well plate. Day 1, 2 and 3 correspond to cells harvested 24, 48 and 72 hours after the Day 0 time-point. Serum concentration tested: 0% (blue), 0.1% (orange), 0.5% (grey) and 5% (yellow). Indicated values correspond to mean ± SD over 3 replicate wells per condition.

In addition to preventing cell division, serum starvation is also likely to alter additional cellular pathways (Caro-Maldonado and Muoz-Pinedo 2011; Longo and Mattson 2014). To confirm that this culture condition could be used to study the transcriptional bursting properties of *Bmal1*, RNA samples were extracted up to 30 hours after DEX synchronization, and the levels of *Bmal1-sLuc2* expression were compared between serum starved and 5% serum conditions. Fortunately, serum starved cells did not altered the rhythmic expression of the *Bmal1-sLuc2* transcript (**Figure 2.24**). The amplitude of the rhythms was even slightly higher in absence of serum, notably during the first 24-hours of recording. This observation confirms that cells unable to divide display more robust circadian rhythms both at the amplitude and at the duration level (O'Neill and Hastings 2008).

Consequently, culturing cells in complete absence of serum turned out to be a suitable approach to avoid massive division while still maintaining the cells in proper conditions during at least 30 hours. Thus, to prevent side effects of cell division on the circadian cycle and transcriptional

bursting, and to avoid differences in absolute number of transcripts arising from cell-size differences, all following smRNA-FISH data were produced in 0% serum conditions.
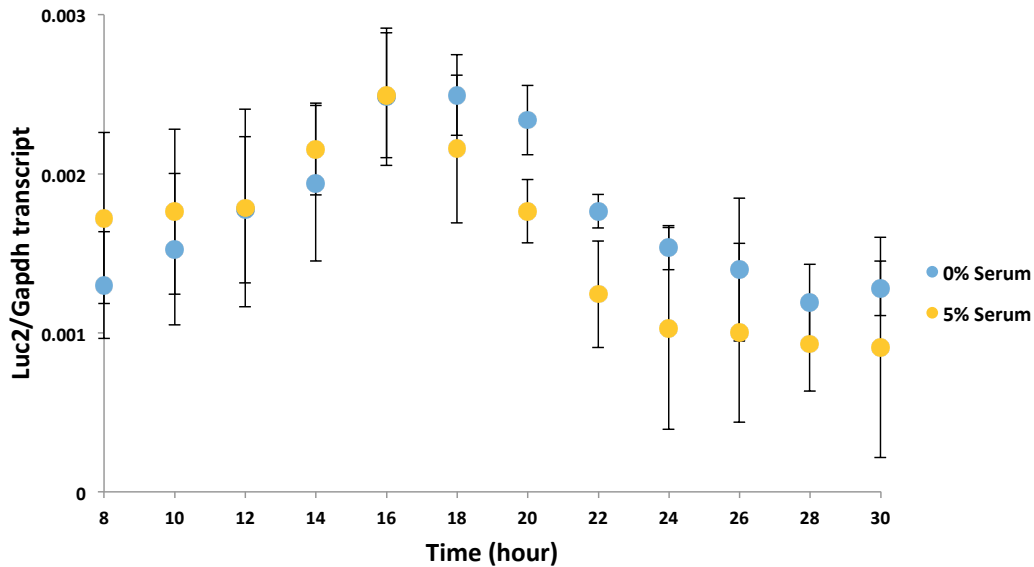


**Figure 2.24 Circadian rhythmicity in serum starved cells**

qPCR analysis of *Bmal1* short-lived luciferase (*Luc2*) transcript abundance in cells extracted every 2 hours between 8 and 30 hours after DEX synchronization. *Luc2* transcripts were normalized to Gapdh transcripts (2^-ΔCt). The cells derived from the L clone were cultured both in serum starved (0%, blue) or 5% serum (yellow) conditions. Cells were kept in respective serum condition from plating to RNA extraction. The colored dots represent the mean and the error bars the SD between 3 replicates.

### 2.3.3 Quantifying the absolute number of short-lived luciferase transcripts per cell

After testing the quality of the *Luc2* smRNA-FISH probes, the best dye to mark the shape of individual cells and the optimal culture conditions to block cell division, smRNA-FISH could be used to determine the absolute number of *Bmal1-sLuc2* transcripts present per cell. To compare the smRNA-FISH approach with previously obtained results using time-laps luminescence recording, cells in the following conditions were fixed and hybridized with Luc2 probes:

1) Cells of the H clones fixed during the peak and trough of *Bmal1* mRNA accumulation to identify changes in bursting properties along the circadian cycle.
2) Unsynchronized cells of clones H and M to compare transcriptional bursting variations driven by the reporter integration site in the two differentially expressing clones while eliminating bursting modulation caused by the circadian cycle.

Although the half-live of the short-lived luciferase protein is small and the peak of mRNA accumulation preceded that of the protein by ~20 minutes, the mRNA accumulation levels of the transcript were first measured by qPCR on reverse-transcribed RNA to confidently identify the mRNA expression peak and trough. Cells of the three selected *Bmal1-sLuc2* clones were synchronized, and RNA was harvested every 2 hours between 8 and 30 hours post-

synchronization. In each clone, accumulation of the *Bmal1* transcript followed a circadian pattern (**Figure 2.25A**). As observed in luminescence, expression levels in clone H is ~1.5 and ~2.5-folds higher than clones M and L respectively. After normalizing expression levels for each individual clones to the time-point with the highest mRNA accumulation, CT 16 consistently correspond the mRNA expression peak, while the expression trough was around CT 4 (**Figure 2.25B**). Thus, in smRNA-FISH experiments, cells were fixed at these time-points corresponding to *Bmal1-sLuc2* circadian peak and trough of expression.
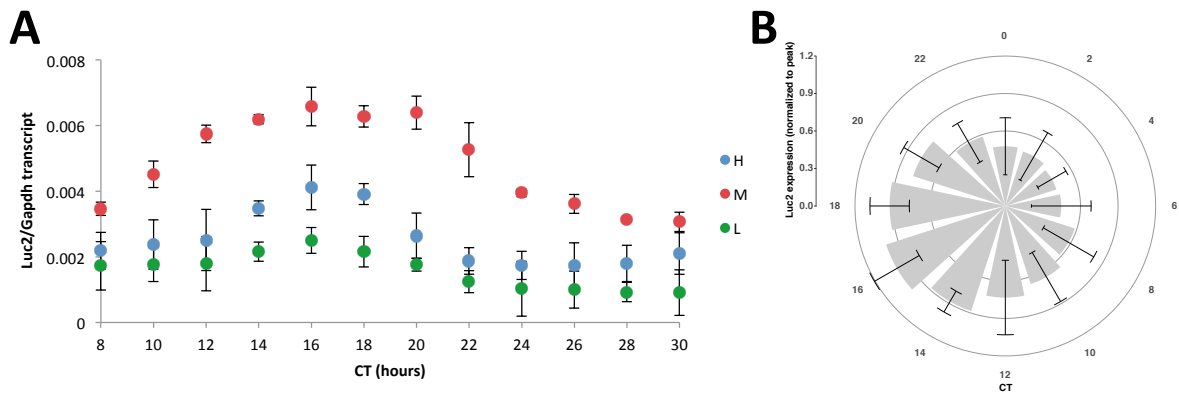


**Figure 2.25 Rhythmic accumulation of *Bmal1-sLuc2* transcripts**

(**A**) qPCR analysis of reverse-transcribed *Bmal1* short-lived luciferase (*Luc2*) transcript abundance in cells of the H (red), M (blue) and L (green) clones extracted every 2 hours between 8 and 30 hours after DEX synchronization. *Luc2* transcripts were normalized to *Gapdh* transcripts ($2^{-\Delta Ct}$). The colored dots represent the mean, and the error bars the standard deviation between 3 replicates. (**B**) Pooled mRNA accumulation of *Bmal1-sLuc2* transcripts in the H, M and L clones. Expression levels are displayed in fraction of expression peak (CT 16) in mean ± SD over 9 replicates (3 per clone).

Knowing the precise oscillatory accumulation pattern of *Bmal1-sLuc2* transcripts, cells of four experimental conditions were fixed for *Luc2* transcripts detection using smRNA-FISH (**Figure 2.26**). To identify transcriptional bursting variations occurring over the circadian cycle, cells of the H clone were harvested at *Bmal1* peak and trough of expression (**Figure 2.26A and B**). To focus on reporter integration site-mediated changes in transcriptional bursting properties, unsynchronized cells of the H and M clones were simultaneously harvested (**Figure 2.26C and D**). In each of these conditions, dots in the far-red channel corresponding to individual *Bmal1-sLuc2* transcripts could easily be identified by eye. To automatize the dot detection and their attribution to a cell, we used the CellProfiler open-source software (Carpenter et al. 2006). This software enabled precise segmentation and area estimation of both nuclei using DAPI signal (**Figure 2.27A**), and the entire cells using the HCS CellMask channel (**Figure 2.27B**). Individual transcripts were identified using the far-red channel (**Figure 2.27C and D**) and assigned to a cell.
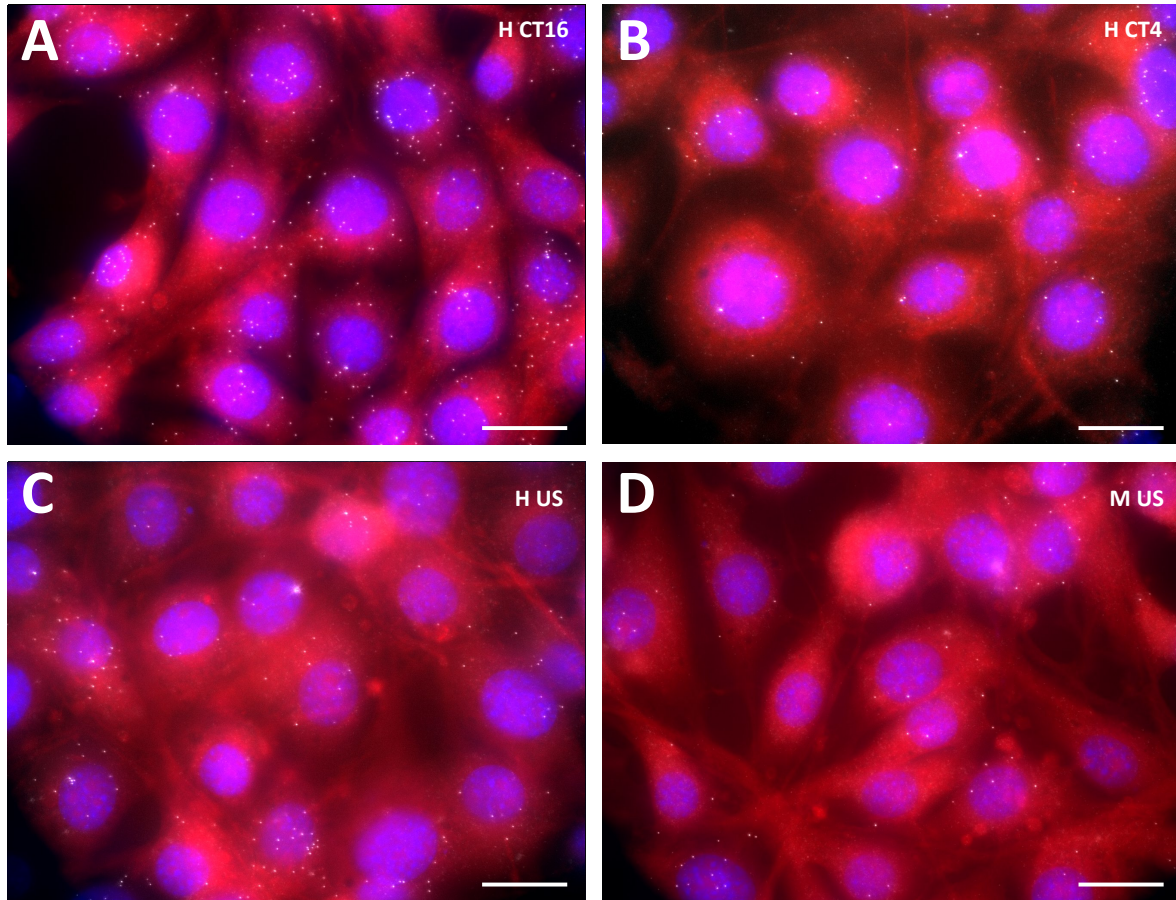
**Figure 2.26 smRNA-FISH detection of *Bmal1-sLuc2* mRNA in four experimental conditions**
Typical images resulting from the smRNA-FISH detection of *Bmal1* short-lived luciferase (*Luc2*) transcripts. Nuclei were stained with DAPI (blue) and cell area with HCS CellMask (red). *Luc2* transcripts detected in the far-red channel by the smRNA-FISH probes are shown as white dots. The different experimental conditions tested correspond to the mRNA expression peak (**A**, CT16) and trough (**B**, CT4) of *Bmal1-sLuc2* in Clone H, and unsynchronized (US) cells from the H (**C**) and M (**D**) clones harvested 16 hours after seeding. Images correspond to Z-stack projections of at least 30 stacks (separated by 0.3 μm). The white scale corresponds to 20 μm.

The quantification of individual *Bmal1-sLuc2* transcripts in individual cells permitted the obtaining of mRNA per cell distribution in a population. To validate the two-state telegraph model of gene expression and its capacity to reliably infer short-lived luciferase transcript numbers from luminescence signal, mRNA distributions per cell obtained with the telegraph model on luminescence traces were compared to that of the smRNA-FISH for the unsynchronized H and M clones. In both cases, the model-inferred distribution remarkably fit the smRNA-FISH distributions (**Figure 2.28**). Thus, the telegraph model accurately estimated the number of short-lived luciferase mRNA corresponding to a luminescent signal. Consequently, despite experimentally immeasurable parameters and the additional layer of complexity conferred by measurement done at the protein level, the short-lived luciferase system showed to present a trustworthy estimate of the transcriptional process, and a valid tool for the study of transcriptional bursting.
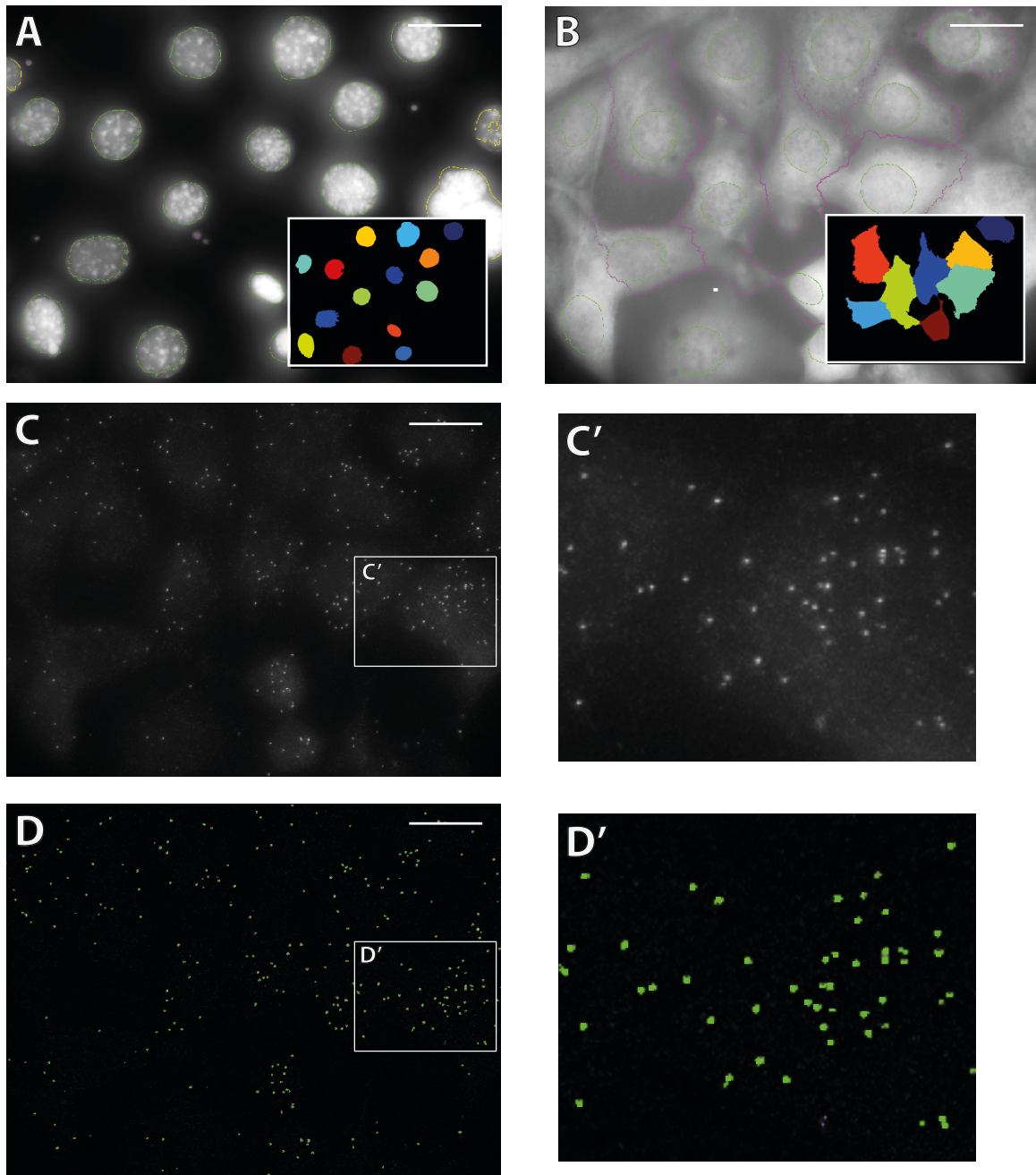
**Figure 2.27 Steps of the CellProfiler automated procedure for smRNA-FISH signal detection and attribution to cells**
To identify individual transcripts from smRNA-FISH images and assign them to a parental cell, all color channels composing the image were first split. The DAPI image served to segment nuclei (**A**, inset corresponds to the segmented image, nuclei touching image borders are not included). HCS CellMask images (green channel) were used to delimitate the cell shape (**B**, inset corresponds to the segmented image, cells touching image borders are not included). The raw far-red channel (**C**) was used to identify single mRNA molecules (**D**). (**C'**) and (**D'**) correspond to zoomed versions of the white rectangle in (C) and (D) respectively. Images correspond to Z-stack projections of at least 30 stacks (separated by 0.3 μm). The white scales correspond to 20 μm.
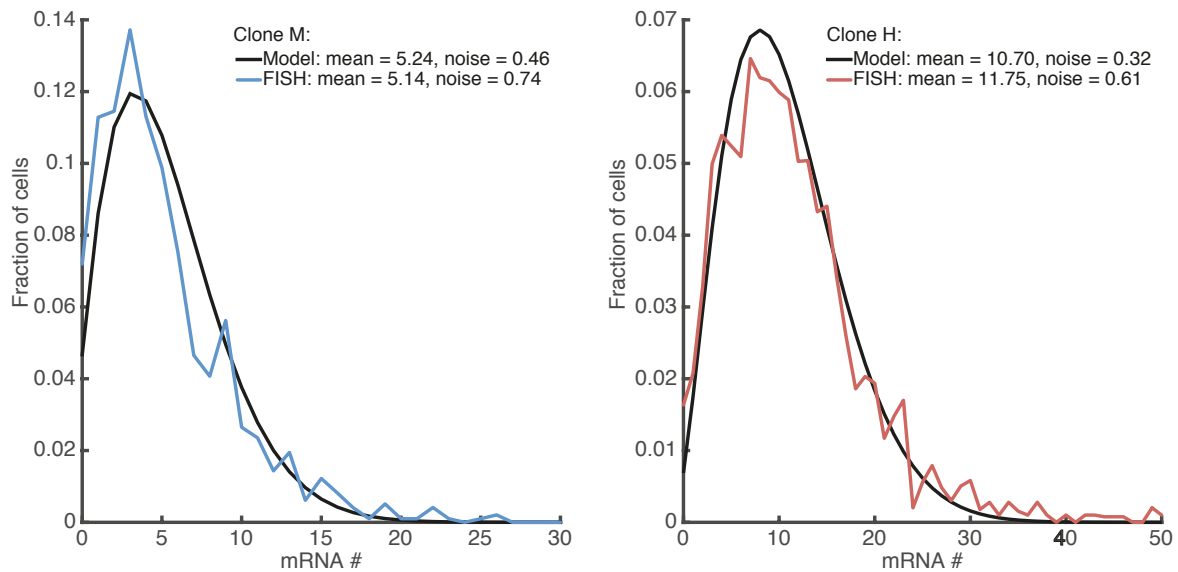
**Figure 2.28 Comparison between mRNA distributions per cell inferred from modeling on luminescence traces or obtained from smRNA-FISH**

The distributions represent the faction of cells (Y-axis) containing a specific number of transcripts (X-axis). The transcripts per cell distributions obtained from an unsynchronized populations of clone M using smRNA-FISH (blue) is compared with the inferred distribution from pooled single-cell luminescence traces of the same clone including all time-points and using the two-state model of gene expression (black left). A similar comparison is done for the H clone (red and black right). For both model-inferred and smRNA-FISH-resulting distributions, the mean number or transcript per cell and the expression noise (squared coefficient of variation) are indicated. Model distributions were generated from a minimum of 119 single-cell traces, and smRNA-FISH distributions from 950 individual cells for both clones.

## 2.3.4    Estimating bursting parameters from smRNA-FISH distributions

Ultimately, smRNA-FISH distributions were generated to provide an alternative approach for the estimation of the transcriptional bursting parameter. In particular, we aimed at confirming the trends observed using two-state model on luminescence traces showing that the circadian phase predominantly modulates the burst frequency while the integration site dictates the burst size.

Different approaches can be used to infer transcriptional bursting parameters from transcript distributions per cell. The gold-standard consists in using the two-state model of gene expression to calculate the likelihood of observing the distribution for given bursting parameters (Raj et al. 2006; Neuert et al. 2013; Dey et al. 2015). With the assumption that the "on"state is largely shorter than the transcript half-life, the steady state mRNA distributions corresponds to a negative binomial distribution whose parameters can be estimated using maximum likelihood approaches (Raj et al. 2006).

By further assuming that in addition to the mRNA half-life, the "on"states are also considerably briefer than "off" states, and that the number of transcripts per bursts (burst size) is substantial, the normalized burst frequency (in units of burst events per transcript half-life) and burst size can both be easily estimated by only using the mean and variance of the mRNA copy numbers (Weinberger et al. 2012; Dey et al. 2015). The normalized burst frequency is then inversely

proportional to the coefficient of variation squared (SD/mean)$^2$, while the burst size is proportional to the mean number of transcripts divided by the burst frequency (**Equation 4.8**).

We first applied this simplified approach to the *Bmal1-sLuc2* smRNA-FISH distributions. The resulting transcriptional bursting parameters were comparable with those obtained with real-time luminescence analysis (**Figure 2.29**). Indeed, although absolute values are not comparable since parameters inferred from smRNA-FISH distributions are expressed in transcript lifespan instead of absolute time units, the circadian phase again predominantly affected the burst frequency (although a minor variations of burst size could also be noticed between the time-points) while the difference between the H and M clones mainly arose from variations in burst size.
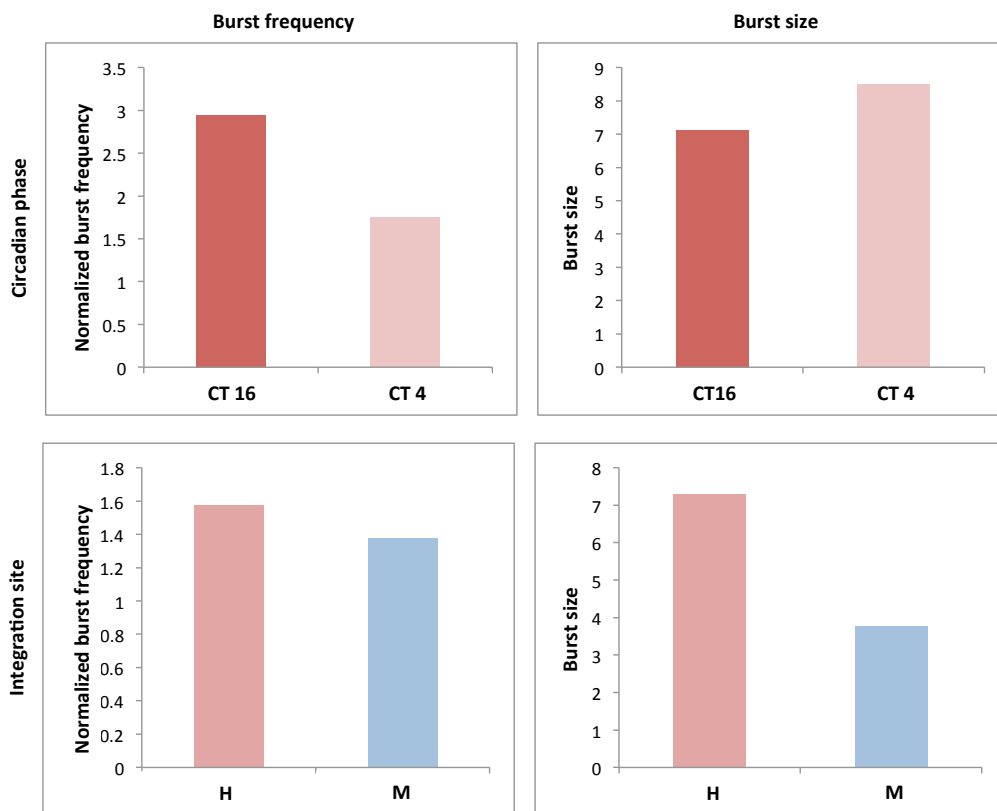


**Figure 2.29 Transactional bursting parameters inferred from smRNA-FISH distribution using "mean and variance" simplified strategy**

The burst frequency (left boxes, normalized to the transcript lifespan) and burst size (right boxes) were inferred from smRNA-FISH distributions using only the mean and variance values, where the normalized burst frequency is then inversely proportional to the coefficient of variation squared and the burst size corresponds to the mean number of transcripts divided by the burst frequency. This strategy was applied to compare the transcriptional bursting parameters between the circadian phases (upper boxes, expression peak at CT 16 and trough at CT 4) and between integration sites (lower boxes, clones H and M). Bursting parameters were estimated from a minimum of 340 cells per condition from a single experiment.
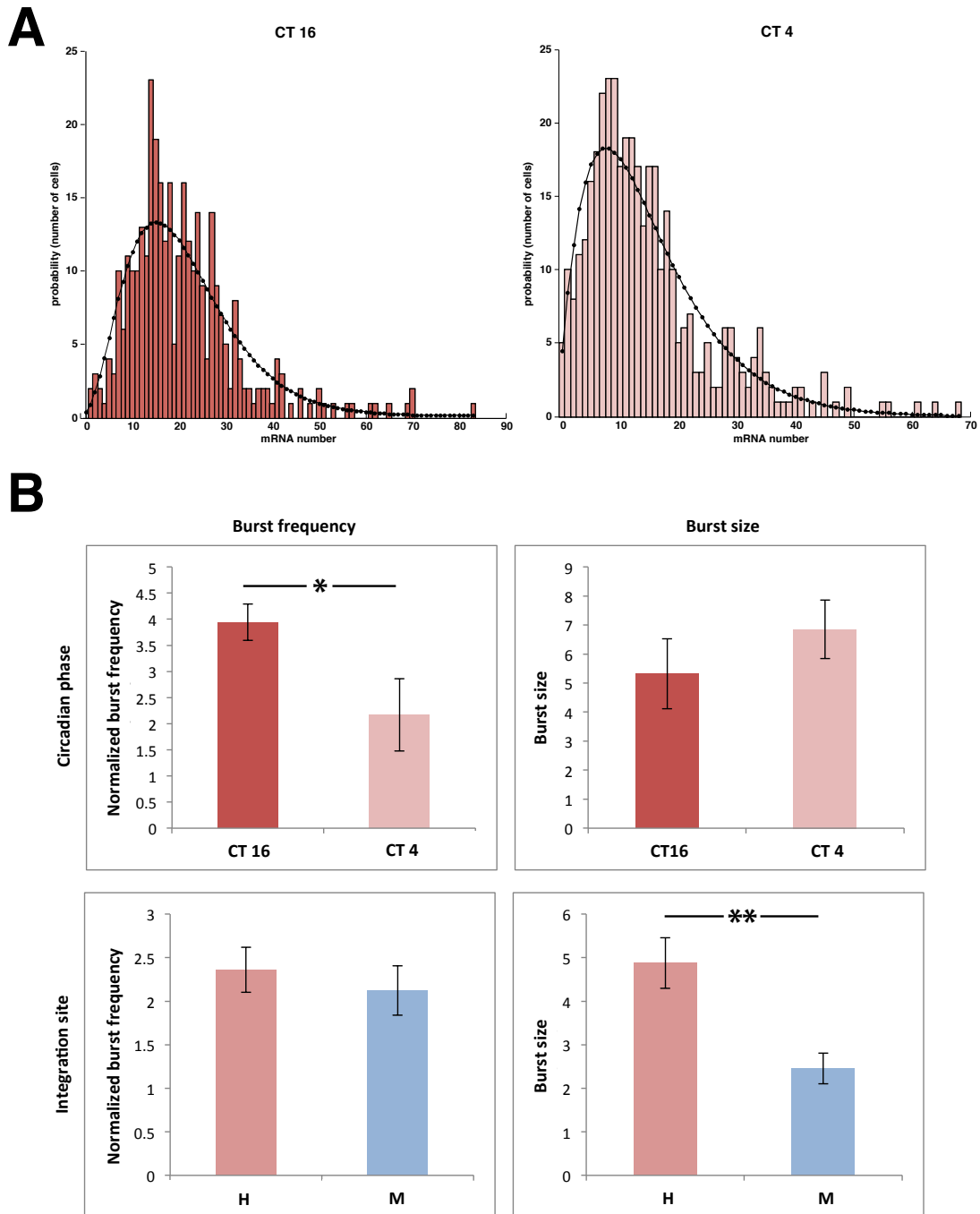
**Figure 2.30 Transactional bursting parameters inferred by fitting a two-state model of gene expression to smRNA-FISH distributions**

(**A**) Examples of smRNA-FISH distributions expressed in probability (absolute number of cells) containing N mRNA (red bars) with their corresponding fit (black curve). The two distribution displayed correspond to the H clone fixed at *Bmal1* expression peak (CT 16, left) or trough (CT 4, right). N=340 and 390 cells for CT 16 and 4 respectively. (**B**) Transactional bursting parameters maximizing the likelihood of explaining the experimental distribution. The burst frequency (left boxes, normalized to the transcript lifespan) and burst size (right boxes) are compared between circadian phases in the H clone (upper boxes, expression peak at CT 16 and trough at CT 4) and between integration sites (lower boxes, clones H and M). Bursting parameters were estimated from a minimum of 340 cells per condition. The values correspond to the mean and 5 and 95 percentiles over 20'000 MCMC iterations. * = p-value <0.05, **= p-value <0.01.

These results were further confirmed by fitting a negative binomial function to the smRNA-FISH distributions (**Figure 2.30A**) to infer the transcriptional bursting parameters having the maximal likelihood of explaining the experimental distribution (**Figure 2.30B, Equation 4.7**). This more elaborated analytical strategy provided highly similar results than the simplified version based on the distribution mean and variance, suggesting that both analytical strategies can be used to infer transcriptional bursting parameters from smRNA-FISH distributions. The statistics resulting from the multiple Markov chain Monte Carlo (MCMC) iterations indicated that only the burst frequency significantly varied between *Bmal1* peak and trough of expression. On the other hand, differences between smRNA-FISH distributions corresponding to the M and H clones were very significantly arising from burs size variations.

## 2.3.5 Summary

To substantiate the results previously obtained on transcriptional bursting modulation between different experimental conditions (circadian time and integration site) using a real-time luminescence reporter, we implemented smRNA-FISH. Although this experimental approach does not permit real-time monitoring of transcription, it quantifies mRNA, the direct product of transcription.

Several experimental details were optimized to permit reliable use of this approach to infer transcriptional bursting parameters of a circadian reporter. Notably, we used HCS CellMask, a dye that marks the inner-cell region in an even and regular manner to facilitate cell segmentation and the attribution of transcript to a cell (**Figure 2.21**). To block cell division and avoid cell-size variations between the time-points that bias comparison of absolute transcripts number per cell, serum-free culture conditions were adopted. Although it eventually provokes cell death, we observed that at short-term, cells viability and circadian rhythmicity were not affected (**Figure 2.23 and 2.24**).

Four conditions were probed by smRNA-FISH to quantify the absolute number of *Bmal1-sLuc2* mRNA per cell: the H clone fixed during the mRNA expression peak and trough (CT 16 and 4 respectively) to compare the transcripts distributions between the circadian time-points, and unsynchronized cells of the H and M clones to compare mRNA distributions in cells with a different reporter integration site (**Figure 2.26**).

The resulting transcript per cell distributions were compared to that of section 2.2 inferred from the single-cell luminescence traces (**Figure 2.288**). Since both smRNA-FISH and modeling on luminescence traces provided remarkably similar distributions, both approaches were considered appropriated for the study of transcriptional bursting.

Two types of analyzes were then tested to infer transitional bursting kinetics from smRNA-FISH distributions: a simplified approach only using the distribution mean and variance (**Figure 2.29**), or by fitting a negative binomial function (**Figure 2.30**). Both strategies converged to similar results, although the complex fit allowed a more statistically detailed analysis. These analyzed confirmed the results previously observed using short-lived luciferase live reporter: during the circadian cycle, variations in expression levels were significantly caused by changes in burst frequency, and between the clones, the aspect of transcriptional bursting significantly differing was the burst size.

Thus, both smRNA-FISH and modeling on single-cell luminescence traces represented alternative and complementary approaches to infer transcriptional bursting, and they both led to similar conclusions.

Interestingly, these results imply that both the burst size and burst frequency can be modulated to vary the expression levels. However, both strategies likely involve different transcriptional regulatory pathways. Thus, in the following sections, I will focus on the identifications of molecular markers differing between the experimental conditions (between the integration sites and between the circadian time-points) that could cause (or at least co-vary with) the observed differences in transcriptional bursting pattern.

## 2.4     Identification of molecular markers correlating with the burst size

Considering that the burst frequency and the burst size can both be independently tuned to modulate *Bmal1* expression levels strongly implies that molecularly distinct phenomenon participate in the regulation of this promoter. The molecular mechanisms varying the expression levels of this circadian promoter along the daily period preferentially modulate the burst frequency, while the transcriptional regulatory changes conferred by the integration site primarily affect the size of the bursts.

Thus, we aimed at identify specific molecular markers and more specifically mechanisms involved in either of these two expression modulation strategies. Focus was first placed in the identification of molecular mechanisms varying between the integration sites of the L, M and H clones and thus potentially participating in modulating the burst size.

### 2.4.1   Reporter integration site likely impacts the burst size

Both technical approaches used to estimate transcriptional bursting properties of the *Bmal1* promoters (i.e. the short-lived luciferase reporter and smRNA-FISH distributions) revealed that the burst size only marginally varied between the circadian time-points but significantly differed between the H, M and L clones. To confirm that this effect was related to the integration site of the reporter and could not be attributed to global differential regulation of the circadian clock in the three clones, the expression of several clock genes was tested. mRNA levels of three endogenous clock genes were measured by qPCR on reverse-transcribed RNA at their respective circadian expression peak. While the amount of *Bmal1-sLuc2* mRNA strongly varieed between the three clones (**Figure 2.31A**), *Cry1* and *Dbp* mRNA levels remained fairly similar (**Figure 2.31B and C**). As for the reporter, endogenous *Bmal1* levels significantly differed between the three clones (**Figure 2.31D**). However, since endogenous *Bmal1* expression levels were inversely proportional to those of *Bmal1-sLuc2*, the observed variations between the clones were unlikely caused by clone-specific features affecting the circadian clock and more particularly *Bmal1* expression. The variations in endogenous *Bmal1* expression levels could in contrast result from a competition phenomenon where limiting *Bmal1* regulators would be less available at the endogenous loci in clones displaying high expression levels of the reporter. However, this hypothesis could not explain the variation in expression levels observed at the reporter level. Since essential

components of the circadian clock were similarly regulated in the three clones, the differences observed at the *Bmal1-sLuc2* expression level (and particularly at the burst size level) likely directly arose from the clone-specific integration site of the reporter.
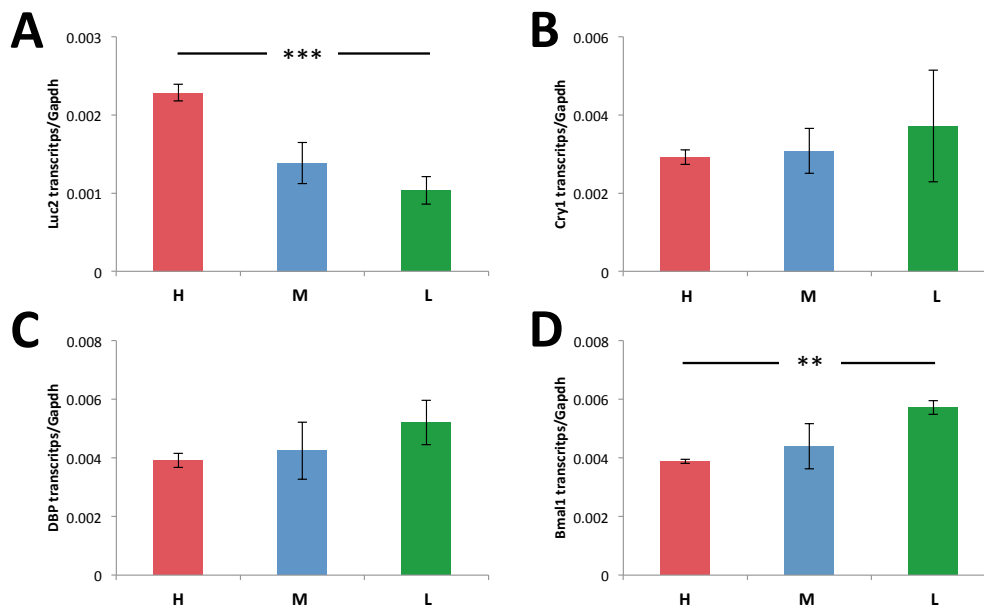


**Figure 2.31 Comparative expression of clock genes in H, M and L clones**
qPCR analysis of the expression levels of various circadian genes at their respective expression peak in the H (red), M (blue) or L (green) clones: (**A**) *Bmal1* short-lived luciferase reporter (*Luc2*), (**B**) endogenous *Cry1*, (**C**) endogenous *Dbp* and (**D**) endogenous *Bmal1*. Ct values for the selected genes were normalized to Ct values of *Gapdh* transcripts (2^-ΔCt). Statistics were performed using one way Anova: \*\*= p-value <0.01, \* \*\*= p-value <0.001.

## 2.4.2   Molecular marker enrichment at the promoter of the reporter

To identify molecular markers displaying marked enrichment variations at the promoter region of the luminescence reporter in each of the three clones, we implemented a Chromatin ImmunoPrecipitation assay (ChIP). Theoretically, this approach could identify markers which abundance correlates with burst size levels. Such markers could thus possibly be linked to the molecular mechanisms responsible for the burst size-driven expression modulation. Since the sequence of the reporter was identical in each clone, these burst size differences likely arose from general chromatin context differences of the integration site. Thus, three histone marks characteristic of specific chromatin contexts were selected as candidate markers. H3K27ac is an established marked of active promoters. In contrast, H3K27me3 and H3K9me3 are both associated with repressed chromatin state, as they are associate with Polycomb and HP1 respectively (Lawrence et al. 2016). In addition, we also tested the abundance of histone H3 to discriminate between authentic enrichment of a specific histone post-translational modification and variations in nucleosome (and thus H3) occupancy. The presence of these histone marks was tested in all three clones (H, M and L), and at three loci:

1) The promoter region of the *Bmal1* luminescence reporter to examine the variation of histone mark enrichment between the three clones. Within the *Bmal1-sLuc2* locus, two

locations termed regions A and B and located in -149/-270 and -1274/-1357 to the TSS respectively, were analyzed using specific primers.

2) The promoter region of a *Actb*, a highly expressed gene used as a control for elevated burst size (Bahar Halpern et al. 2015b).

3) The promoter region of *Sox2*, a transcription factor involved in pluripotency and not expressed in NIH-3T3 cells. It was thus used as a negative control (He et al. 2004; Plautz et al. 2011).

The enrichment of active histone marks at the *Actb* promoter and the presence of repressive marks around *Sox2* locus suggested that the ChIP experiment worked properly (**Figure 2.32**). Additionally, the similarity in histone H3 abundance observed between each clone indicated comparable nucleosome occupancy at the *Bmal1* reporter loci (**Figure 2.32D**). Nevertheless, no obvious enrichment of H3 post-translational marks correlated with previously measured changes in burst size. Indeed, the high variability between replicates and the inconsistencies between histone marks levels quantified at the two *Bmal1-sLuc2* promoter regions despite close physical proximity challenged the use of ChIP-qPCR to highlight variations in histone marks enrichment. This approach was indeed not sufficiently sensitive to detect putatively slight differences in molecular marker abundance that could explain the clone-specific burst sizes.
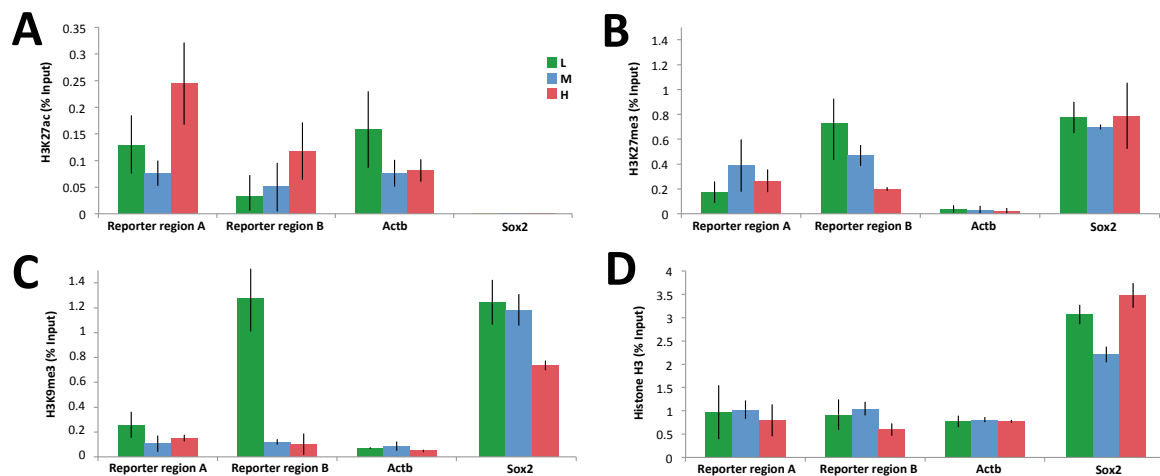


**Figure 2.32 Histone H3 marks enrichment at the reporter promoter level of the three clones**
ChIP-qPCR analysis to determine the abundance of (**A**) H3K27ac, (**B**) H3K27me3, (**C**) H3K9me3 or (**D**) histone H3 in the three clones L (green), M (blue) and H (red) and at four genomic location: the +149/270 region of the *Bmal1* reporter (Reporter region A), the +1274/1357 region of the *Bmal1* reporter (Reporter region B), the promoter region of the expressed *Actb* gene and the promoter region of the inactive *Sox2* gene. For all clones, cells were harvested at CT 16 (*Bmal1* expression peak). All enrichments are displayed in immunoprecipitated percentage of the input chromatin material. The values correspond to the mean ± SD over three replicates.

### 2.4.3    Molecular marker enrichment at the integration site

Because the nature and sequence of the reporter integrated into the genome is identical in the three clones, the locus-specific features influencing the burst size were likely already present at these genomic locations before the reporter integration. Knowing the precise integration site of

the reporter would then provide information on specific characteristics of these genomic regions, notably by using public datasets such as ChIP-seq.

The exact integration site of each clone was determined by inverse PCR (iPCR). This approach consisted in digesting gDNA with a restriction enzyme to release hybrid DNA fragments containing both a fraction of mouse gDNA and a portion of integrated reporter (Jong et al. 2002). After circularization of the DNA fragments using diluted ligation, amplifications of hybrid fragments was achieved using inverse primers targeting known sequence in the reporter fragment. Amplified regions were then sequenced to reveal the identity of the unknown mouse genomic region.

In addition to the usual clones H, M and L, the integration site was also determined for two additional clones displaying low expression level (clones 1 and 8) to increase the robustness of the subsequent analysis. Once the exact location of the FRT cassette determined (**Table 2.2**), public functional genomics databases were used to correlate the reporter burst size of the five clones with various markers enrichment around their integration site. For NIH-3T3 cells, such public repositories comprised DNase I hypersensitive site datasets (Dunham et al. 2012) as well as ChIP-seq dataset for various histone post-translational modifications or transcription factors such as H3K4me1, H3K27ac, H4ac and H3K9me3 (Zhu et al. 2012) as well as CTCF, H3K4me3, H3K36me3 and H3K27me3 (Zullo et al. 2012).

| Clone | Integration site (mm9) | Orientation |
|---|---|---|
| **Clone 1** | chr5:125,446,362 | Sense |
| **Clone 4 (M)** | chr10:4,954,129 | Antisense |
| **Clone 7 (L)** | chr5:14,911,612 | Antisense |
| **Clone 8** | chr2:125,929,533 | Sense |
| **Clone 15 (H)** | chr2:46,374,969 | Sense |

**Table 2.2 Genomic location of the integration sites of various clones**

Inverse PCRs (iPCR) were used to determine the integration site of the FRT cassette (and thus the *Bmal1-sLuc2* reporter) on the mm9 version of the mouse genome. The orientation of the FRT cassette on the genome is also indicated.

Comparison between molecular markers abundance at different windows around the integration site and the burst frequency of the *Bmal1-sLuc2* in the same clones resulted in a correlation heat map (**Figure 2.33A**). None of the assessed molecular marker provided a significantly positive or negative correlation with burst size (p-values ≥ 0.149), and most correlation coefficients were small. Also, the low number of clones present in this study considerably reduces the robustness of the analysis. However, some trends were perceptible. Notably, most molecular markers with known roles in chromatin compaction such as H3K27me3 or H3K9me3 were depleted in clones with elevated burst sizes (**Figure 2.33D**). The inverse observation is not necessarily true, since most markers positively affecting transcription such as H3K27ac or H4ac did not correlate with the burst size (**Figure 2.33C**). A notable exception is the presence of DNase I hypersensitive sites, which positively correlated with the burst size (**Figure 2.33C**). However, the correlation coefficient are mainly influenced by the H clone displaying a higher burst size than the other clones, and may thus not represent general trends.
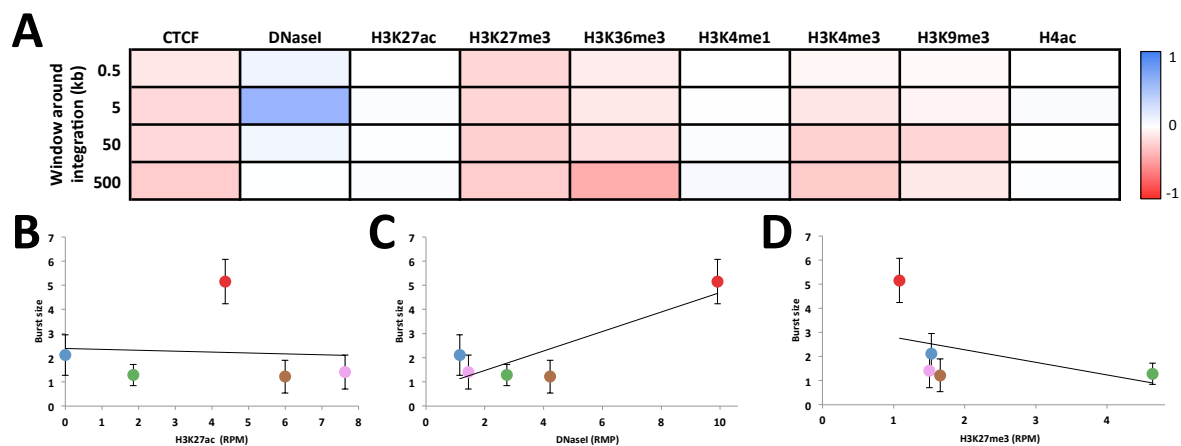
**Figure 2.33 Correlation between abundance of molecular markers at integration site and burst size**

Burst size values corresponding to five *Bmal1-sLuc2* clones (inferred using the real-time short-lived luciferase approach) were compared with enrichment of functional genomic markers in four windows around the integration site (0.5, 5, 50 and 500 kb) in reads per million mapped reads (RPM). (**A**) Resulting correlation heat map for the 9 molecular markers, with positive correlations indicated in blue and negative ones in red. (**B-D**) Examples of correlation (corresponding to the 5kb window around integration site) including the five clones (Clone 1 in brown, Clone 4 (M) in blue, Clone 7 (L) in green, Clone 8 in pink and Clone 15 (H) in red). The trend line is indicated in black. Specific examples are shown corresponding to (**B**) an absence of correlation (H3K27ac, $R^2$=0.00475), (**C**) a positive correlation (DNaseI, $R^2$=0.74957) and (**D**) a negative correlation (H3K27me3, $R^2$=0.20482).

In conclusion, although no clear molecular marker was found to efficiently correlate with the burst size of the different clones, a trend indicated that this transcriptional bursting property could be affected by the global chromatin state. In particular, the presence of heterochromatin around the integration site seemed to participate in the establishment of a low reporter burst size. Inversely, the correlation between DNase I hypersensitive sites and burst size indicated that the presence of neighboring DNA binding sites positively affects the burst size of the reporter.

### 2.4.4 Summary

After identifying the burst size as the bursting parameter most sensitive to expression differences between the *Bmal1-sLuc2* clones, several approaches were tested to identify molecular markers correlating with integration site-specific burst size. Such markers could indeed participate in the molecular mechanisms involved in burst size modulation.

Since the expression level of endogenous circadian genes was comparable in all three tested clones (**Figure 2.31**), the observed differences in burst sizes were considered to arise from integration site specificities. Global chromatin state markers were thought to be reasonable first-approach candidates, and the presence at the reporter promoter of both heterochromatin and active histone marks were tested using ChIP. Unfortunately, the reproducibility of the assay and consistency between the reporter regions were too poor to drive any conclusions (**Figure 2.32**).

Since the integrated sequences are identical for all clones, the differences causing changes in burst size were probably present at the locus before the integration of the FRT cassette and the reporter. Identification of the genomic sites of integration allowed the quantification of specific

genomic features at each locus using public databases (**Figure 2.33**). Again, no major candidate was found to explain burst size differences between the genomic locations. Although general heterochromatin histone marks may participate in decreasing the burst size while the presence of bound DNA elements in a close neighborhood may be linked to slightly higher burst sizes, molecular markers distinctly correlating with the size of the bursts remain to be identified. The possibility that burst size is modulated by biological phenomenon undetectable with functional genomic markers should also be considered.

## 2.5    Identification of molecular markers correlating with burst frequency

As previously attempted with the burst size, we next thought to identify molecular mechanisms involved in expression level variation through modulation of the burst frequency. From our previous analysis, it appeared that such expression regulation phenomenon occurs over the circadian period in the *Bmal1-sLuc2* reporter. Indeed, both single-cell luminescence traces and smRNA-FISH approaches highlighted that during the peak of expression, the promoter was bursting more frequently than during the circadian trough while the burst size was only marginally varying over the circadian period. Based on prior knowledge on the circadian regulation of the *Bmal1* promoter, we tested the impact of various factor such as nuclear receptor transcription factors, RORE regulatory DNA elements or histone acetylation on the burst frequency.

### 2.5.1    Modulating the activity of nuclear receptor transcription factors

The rhythmic expression of *Bmal1* is tightly linked to the presence of two RORE close to the TSS (Preitner et al. 2002; Yin and Lazar 2005). These regulatory elements recruit two families of nuclear receptors transcription factors: REV-ERBs that repress *Bmal1* expression (Harding and Lazar 1993; Sierk et al. 2001), and the RORs activators (Akashi and Takumi 2005; Guillaumond et al. 2005). Long considered as orphan nuclear receptors, it is now known that heme acts as a natural ligand for the REV-ERB family (Raghuram et al. 2007; Yin et al. 2007), whereas the ligands of RORα and RORγ consist in various sterol compounds (Kallen et al. 2002; Wang et al. 2010a). Since these discoveries, many efforts have been invested in the development of synthetic ligands able to modulate the activity of RORs and REV-ERBs (Kojetin and Burris 2014). Considering their role in *Bmal1* circadian regulation, the presence and activity of these nuclear receptors was considered to be good candidate for co-varying with the burst frequency variations of the *Bmal1* reporter along the circadian cycle. To modulate the activity of RORs and REV-ERBs and assess their impact on *Bmal1* transcriptional bursting pattern, several nuclear-receptor ligands were tested. The effect of these compounds on *Bmal1-sLuc2* expression was tested at the population level by recording luminescence signal over three days in presence of various concentrations of the drugs (**Figure 2.34**). SR8278 was selected for its antagonistic properties on REV-ERBα and REV-ERBβ (Kojetin et al. 2011). Consequently, increased expression levels of the *Bmal1* reporter should be observed upon SR8278 treatments as the compound blocks the activity of the repressor.  Indeed, although SR8278 caused cytotoxicity at high concentrations, appropriate doses (around 5 μM) positively impacted *Bmal1* expression levels (**Figure 2.34A**). Unfortunately, the impact of the drug

was not visible before 1.5 days of recording. Consequently, the direct effect of SR8278 on the observed modulation of *Bmal1* reporter expression could not be certified. Indeed, a global missregulation of the circadian clock in presence of the drug could also cause a similar phenotype. Additional compounds affecting the activity of nuclear receptors such as 7-ketocholesterol, a natural ligand of RORα and RORγ (Wang et al. 2010b), were tested. Since expression activation properties of RORs on downstream gene targets are inhibited by ligand binding (Wang et al. 2010b, a), presence of this compound was expected to decrease *Bmal1-sLuc2* expression level. Similarly, T0901317 is an inverse agonist of RORα and RORγ known to suppresses expression in target promoters (Kumar et al. 2010). Unfortunately, in both cases low concentration did not impact the expression of the *Bmal1* reporter whereas high concentrations caused massive cell death (**Figure 2.34B-C**). While no in-between effects were observed with 7-ketocholesterol, intermediate concentrations of T0901317 (around 10 µM) caused a mild decrease in *Bmal1* expression that could possibly correspond to the predicted impact of the drug. Unfortunately, using similar concentrations on an H1 promoter whose regulation did not depend on RORs activity led to a similar decrease in luminescence signal (**Figure 2.34D**). Thus, the impact of T0901317 observed on *Bmal1* expression at a drug concentration of 10 µM was likely already reflects cytotoxicity.
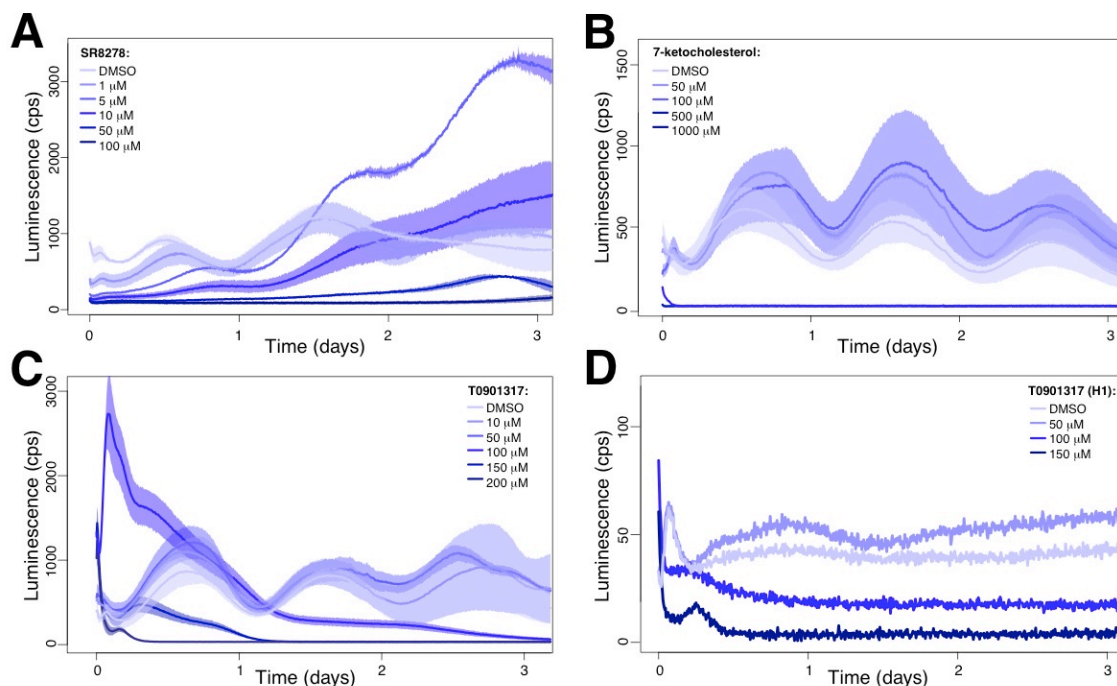


**Figure 2.34 Effect of nuclear receptor ligands on *Bmal1-sLuc2* reporter**

Impact on *Bmal1lsLuc2* expression of compounds affecting the activity of nuclear receptors were tested at different concentrations (blue gradients) on a population of M clone cells (**A-C**). Populations were recorded for three days in a lumicycle. The traces correspond to the average ± SD over 3 replicates. Compounds tested are (**A**) SR8278 (REV-ERBs antagonist), (**B**) 7-ketocholesterol (RORs natural ligand) and (**C**) T0901317 (RORs inverse agonist). (**D**) Similar concentrations of T0901317 tested on an H1 promoter driving short-lived luciferase expression. Lines correspond to a single recording.

In summary, none of the compounds tested to modulate the activity of REV-ERBs and RORs provided satisfying results. Indeed, their effects were either delayed (SR8278) or nonexistent (7-ketocholesterol and T0901317) in our system. Alternative strategies where thus chosen to modulate the rhythmic expression of *Bmal1*.

## 2.5.2 Mutating ROREs in *Bmal1* promoter region

Since we could not affect the expression of the luciferase *Bmal1* short-lived reporter by modulating the activity of its direct regulators, we tried to prevent their binding onto the promoter. Both RORs and REV-ERBs bind to the same DNA motive, RORE, present in two copies in the promoter region of *Bmal1*. To test the role of this promoter in driving *Bmal1* expression, both ROREs were mutated to completely prevent RORs and REV-ERBs binding (**Figure 2.35**) (Akashi and Takumi 2005; Guillaumond et al. 2005). The mutated *Bmal1-sLuc2* promoter was then stably integrated into the FRT site of Clone H to be compared with the WT expression pattern.
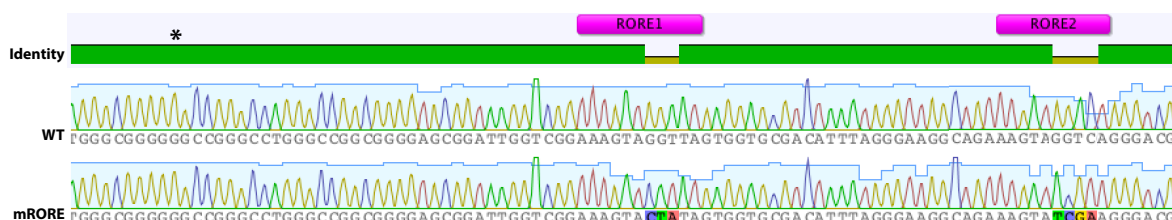


**Figure 2.35 Mutation of the RORE on *Bmal1* promoter (mRORE)**

Sequencing results confirming the presence of the two mutated ROREs in the promoter region of *Bmal1* (lower sequence) compared to the WT version (upper sequence). Both ROREs are represented by pink boxes, and the TSS is indicated as a star.

At the population level, the double RORE mutation (mRORE) clearly abrogated the rhythmic expression pattern of the WT reporter (**Figure 2.36**). Interestingly, mRORE expression levels were maintained elevated, at level close to that of WT circadian peaks. This indicated that the rhythmic regulation of *Bmal1* by the ROREs is mainly driven by repression. Thus, the RORα and RORγ activators probably only play minor roles in *Bmal1* regulation in NIH-3T3 cells compared to the REV-ERBs repressors. The elevated yet stable mRORE expression level also implied that additional regulatory elements responsible for basal expression rather than driving rhythmic oscillations were present on the *Bmal1* promoter. However, their identity, to our knowledge, remains to be determined.

The impact of the double RORE mutation was also assessed at the transcriptional bursting level. To this aim, single-cell luminescence traces of the mRORE condition were monitored. The 24 hours traces were then separated into two 12 hours fractions corresponding to *Bmal1* peak and trough of expression (centered on CT 16 and 4 respectively). These traces corresponding to semi-circadian periods were analyzed separately to infer the transcriptional bursting parameters corresponding to the WT reporter peak and trough as well as the corresponding periods in mRORE. As previously observed, the expression variations between *Bmal1* peak and trough could essentially by explained by changes in burst frequency over the circadian cycle (**Figure 2.37**). In the double mutant however the transcriptional bursting characteristics remained remarkably unchanged

between the CT 16 and CT 4 phases. Although not identical, the mRORE bursting characteristics were much closer to those of WT *Bmal1-sLuc2* circadian peak phase. Together, these results indicate that the RORE motifs in the *Bmal1* promoter are required for the decrease in burst frequency characteristic of the trough expression time.
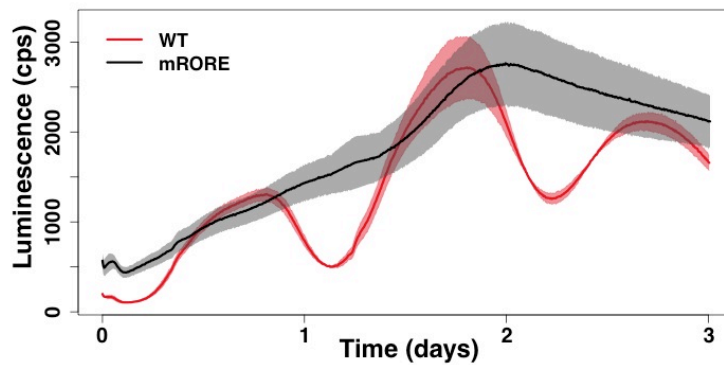


**Figure 2.36 Population luminescence driven by the mRORE *Bmal1* promoter**
Real-time luminescence comparison between populations of Clone H cells stably carrying the WT or double RORE mutant (mRORE) version of *Bmal1* promoter driving the expression of the short-lived luciferase. The data are displayed as mean ± SD over 3 replicates.
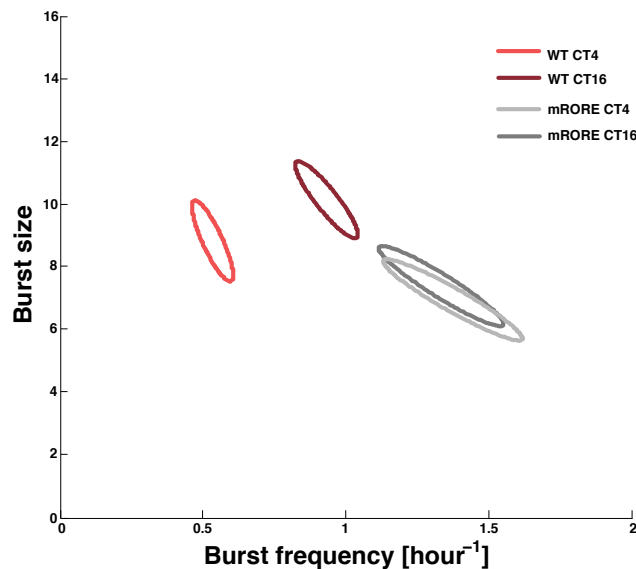


**Figure 2.37 Transcriptional bursting parameters of the mRORE *Bmal1* promoter**
12-hour traces centered on CT 16 (WT *Bmal1* expression peak, dark color) or CT 4 (WT *Bmal1* expression trough, light color) were used to infer transcriptional bursting parameters using the telegraph model. Comparison between burst frequency and size variations in the WT (red) or mRORE (grey) *Bmal1* promoter. Ellipses delimit the 5th and 95th percentiles of the estimate over all MCMC iterations. N>100 single-cells per condition.

### 2.5.3  Rhythmic regulation of *Bmal1* involves histone acetylation

Since the mRORE *Bmal1* arrhythmic expression pattern resembled more that of *Bmal1* WT expression peak, the double mutation likely prevented the binding of REV-ERBα and/or REV-ERBβ

to the *Bmal1* promoter. The mode of action of these repressors typically involves the recruitment of NCoR1 followed by HDAC3 to de-acetylate histones around the promoter (Everett and Lazar 2014). Since the RORE motifs and the recruitment of the REV-ERBs transcription factors are linked to the rhythmic changes of burst frequency in *Bmal1* expression, the downstream molecular consequences of their recruitment are also likely involved. Thus, we tested the impact of histone acetylation state on the *Bmal1* reporter expression. Populations of cells expressing *Bmal1-sLuc2* were treated with Trichostatin A (TSA), a histone deacetylase inhibitor notably affecting HDAC3 activity (Yoshida et al. 1990). Since histone acetylation is a mark of active transcription, and the role of REV-ERBs on *Bmal1* promoter is to eventually deacetylate the locus, the HDAC inhibitors was expected to globally increase expression levels by preventing the removal of the acetylation groups on histone tails. We indeed observed that high concentrations of TSA (corresponding to 1 µM final concentration) significantly increased *Bmal1* expression levels (**Figure 2.38**). Intermediate concentrations (0.5 µM) also activated *Bmal1* expression, but only during circadian troughs. Indeed, moderate concentrations of this compound abrogated the rhythmicity, while keeping the expression constant at levels comparable to the circadian peak of the untreated condition. The similarity between these phenotypes triggered by intermediate concentration of TSA and REV-ERBs incapacity to bind to *Bmal1* promoter (mRORE condition) implies that the expression pattern of the double RORE mutant simply reflects impaired histone acetylation state. Thus, histone acetylation levels were considered as candidate marks co-oscillating with burst frequency variation.
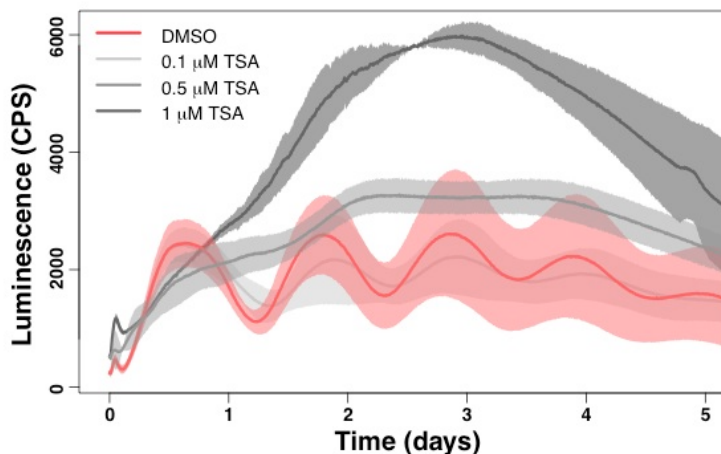


**Figure 2.38 Trichostatin A (TSA) impact on *Bmal1* promoter**

Real-time luminescence monitoring of *Bmal1-sLuc2* expressing cells from clone H in presence of various concentrations of the Trichostatin A (TSA) histone inhibitor. The data are displayed as mean ± SD over 3 replicates.

### 2.5.4 Rhythmic histone acetylation at *Bmal1* promoter

Since both histone deacetylase inhibitor and the mutation of the two RORE in *Bmal1* promoter displayed a similar arrhythmic expression pattern corresponding to the expression peak of WT conditions, histone acetylation state was considered as a candidate to correlate with burst frequency. Indeed, the burst frequency variations over the circadian cycle observed in the *Bmal1* promoter were lost in the mRORE condition that, at the population level, behaved similarly as the

TSA-treated reporter. To verify that histones around *Bmal1* promoter were rhythmically acetylated during the circadian cycle, we used Chromatin ImmunoPrecipitation sequencing (ChIP-seq) to assess the H3K27 acetylation levels in the H, M and L clones at two time-points corresponding to *Bmal1* peak and trough of expression. The reads were mapped on a custom genome corresponding to a ~10 kb region of the genomic short-lived luciferase expression cassette centered on the *Bmal1* promoter, and normalized to the total number of mapped reads on the reference mouse genome for each condition.

If histone acetylation state truly correlated with burst frequency, higher histone acetylation levels would be expected during *Bmal1* peak of expression. However, since most inter-clonal variability in transcriptional properties resided at the burst size level, little variations of histone acetylation should be observed between the clones. Reads mapping onto the stably integrated reporter region revealed that in each clones and at both time-points, most of the H3K27ac signal located within the *Bmal1* promoter, in a 3kb region upstream of *Bmal1* promoter, or downstream of the luciferase gene (**Figure 2.39A**). Since the selected 10kb region mainly contained untranscribed regions or genes used for the maintenance of the cell line (such as antibiotic resistances), a ~2kb region centered on the *Bmal1* promoter was further selected to be analyzed in depth. Within the *Bmal1* sequence, three H3K27ac peaks could be identified. The two upstream ones corresponded to the *Bmal1* promoter region, while the third peak targeted *Bmal1* first exon. For each of these peaks, variations in H3K27ac levels could be observed between the two-time points in all three clones, with CT 16 displaying higher acetylation signal than the circadian trough time-point (**Figure 2.39B**). Unfortunately, these three peaks mapped to unspecific regions of the promoter, present both on the endogenous *Bmal1* loci and in the integrated reporter. To specifically focus on the histone acetylation state of the *Bmal1-sLuc2* reporter, a 100 bp reporter-specific region downstream of *Bmal1* sequences was selected to quantify the histone acetylation state. Considerably fewer reads mapped to this region since a single copy of the *Bmal1-sLuc2* reporter was present per cell for four copies of the endogenous *Bmal1* loci in tetraploid NIH-3T3 cells (Leibiger et al. 2013). Despite the weak H3K27ac signal in the reporter-specific region, the enrichment of histone acetylation varied between the circadian time-points (**Figure 2.39C**). Indeed, although these observations were not significant, in each clone the peak of expression globally displayed more histone acetylation signal than the trough. The variations between time-points were not perfectly equivalent in each clone, but the overall histone acetylation state measured in each clone after pooling the time-points were highly comparable. Thus, these results suggested that in the *Bmal1* promoter region common to the reporter and endogenous locus as well as in a reporter-specific fragment, histones around *Bmal1* TSS are more acetylated during the peak expression phase in all three clones. However, no obvious difference was observed in the global histone acetylation enrichment between the three clones after the pooling of both time-points. Thus, just like burst frequency, histone acetylation states on the *Bmal1* promoter varied over the circadian time but remain strikingly similar between the three clones yet expressing luminescence at different levels.
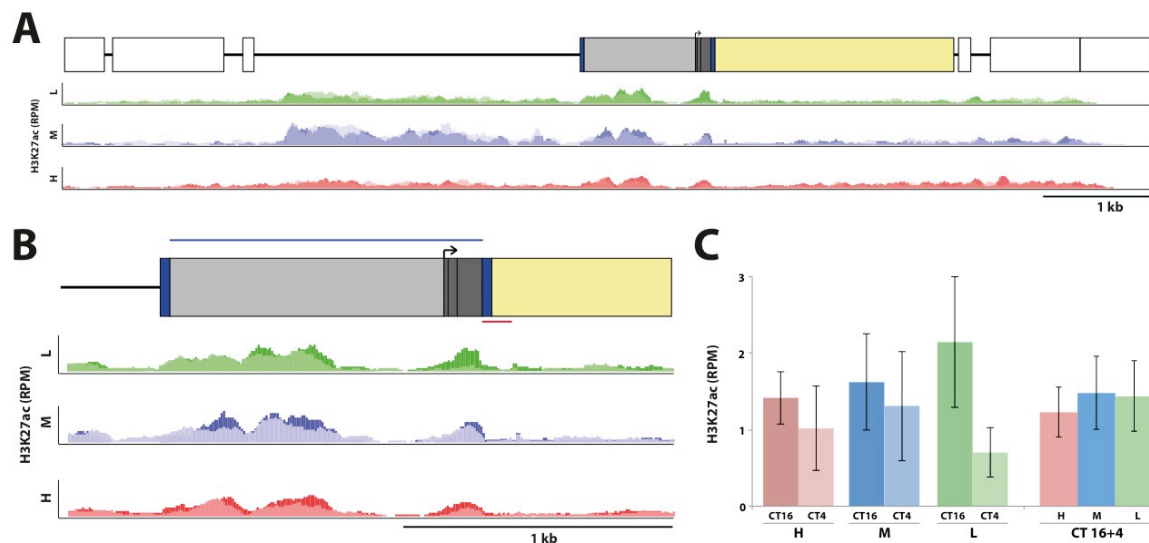
**Figure 2.39 H3K27 acetylation levels of the *Bmal1-sLuc2* reporter**

ChIP-seq peaks corresponding to H3K27ac signal mapped onto a 10kb region of the integrated *Bmal1-sLuc2* reporter. Samples were harvested at two time-points corresponding to peak and trough of *Bmal1* expression (CT16 and CT4, in dark and light colors respectively). The three clones are displayed in red (clone H), blue (clone M) and green (clone L). (**A**) Representative signal for each condition in the full 10kb region. The upper part schematizes the elements of the reporter region, with maintenance genes (such as resistance genes, white boxes), the *attB* reporter-specific cloning sites (dark blue boxes), *Bmal1* promoter region (grey box) and first intron (dark grey box, with the two RORE indicated as black bars) preceding the luciferase CDS (yellow). ChIP-seq signal is displayed in reads per million mapped reads onto the entire mouse genome (RPM), with the left vertical bar corresponding to 20 RPM. (**B**) Enlargement of a ~2kb region centered on the *Bmal1* promoter. The blue bar on top of the schematized promoter corresponds to the unspecific region present both at the endogenous *Bmal1* and on at *Bmal1-sLuc2* reporter. The red bar bellow corresponds to the 100 bp region used to quantify H3K27ac enrichment in the replicates of each experimental conditions (2 time-points, 3 clones) displayed in (**C**). Left panel represents the signal variation quantified between CT16 and CT4 in the H, M and L clones. Right panel corresponds to the pooled time-points signal in each clone. The bars represent the mean ± SD over 2 replicates.

## 2.5.5 Summary

In this section, we looked for molecular events at the *Bmal1-sLuc2* reporter co-occurring with changes in burst frequency, and therefore varying along the circadian period while remaining similar in the three H, M and L clones. The search was thus oriented towards known molecular events fluctuating at *Bmal1* promoter level along the circadian cycle.

Since tuning the transcriptional activity of the RORs and REV-ERBs circadian regulators of *Bmal1* could not be achieved using chemical ligands (**Figure 3.34**), an mRORE mutant version of *Bmal1* promoter region was generated to prevent their binding onto DNA. The mutations, in addition to maintaining *Bmal1* expression levels constitutively elevated (**Figure 3.36**), prevented the circadian cycle-dependent variations in burst frequency (**Figure 3.37**).

The link between the nuclear receptors (in particular REV-ERBs repressors) binding to the promoter and the variations in burst frequencies was further investigated by altering the histone acetylation state, since REV-ERBs mode of action results in the de-acetylation of the *Bmal1* promoter upon HDAC3 recruitment. Since inhibition of histone deacetylases using TSA resulted in

a similar phenotype than the mRORE mutant (**Figure 3.38**), the histone acetylation state of the *Bmal1-sLuc2* promoter at two circadian time-points and in three clones was assessed (**Figure 3.39**). H3K27ac signal at the integrated *Bmal1-sLuc2* reporter locus were more variable between the peak and trough expression phases than between the clones. Interestingly, this phenotype corresponded well to that of *Bmal1* burst frequency. Thus, histone H3 acetylation state was considered to be a very plausible candidate to co-vary with the promoter burst frequency. In fact, H3K27ac could even play a direct role in defining the frequencies of the burst. In the following sections, we will extend these observations by comparing histone acetylation state and burst frequency in other genes.

## 2.6     Transcriptional bursting behavior of endogenous circadian genes

Following the observation that both the burst frequency and the histone acetylation state of the *Bmal1-sLuc2* reporter synchronously oscillated over the circadian period, we thought to verify whether similar phenomena could be detected in other genes. Indeed, these observations done on the *Bmal1* reporter could reflect general features of gene expression, or be characteristic to this promoter.

To assess whether other rhythmically expressed genes modulate their burst frequency rather than burst size between the different circadian phases and if this phenomenon is correlated with variation in histone acetylation states, we focused on endogenous circadian genes. More specifically, we estimated their transcriptional bursting properties at the peak and trough of their rhythmic expression, and their histone acetylation state around the promoter.

To evaluate the extend of the phenomenon described with the *Bmal1* reporter, three genes representing broad ranges of circadian expression phase and transcriptional regulatory mechanisms were tested:

1) *Bmal1*: the endogenous *Bmal1* was selected to confirm that observations done with the *Bmal1-sLuc2* reporter also apply to the endogenous gene. A high degree of similarities between transcriptional properties of both the endogenous and the synthetic versions of *Bmal1* would confirm the biological relevance of the observations done using *Bmal1-sLuc2*, and confirm that the reporter could efficiently reproduce transcriptional behaviors of endogenous genes.

2) *Cry1*: expressed in most tissues ~6 hours before *Bmal1* (Panda et al. 2002; Zhang et al. 2014), the promoter of *Cry1* is composed of RORE but also complementary circadian regulatory motifs that dictate its expression phase (Ueda et al. 2005; Ukai-Tadenuma et al. 2011; Takeda et al. 2012). Thus, this gene was selected for partially sharing transcription regulation mechanisms with *Bmal1*.

3) *Dbp*: The expression of this gene is thought to be entirely regulated through E-boxes, the regulatory motif notably recruiting BMAL1 and CLOCK (Yamaguchi et al. 2000; Ueda et al. 2005; Stratmann et al. 2010). Consequently, its expression phase is anti-phasic to *Bmal1*. Additionally, despite their rhythmic circadian expression, promoter of *Bmal1* and *Dbp* do

not contain any common cis-regulatory motifs. Thus, *Dbp* was selected to assess the transcriptional bursting properties of circadian gene regulated in a RORE-independent manner.

### 2.6.1 smRNA-FISH on endogenous circadian genes

Comparison between real-time short-lived reporter measurements and smRNA-FISH highlighted the robustness of both strategies to infer the transcriptional bursting properties of promoters. However, unlike the short-lived luminescence reporter necessitating cloning steps and generation of new stable cell lines, the smRNA-FISH can straightforwardly be applied to any endogenous gene by designing probes specifically targeting their mRNA. Thus, this strategy was chosen to infer *Bmal1*, *Cry1* and *Dbp* burst size and frequency from mRNA distributions per cell at their expression peak and trough.

Since precise measurement of their rhythmic expression pattern is key to perform smRNA-FISH at time-points corresponding to the expression peak and trough, the circadian phase of these three genes was determined experimentally. NIH-3T3 mRNA was extracted every 2 hours for 24 hours, and after reverse transcription, expression levels of *Bmal1*, *Cry1* and *Dbp* were measured by qPCR. As expected, the mRNA accumulation of each transcript was displaying clear oscillations (**Figure 2.40A**). To more reliably identify the expression peak, pre-mRNA levels of these three genes were also assessed (**Figure 2.40B**). Consistent with the short half-lives of circadian genes, pre-mRNA accumulation peaked slightly earlier than the mature transcripts with an average phase advance of 2 hours (Kojima et al. 2012; Lück et al. 2014). From both mature mRNA and pre-mRNA accumulation data, peaks and troughs of circadian expression were estimated to correspond to CT 16 and 4 for *Bmal1*, CT 10 and 22 for *Cry1*, and CT 6 and 18 for *Dbp* respectively, confirming that this selection of circadian genes covered a broad range of expression phases (**Figure 2.40C and D**).

For the three endogenous genes, the estimated expression peaks and troughs were used as smRNA-FISH time-points. Comparison between mRNA per cell for the three circadian genes highlighted noticeable differences in the distributions shapes, notably between *Bmal1* displaying an average of 48.8 transcripts per cell and *Dbp* with 9.4 transcripts at their respective expression peaks (**Figure 2.41**). The average number of *Cry1* transcript per cell at its expression peak was 17.7.
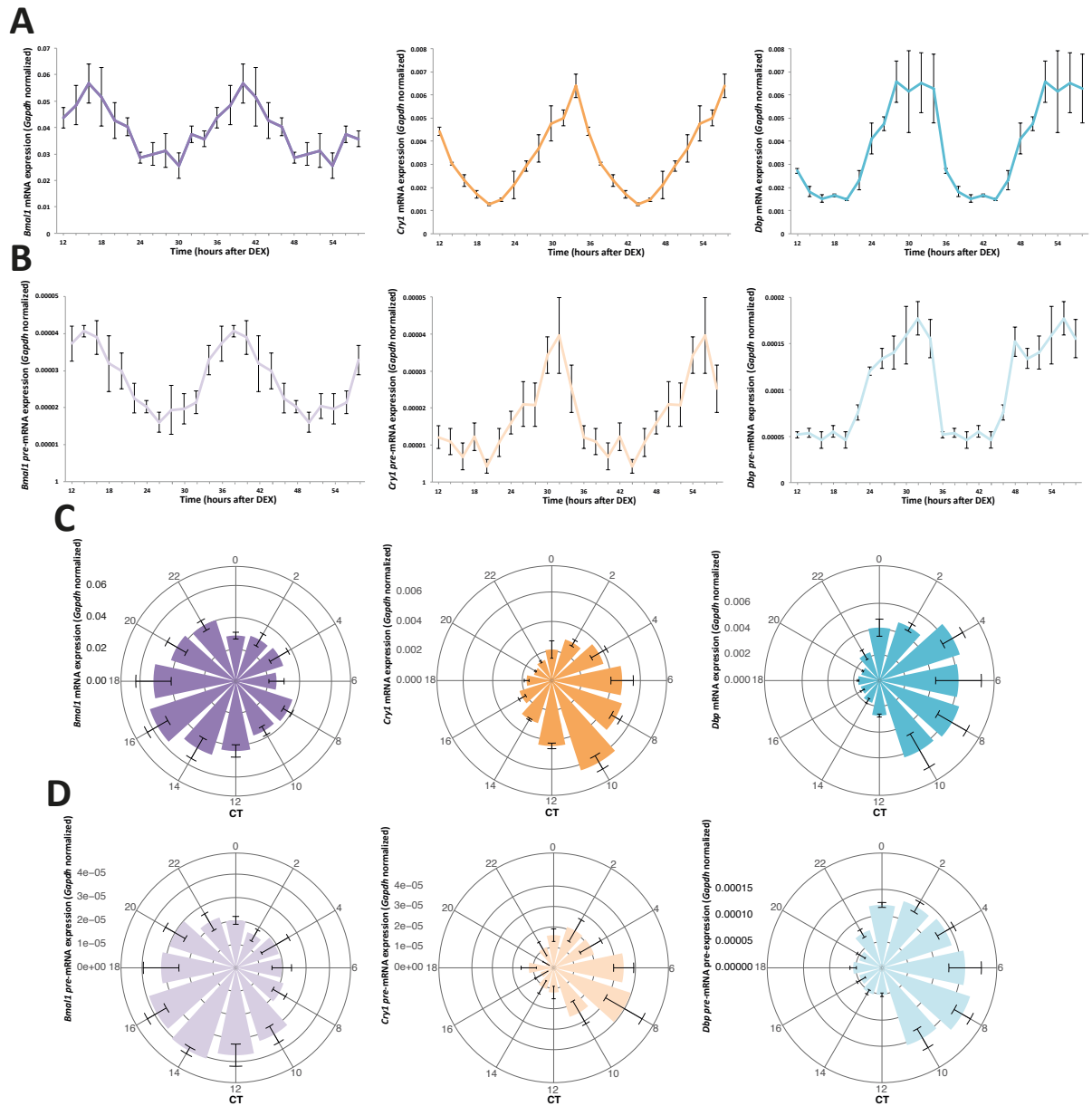
**Figure 2.40 mRNA accumulation of endogenous *Bmal1*, *Cry1* and *Dbp***

Rhythmic accumulation of *Bmal1* (purple), *Cry1* (orange) and *Dbp* (turquoise) transcripts in NIH-3T3 were measured by qPCR on reverse transcribed mRNA extracted every 2 hours between 8 and 30 hours after DEX synchronization. Transcripts values are normalized to Gapdh mRNA abundance (2^-ΔCt). Both mature mRNA (**A**) and pre-mRNA (**B**) were quantified. (**C-D**) Circular histograms representing the distribution phases of mature mRNA (**C**) and pre-mRNA (**D**) around the circadian period. Results displayed as mean ± SD over three replicates.
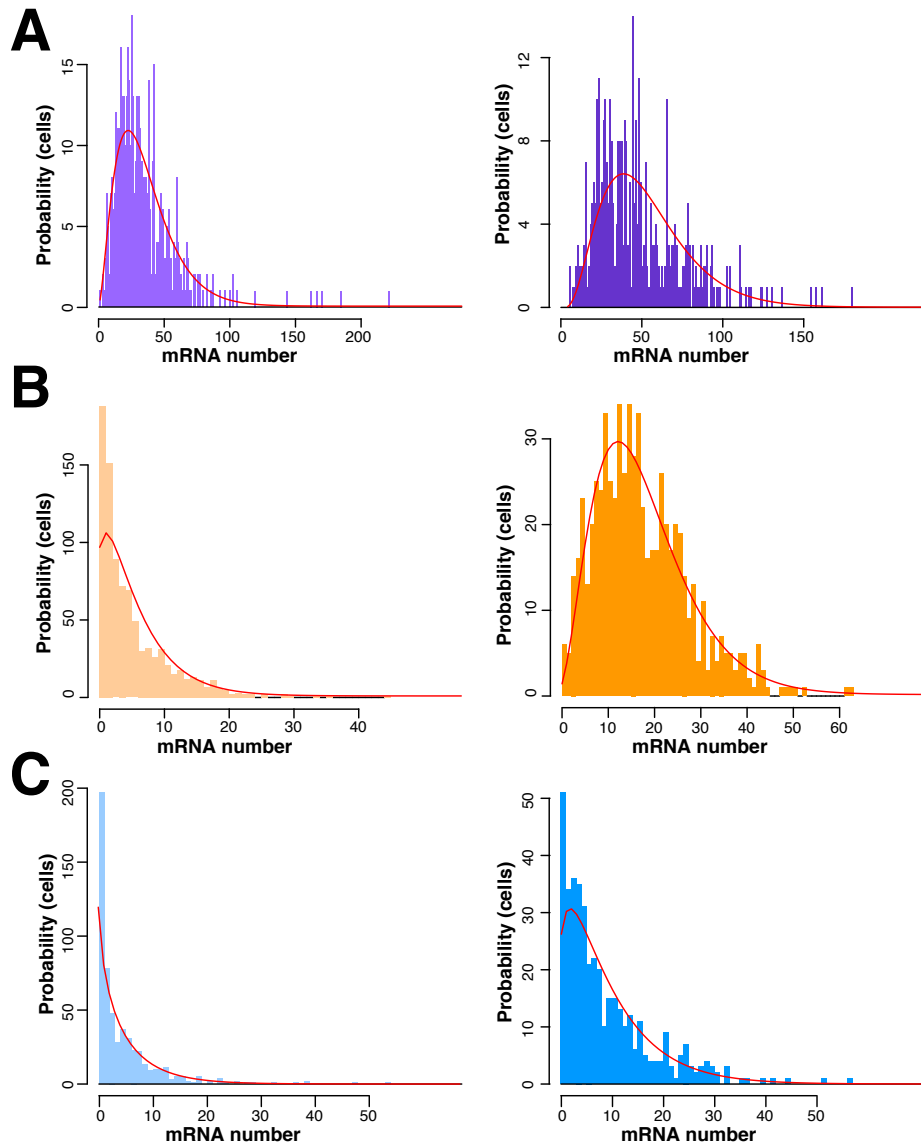
**Figure 2.41 smRNA-FISH distributions of endogenous circadian genes**

Distributions corresponding to the amount of transcript detected per cell (probability, corresponding to the absolute number of cells containing a given number of transcripts) using smRNA-FISH are shown for two circadian time-points corresponding to the expression trough (light colors) and peak (dark colors). The red line corresponds to the negative binomial fit used to infer bursting parameters. List of circadian genes probed in this figure, and their corresponding fixation trough and peak time-points: (**A**) *Bmal1* (purple, CT 4 and 16, n=490 and 378 respectively), (**B**) *Cry1* (orange, CT 22 and 10, n=672 and 874 respectively) and (**C**) *Dbp* (light blue, CT 18 and 6, n=491 and 411 respectively).

A telegraph model was used to infer the transcriptional busting parameters that best explained each of the observed smRNA-FISH distributions (**Equation 4.7**). Beside uncertainties in the fitted parameters, transcriptional regulation between expression peaks and troughs of *Bmal1*, *Cry1* and *Dbp* showed similarities, as for all genes the burst frequency was the only significantly varying parameter between the two time-points (**Figure 2.42**). Thus, despite large variations in their expression phase, their considerably dissimilar smRNA-FISH distributions and their different

promoter regulatory motifs, all three genes seemed to undergo similar modulations of their bursting parameters over the circadian phase. Although the phase-specific transcriptional bursting variations in the endogenous *Bmal1* gene were less pronounced than in the *Bmal1-sLuc2* version, they corresponded well with that of the reporter also estimated from smRNA-FISH. Thus, the *Bmal1* reporter likely trustfully recapitulates the transcriptional behavior of the endogenous gene.
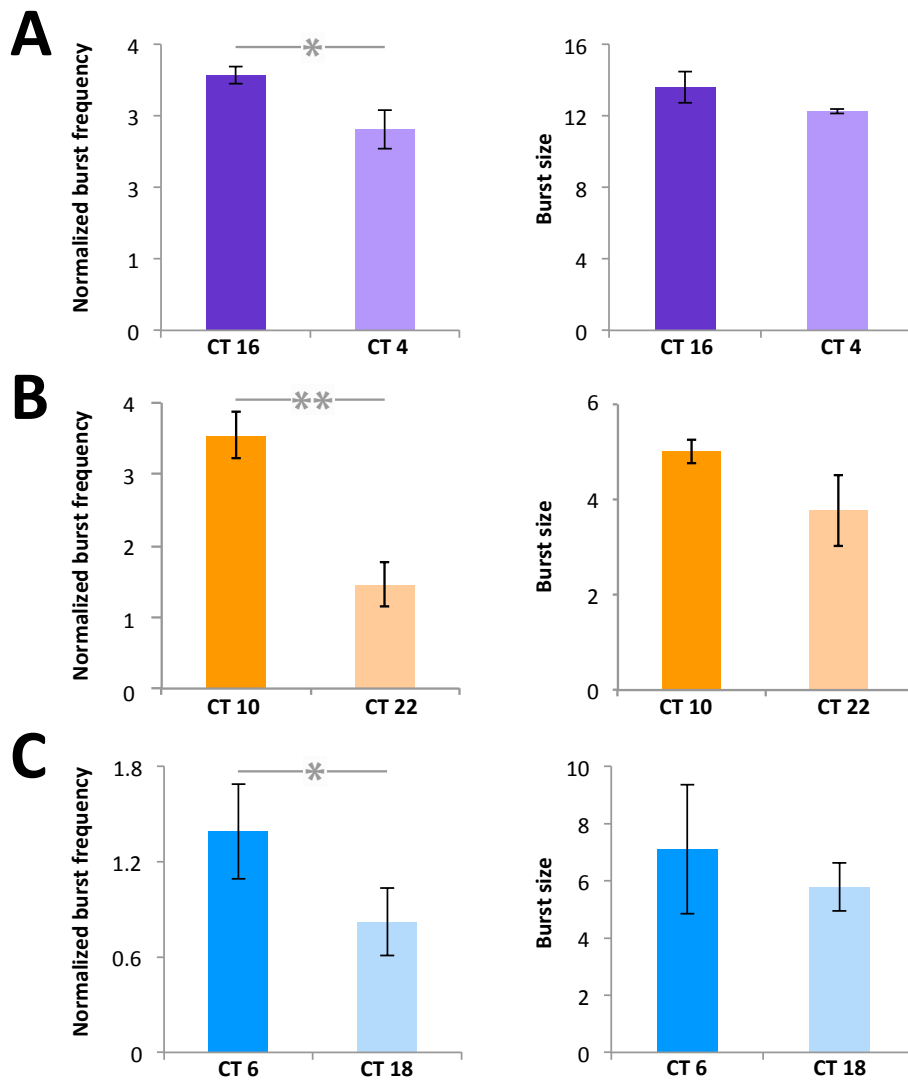


**Figure 2.42 Transcriptional bursting parameter of endogenous circadian genes**
Burst frequency (left, in transcript lifespan units of time) and burst size inferred from the smRNA-FISH distributions of transcript per cell corresponding to the peak (dark color) and trough (light color) of expression. Transcriptional bursting parameters were estimated for (**A**) *Bmal1* (purple), (**B**) *Cry1* (orange) and (**C**) *Dbp* (light blue). Bursting parameters were estimated from N>380 cells per condition. * = p-value <0.05, **= p-value <0.01.

## 2.6.2   Histone acetylation state of endogenous circadian genes

To confirm the link between histone acetylation state and burst frequency observed in *Bmal1-sLuc2*, the H3K27ac levels of endogenous *Bmal1*, *Cry1* and *Dbp* were also assessed. To this aim, we used the H3K27ac ChIP-seq datasets generated to compare the histone acetylation states at

*Bmal1-sLuc* promoter between its expression peak and trough. In addition to mapping the reads to the integrated reporter sequence, they were also mapped to the mouse reference genome to extract genome-wide information on H3K27ac enrichments. Fortunately, the time-points corresponding to the *Bmal1-sLuc2* reporter peak and trough of expression rationally corresponded to those of the endogenous *Bmal1*, but also to *Dbp*. Indeed, since this gene is antiphasic to *Bmal1*, the CT 16 roughly corresponds to its expression trough and CT 4 to its expression peak. However, these two time-points corresponded to *Cry1* intermediate expression level phase. Thus, limited information could be extracted from this dataset regarding the differential histone acetylation states of at *Cry1* promoter between its expression peak and trough.

Independently of the time-points, acetylation of histones was detected around the promoter region of all three circadian genes (**Figure 2.43**). However, H3K27ac spread differently in these genes. In *Bmal1* and *Cry1*, acetylation mainly accumulated in regions within 500 bp upstream of the promoter (in two distinct peaks for *Bmal1* and a single one for *Cry1*), as well as up to 1kb downstream of the TSS. The gene body, however, was essentially signal-depleted. Interestingly, *Dbp* acetylation pattern was strikingly different, with little or no accumulation observed within the 1kb window around the TSS. In contrast, histones located in the gene body were heavily acetylated. These differences in the spreading of H3K27ac signal along the genes may reflect the diversity regulatory mechanisms involved in their regulation.
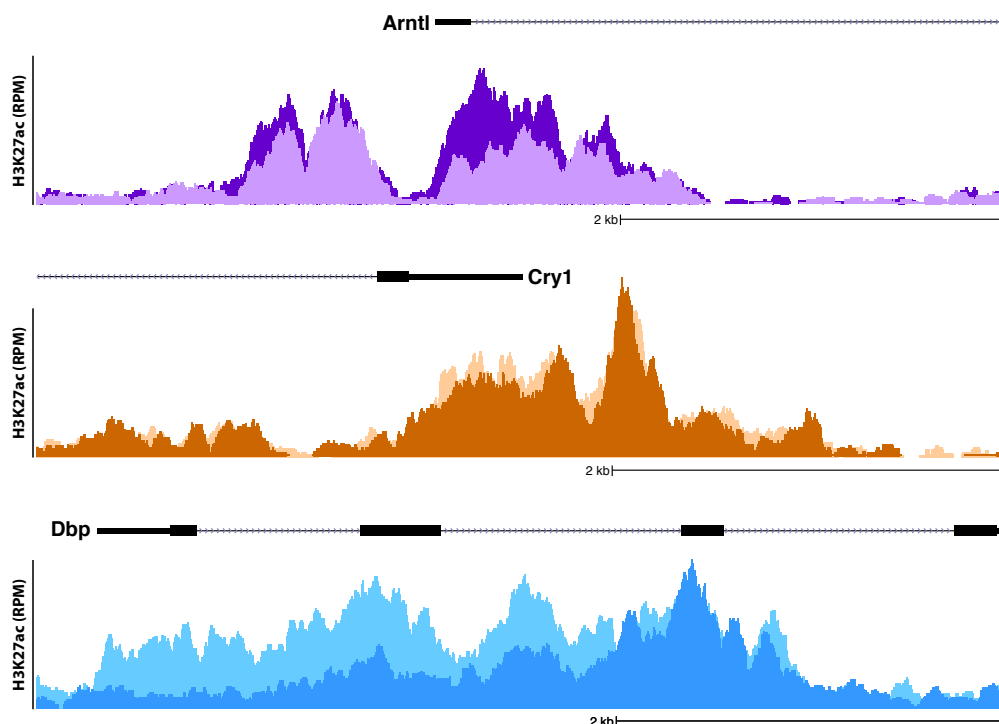


**Figure 2.43 Histone acetylation state of endogenous circadian genes at CT 16 and 4**
H3K27ac ChIP-seq signal from cells corresponding to clone H and fixed at both CT 16 (dark colors) and 4 (light colors) after DEX synchronization. Regions displayed correspond to *Bmal1* promoter (*Arntl*, purple, top), *Cry1* promoter (orange, center) and *Dbp* promoter and gene body (light blue, bottom). Positions of the TSS and elements of the 5' region of the gene are indicated on top (black boxes). ChIP-seq signal is displayed in reads per million mapped reads (RPM), with the left vertical bar corresponding to 10 RPM.

Comparison between the two time-points highlighted obvious differences in acetylation states for both *Bmal1* and *Dbp*, the two genes whose expression peaks and troughs corresponded well to the CT 16 and 4 time-points. Although the two peaks of H3K27ac signal upstream of *Bmal1* TSS only marginally varied, the downstream one did by ~2 folds. Similarly, important variations of histone acetylation levels were observed in the gene body of *Dbp* and especially in its 5' region. As expected, no clear differences were observed in the histone acetylation state of *Cry1* between the two time-points. However, this observation can likely be explained by the sub-optimal selection of cell fixation time-points for the ChIP-Seq that did not correspond to *Cry1* expression peak and trough, since levels of histone acetylation at *Cry1* promoter can vary of up to 4-fold between its expression peak and trough in other tissues (Etchegaray et al. 2003; Koike et al. 2012; Vollmers et al. 2012; Fang et al. 2014).

Thus, altogether, these data indicate that for both *Bmal1* and *Dbp*, expression modulation between the peak and trough time-points is mainly mediated by variations in their burst frequency and histone acetylation state around their promoter similarly varies over the same period. In both cases, the expression peak displaying the highest burst frequency corresponded to the most H3K27ac-enriched phase. If for *Cry1*, burst frequency also significantly varies around the circadian period, the dynamics of histone acetylation at its promoter remains untested in NIH-3T3 cells.

### 2.6.3   Summary

Since the observations that burst frequency rather than burst size varied along the daily period to modulate the circadian expression levels and that changes in burst frequency were correlated with histone acetylation states only relied on the observation of the *Bmal1-sLuc2* reporter, we thought to expand it to other genes. Since smRNA-FISH permits the estimation of transcriptional bursting parameters for endogenous circadian genes (Senecal et al. 2014; Singer et al. 2014; Ochiai et al. 2014; Bahar Halpern et al. 2015a), we tested their modulation between expression peak and trough for the endogenous *Bmal1*, *Cry1* and *Dbp*. Although *Dbp* and *Bmal1* differed in many aspects such as their expression phase and promoter regulatory motifs, both displayed burst frequency variations between the two time-points (**Figure 2.42**). Similarly, although the genomic repartition of the acetylated histones considerably differed between the two genes with *Bmal1* displaying H3K27ac enrichment around the TSS and *Dbp* within the gene body, the global histone acetylation state of both genes was highly rhythmic (**Figure 2.43**).

Concerning *Cry1*, it also mainly modulated its burst frequency between the circadian expression peak and trough. Unfortunately, this variation in bursting properties could not be compared to the histone acetylation state of the gene since H3K27ac ChIP-seq datasets at CT 16 and 4 corresponded to *Cry1 i*ntermediate expression level phases (**Figure 2.40**).


## 2.7   Broad correlation between burst frequency and acetylation state

Burst frequency variations appeared to be a common way to modulate expression levels between peak and trough in circadian genes. Indeed, *Bmal1*-sLuc2 as well as the endogenous *Bmal1*, Cry1

and *Dbp* used this regulatory mechanism despite fundamentally different promoter structures and for *Bmal1* and *Dbp* antiphasic expression phases. Another common feature between these genes is the variations of histone acetylation levels observed at their loci between the circadian time-points. Whether this correlation between burst frequency and histone acetylation levels is a broad phenomenon common to many genes, a specific regulatory phenomenon only observed in a subset genes or a biologically irrelevant correlation remained to be determined. Thus, we thought to assess whether such correlation could be observed in other, not necessarily circadian, genes.

A major limiting factor in assessing the correlation between transcriptional bursting features and the histone acetylation levels is the availability of a large panel of genes with precisely measured transcriptional bursting parameters. Indeed, estimating of transcriptional bursting parameters require laborious experimental approaches and a large majority of previous studies are focusing on one gene or few genes of interest (Lionnet and Singer 2012; Yao 2017). However, in this section, we use two datasets suitable to correlate burst frequency with histone acetylation in larger datasets.

### 2.7.1   Correlation between burst frequency and histone acetylation in a limited number of genes with precisely measured transcriptional bursting parameter

A previous study from our group characterized the transcriptional bursting characteristics of eight endogenous or circadian mouse genes in the NIH-3T3 cell line (Suter et al. 2011a). Thus, we thought of using these genes to compare specific transcriptional bursting properties with histone acetylation states around their promoter regions. The H3K27ac signal at the locus of these genes was directly quantified from the ChIP-seq datasets previously generated to evaluate the histone acetylation state at the *Bmal1-sLuc2* reporter level. A 500 bp region around the TSS was selected to quantity the acetylation state of the genes (**Figure 2.44**).

H3K27ac enrichment at the TSS was then compared to the transcriptional kinetics inferred in (Suter et al. 2011a). The burst size was calculated as the product of the time spent in the active state and the transcription rate ($t_{ON}*km$), and the inverse of the time spent in the transcriptionally inactive state was used as a proxy to estimate the burst frequency ($1/T_{off}$).

A correlation trend was visible between the burst frequency and the H3K27ac levels at the TSS (**Figure 2.45**). However, the reduced number of available genes and the rather small heterogeneity in burst frequency values did not lead to a significant relationship. Nevertheless, the correlation between burst frequency and histone acetylation was more pronounced than the one between burst size and H3K27ac levels. Indeed, although burst sizes were more broadly distributed between the genes, they did not correlate with the promoter acetylation state. *NcKap1* was not included in this burst size correlation because of the difficulty to estimate the time this gene spends in the transcriptionally active state (Suter et al. 2011a).

Thus, in this collection of genes, a modest non-significant correlation seemed to exist between the burst frequency and the histone acetylation levels close to the TSS. However, the correlation is nonexistent between the same histone mark and the burst size. Although transcriptional bursting parameters of these genes had been precisely inferred, the limited size of the selection complicated the interpretation of the analysis. Notably, the range of burst frequencies was limited

between the genes, with 5 out of 7 genes bursting on average between 1.4 and 2 times per hour, and only two genes with displaying more frequent burst.
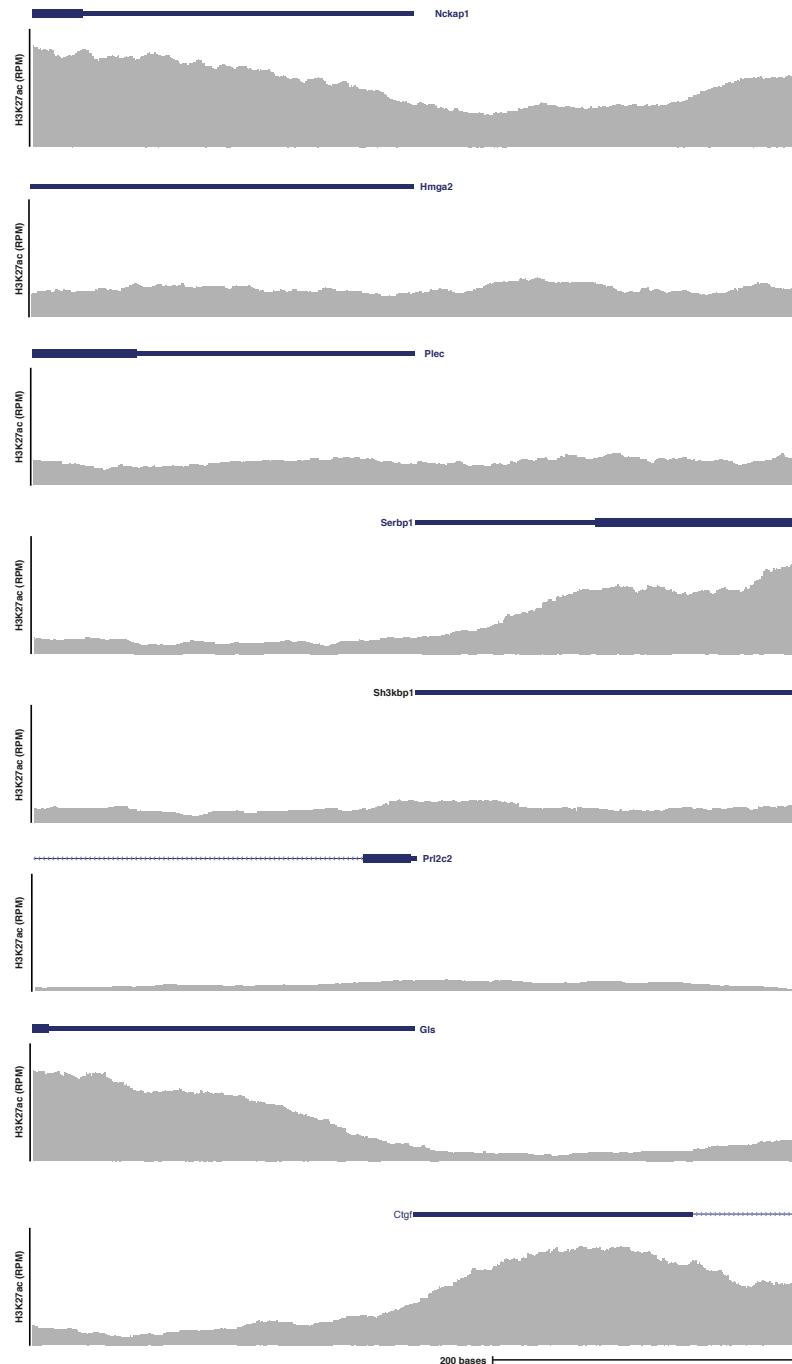


**Figure 2.44 Histone acetylation levels in the TSS region of genes monitored in (Suter et al. 2011a)**
H3K27ac ChIP-seq signal from cells corresponding to clone H and fixed at CT 16. Each row corresponds to a 500bp region around the TSS of genes monitored in (Suter et al. 2011a). Genes are sorted according to their burst frequency (top to bottom). ChIP-seq signal is displayed in reads per million mapped reads (RPM), with the left vertical bar corresponding to 10 RPM.
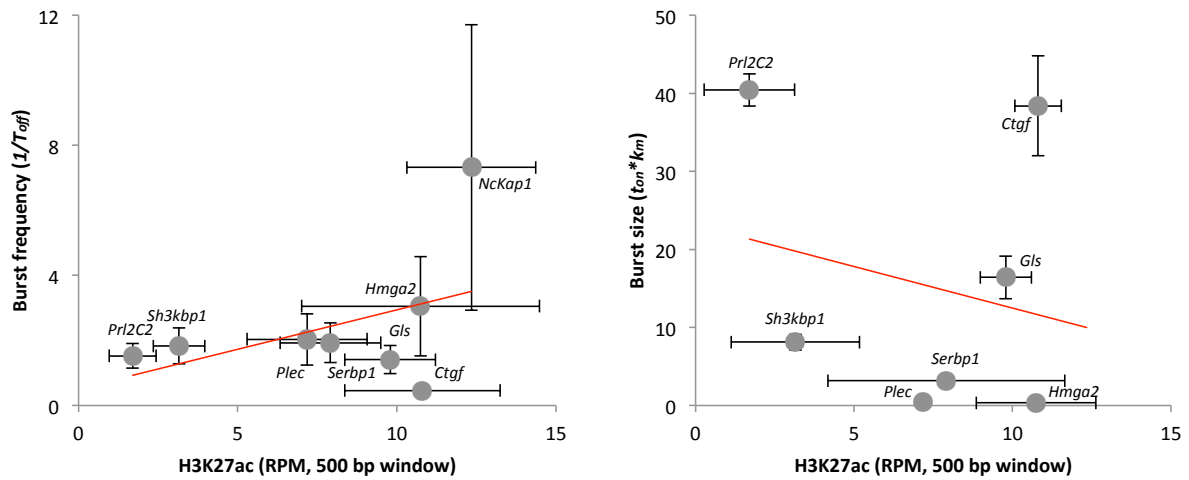
**Figure 2.45 Correlation between histone acetylation levels and transcriptional bursting parameters for genes monitored in (Suter et al. 2011a)**

Correlation between burst frequency ($1/T_{off}$, left panel) or burst size ($t_{on*}km$, right panel) and the H3K27ac ChIP-seq signal from cells corresponding to clone H and fixed at CT 16. Each row corresponds to a 500kb region around the TSS of genes monitored. Genes are sorted according to their burst frequency (top to bottom). The coefficients of correlation $R^2$ correspond to 0.19 and 0.05 for the burst frequency and burst size respectively (linear regression is shown as red line). X-axis values represent the mean ± SD over 12 ChIP-seq replicates (duplicates of 3 clones at two time-points), and the Y-axis the mean ± the 5th and 95th percentiles of the estimate over all MCMC iterations. p-values of the F-statistics correspond to 0.28 and 0.63 for the burst frequency and burst size respectively.

Consequently, to further investigate on the putative existence of a correlation between the acetylation state of a promoter and its burst frequency, we tested a slightly different approach on a larger dataset.

### 2.7.2 Correlation between bursting parameters and histone marks in a larger number of genes using smRNA-FISH distributions

Although data comprising the precise measurement of transcriptional bursting characteristics for a large subset of genes is not yet available, several groups were able to count the number of various mRNA in single-cells using diverse experimental approach such as smRNA-FISH or single-cell RNA-seq (Vera et al. 2016). Since the resulting distributions of transcripts per cell is sufficient to estimate the transcriptional bursting characteristics of the measured genes, we thought to combine this information with ChIP-seq data to identify correlations between presence of histone marks and bursting features. In contrast to smRNA-FISH that requires specific probes for each gene assessed, RNA-seq directly provides information on the entire transcriptome. Unfortunately, the low recovery rate of RNA molecules in single-cell RNA-seq experiments causes non-biological sampling noise, which can mask expression noise (Stegle et al. 2015; Zoller et al. 2015). Thus, we exclusively oriented our research on smRNA-FISH datasets, for which burst size and frequency estimation using distributions of RNA per cell is more reliable. Concretely, we were aiming at finding a cell type for which the following types of data were publically available:

1) smRNA-FISH data comprising  a large subset of genes and in which the burst size and frequency could be estimated from transcripts per cell distributions.
2) mRNA half-lives datasets to calculate the proper burst size and frequency instead of a normalized version on the transcript lifespan.
3) ChIP-seq data of histone marks (notably histone acetylation) and conceivably additional factors to be then correlated with specific bursting characteristics.

HeLa cells fulfilled these requirements. Indeed, in this cell type, over 900 mRNA were targeted using smRNA-FISH approaches (Battich et al. 2013), the half-live of 11,000 genes had been measured (Tani et al. 2012), and various ChIP-seq experiments assessing various molecular factors involved in transcription regulation were performed (Dunham et al. 2012; Berguet et al. 2014). However, this cell type comprises risks linked to its abnormal karyotype and its numerous genomic mutations (Landry et al. 2013). Indeed, in addition to the important chances of suffering differences between the HeLa strains used in these independent experiments, the unpredictable number of alleles present for each gene compromises the accurate estimation of bursting parameters. Thus, we perform our analysis on mouse ESCs. 38 genes were probed using smRNA-FISH in such cell line (Singer et al. 2014), and the lifespan of the corresponding transcripts could be extracted from databases massively determining transcripts half-lives (Sharova et al. 2009). Additionally, genome-wide distribution of the signal corresponding to multiple histone tail post-translational modifications had also been assessed (Dunham et al. 2012).

From the 38 available genes probed using smRNA-FISH, the burst frequency and burst size were calculated from the mean and variance of the distributions (**Table 2.3, Equation 4.8**). The burst frequencies were then corrected by taking into account the corresponding transcript half-life, resulting in burst frequencies in standardized time-scales comparable between all genes (Nicolas et al. 2017). In addition to these two transcriptional bursting parameters, the mean number of transcript per cell (mean expression, equivalent to the product of the burst size and frequency) was also included to discriminate between molecular markers globally affecting transcription, and those specifically acting on particular aspects of bursting. These three parameters of transcription (mean expression, bust frequency and burst size) were then compared to the signal of various molecular marker and notably histone marks present in a 500 bp to 500 kb region around the TSS. The resulting heat map indicated that most assessed histone marks positively correlated with the mean expression (**Figure 2.46A**). Noticeable exceptions included the repressive marks H3K9me3 and, more surprisingly, the H3K36me3 mark often found in the gene body of transcribed genes. Similar correlations were observed with the burst frequency. However, for some active transcription marks in a 5 kb window around the TSS such as H3K4me3, H3K9ac and especially H3K27ac, the correlations were better for burst frequency than for mean expression. This suggests that the associations between mean expression levels and these specific histone marks is largely driven by the burst frequency. However, at larger windows around the TSS correlations between histone marks signal and burst frequency diminished. Concerning the burst size, only weak negative correlations were observed. This implies that the burst size is only marginally associated with the presence of any of the selected marks. Among the best anti-correlations, H3K4me1 and

H3K36me3 slightly stood off. This anti-correlation is surprising as both marks are mainly associated with active transcription (Barski et al. 2007; Wagner and Carpenter 2012). Interestingly, H3K9me3 levels also vaguely anti-correlated with the bust size, notably in large windows around the TSS. Thus, the burst size could be partially sensitive to the global chromatin state of the locus.

| Gene name | Mean mRNA/cell | Variance | mRNA half-life | Relative Burst Freq | Burst Freq | Burst size | TSS position (mm9) |
|---|---|---|---|---|---|---|---|
| Dppa4 | 57.91 | 1214.76 | 6.30 | 2.76 | 0.44 | 132.05 | chr16:48283848 |
| Carm1 | 24.68 | 102.94 | 8.35 | 5.91 | 0.71 | 34.84 | chr9:21351338 |
| Cnot3 | 25.24 | 156.31 | 14.97 | 4.08 | 0.27 | 92.71 | chr7:3596871 |
| gp130 | 11.16 | 52.54 | 5.81 | 2.37 | 0.41 | 27.37 | chr13:113254278 |
| Nacc1 | 41.38 | 427.13 | 23.66 | 4.01 | 0.17 | 244.20 | chr8:87194378 |
| Zfp281 | 16.40 | 101.27 | 3.32 | 2.66 | 0.80 | 20.49 | chr1:138526968 |
| Jmjd2c | 11.08 | 38.24 | 4.13 | 3.21 | 0.78 | 14.25 | chr4:74051768 |
| Oct4 | 120.77 | 3686.76 | 4.58 | 3.96 | 0.86 | 139.85 | chr17:35645752 |
| Prmt5 | 29.23 | 206.44 | 7.04 | 4.14 | 0.59 | 49.71 | chr14:55126019 |
| Sall4 | 52.37 | 611.37 | 5.97 | 4.49 | 0.75 | 69.63 | chr2:168573832 |
| Sdha | 34.97 | 477.25 | 8.21 | 2.56 | 0.31 | 112.07 | chr13:74459703 |
| Smarcc1 | 43.56 | 556.49 | 3.64 | 3.41 | 0.94 | 46.48 | chr9:110034528 |
| Tbp | 17.28 | 60.05 | 2.97 | 4.97 | 1.68 | 10.32 | chr17:15636852 |
| Tcf3 | 24.66 | 292.34 | 5.49 | 2.08 | 0.38 | 65.11 | chr10:79871910 |
| Stat3 | 16.53 | 93.18 | 8.13 | 2.93 | 0.36 | 45.81 | chr11:100748120 |
| Rest | 74.93 | 1203.38 | 4.75 | 4.67 | 0.98 | 76.24 | chr5:77694519 |
| Trim28 | 161.91 | 5617.17 | 6.04 | 4.67 | 0.77 | 209.66 | chr7:13609501 |
| Zic3 | 38.59 | 557.89 | 2.36 | 2.67 | 1.13 | 34.09 | chrX:55283805 |
| Pecam | 7.43 | 53.61 | 19.58 | 1.03 | 0.05 | 141.36 | chr11:106515527 |
| Blimp | 5.84 | 27.78 | 0.98 | 1.23 | 1.25 | 4.68 | chr10:44156980 |
| Nr0b1 | 31.66 | 1225.93 | 8.19 | 0.82 | 0.10 | 317.22 | chrX:83437114 |
| Dnmt3b | 69.52 | 5673.00 | 4.19 | 0.85 | 0.20 | 341.91 | chr2:153475185 |
| Fgfr2 | 6.09 | 36.73 | 3.15 | 1.01 | 0.32 | 18.99 | chr7:137305965 |
| Klf4 | 37.20 | 1165.22 | 1.82 | 1.19 | 0.65 | 57.14 | chr4:55540009 |
| Prdm14 | 4.53 | 38.14 | 3.09 | 0.54 | 0.17 | 25.99 | chr1:13103509 |
| Dppa3 | 5.47 | 92.01 | 24.00 | 0.33 | 0.01 | 403.76 | chr6:122576442 |
| Tbx3 | 4.16 | 64.96 | 4.99 | 0.27 | 0.05 | 77.99 | chr5:120120678 |
| Zscan4c | 0.61 | 21.03 | 13.84 | 0.02 | 0.00 | 480.90 | chr7:11591094 |
| T | 0.30 | 1.13 | 1.54 | 0.08 | 0.05 | 5.79 | chr17:8627288 |
| Socs3 | 11.42 | 94.71 | 1.24 | 1.38 | 1.11 | 10.30 | chr11:117827401 |
| Rex1 | 138.37 | 7176.05 | 2.66 | 2.67 | 1.00 | 137.96 | chr8:44380421 |
| Fgf4 | 26.62 | 354.16 | 1.81 | 2.00 | 1.10 | 24.12 | chr7:152047291 |
| Tcl1 | 27.64 | 523.01 | 11.79 | 1.46 | 0.12 | 223.02 | chr12:106454965 |
| Sox2 | 76.47 | 1919.28 | 1.29 | 3.05 | 2.37 | 32.30 | chr3:34548917 |
| Nanog | 74.46 | 3770.77 | 5.62 | 1.47 | 0.26 | 284.46 | chr6:122657507 |
| Lifr | 11.69 | 59.81 | 6.53 | 2.28 | 0.35 | 33.39 | chr15:7090542 |
| Tet1 | 87.97 | 1949.74 | 3.56 | 3.97 | 1.11 | 78.92 | chr10:62267318 |
| Esrrb | 67.52 | 2474.73 | 7.12 | 1.84 | 0.26 | 261.06 | chr12:87762594 |

**Table 2.3 Bursting parameters estimated from (Singer et al. 2014) smRNA-FISH distributions**
smRNA-FISH distributions of 38 mESC genes obtained from (Singer et al. 2014) were used to infer their corresponding transcriptional bursting parameters. Using the mean number of mRNA per cell ($\mu$) and the variance ($\sigma^2$), the realtive burst frequency could be estimated ($f_R=1/CV^2$, in mRNA half-live$^{-1}$). MC1 LIF+ condition of mRNA half-lives databases (Sharova et al. 2009) in hours were used to measure the absolute burst frequency ($f_A= f_R/half\text{-}live$, in hour$^{-1}$). Burst size was measured as $b=\mu/f$ (in absolute mRNA number). The position of the TSS on the mm9 genome versions also indicated.
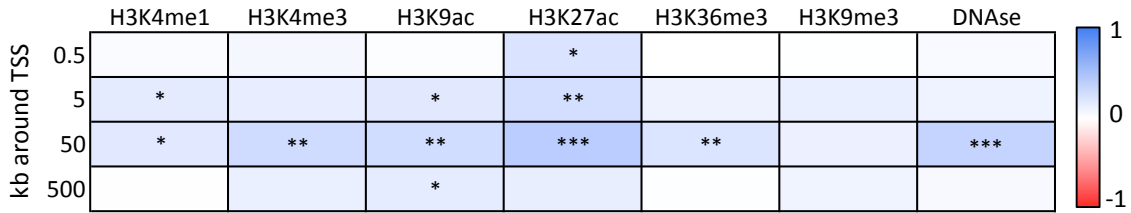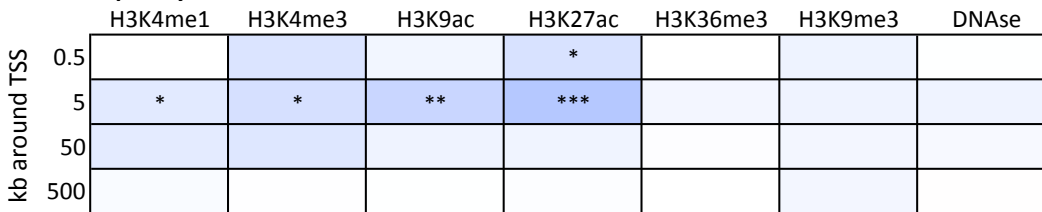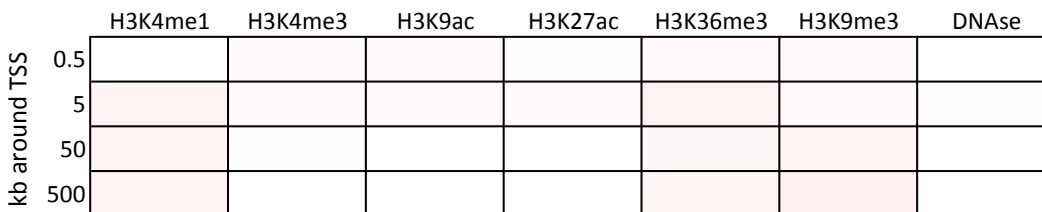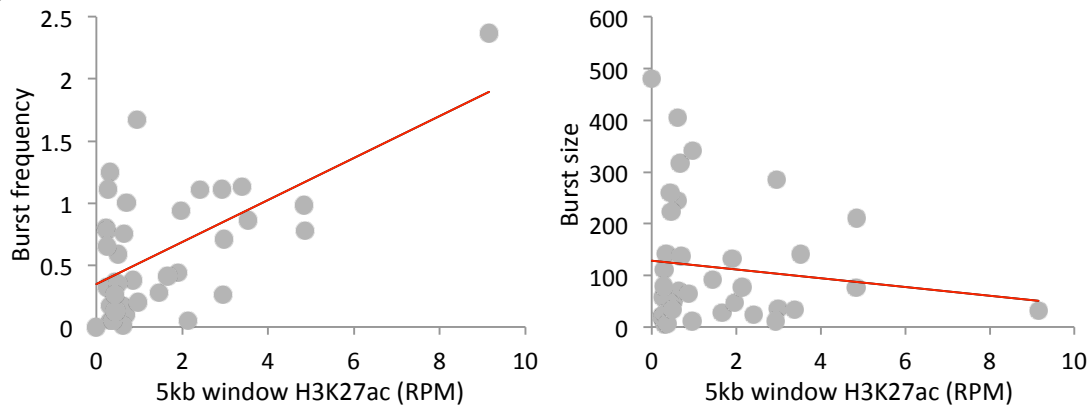
**Figure 2.46 Correlations between histone marks levels and transcriptional bursting parameters using smRNA-FISH distributions**

Correlation between mean expression, burst frequency or burst size and various histone marks as well as DNase I signal in different windows around the TSS. (**A**) Heat map summarizing the result of each comparison. The coefficients of correlation between each levels of each molecular markers in 4 windows around the TSS (0.5, 5, 50 and 500 kb) and the mean expression (top panel), the burst frequency (middle panel) and the burst size (bottom panel) are shown in blue (positive correlation) or red (negative correlation). The color scale is shown in the top right corner. F-statistics: * = p-value <0.05, **= p-value <0.01, ***= p-value <0.001. (**B**) Visualization of correlation examples between the burst frequency (left panel) or burst size (right panel) and H3K27ac levels in a 5 kb region around TSS. The coefficients of correlation $R^2$ correspond to 0.364 and 0.016 for the burst frequency and burst size respectively (linear regression is shown as red line).

Since H3K27ac was the best molecular candidate to come along burst frequency variations in the *Bmal1* reporter, we further focused on this specific histone marks and its possible correlations with various aspects of bursting. When assessed in a reduced region around the TSS, H3K27ac showed the best-observed correlation with the burst frequency. Although mean expression was also associated with this specific marker, the correlation is stronger when restricted to burst frequency only. Indeed, beside few genes displaying both elevated burst frequency and moderate H3K27ac levels, in a majority of genes these two transcriptional aspects were significantly linked ($R^2$ = 0.364, p-value: 6.082e-05) (**Figure 3.46B**). The correlation remained even in absence of the *Sox2* gene displaying both high burst frequency and elevated H3K27ac signal ($R^2$ = 0.13, p-value: 0.03552). In contrast, no correlation was observed between that mark and the burst size ($R^2$ = 0.016, p-value: 0.6916), implying that expression modulation through burst size variations is achieved by markedly different molecular mechanisms.

Finally, since none of the selected molecular marker clearly correlated with burst sizes, we investigated on the role of an alternative biological phenomenon. Some theories propose that burst size arises from PolII pausing during transcription elongation (Dobrzynski and Bruggeman 2009; Kim and Marioni 2013). Indeed, pausing is likely to influence the bursting properties of genes, notably by defining the amount of PolII that can be loaded onto the gene before releasing the transcription. Thus, bursting parameters of the same 38 genes were compared to their PolII proximal pausing properties. Because no ChIP-seq datasets of well characterized pausing factors or specific PolII CTD variants could be found for mESC, the extend of the pausing phenomenon for each gene was estimated using the ratio between PolII signal in the 200bp region downstream of the TSS (paused PolII) and the gene body signal (elongating PolII) (Adelman and Lis 2012). PolII signal along the gene body positively correlated with the burst frequency and to a lesser extend to the mean expression level. Surprisingly, no correlation was observed between elongating PolII signal and the burst size, suggesting that PolII ChIP-signal poorly captured this aspect of bursting (**Figure 2.47**). Focusing on the paused/elongating PolII ratio did not highlight any link between PolII proximal pausing and transcriptional bursting features. Thus, the prevalence of PolII pausing in some genes did not seem to participate in determining neither the burst size nor the frequency. Alternatively, this lack of correlation could also indicate that the strategy used to estimate the occurrence of proximal pausing was not appropriate.

### 2.7.3 Summary

To verify whether the correlation between burst frequency and histone acetylation was a broad feature of gene expression, we performed cross-correlations between transcriptional bursting parameters and functional genomic datasets for different sets of genes. The first set consisted in 8 randomly selected genes for which the bursting parameters had been precisely measured (Suter et al. 2011a). Since bursting features of these genes were measured in the same cell line than the H3K27ac ChiP-seq generated for this study, these data were used to quantify the histone acetylation levels at the promoter regions. A non-significant trend was observed between H3K27ac enrichment around the TSS and the burst frequency, while the burst size was completely independent from the H3K27ac signal (**Figure 2.45**).

**Mean expression**

| Paused PolII | Elongating PolII | Paused/Elongating |
|---|---|---|
|  |  |  |

**Burst frequency**

| Paused PolII | Elongating PolII | Paused/Elongating |
|---|---|---|
|  | ** |  |

**Burst size**

| Paused PolII | Elongating PolII | Paused/Elongating |
|---|---|---|
|  |  |  |
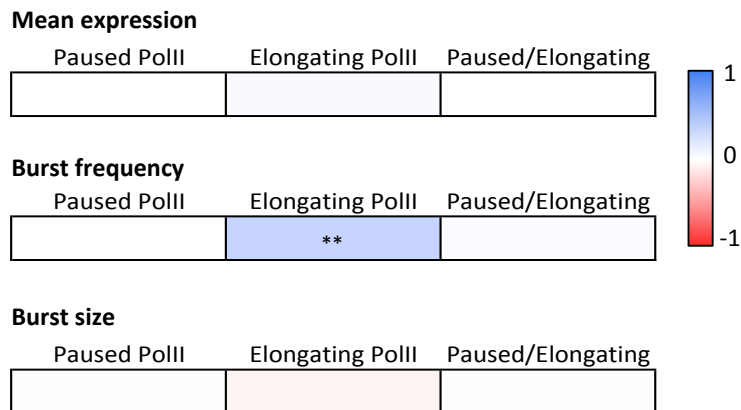
**Figure 2.47 Correlations between PolII signal and transcriptional bursting parameters**
Correlation between mean expression (top), burst frequency (center) or burst size (bottom) and PolII signal enriched in paused PolII (Paused PolII, 200bp window downstream of TSS), Elongating PolII signal (gene body signal) or the Paused/Elongating ratio. Coefficients of correlation are shown in blue (positive correlation) or red (negative correlation). F-statistics: * = p-value <0.05, **= p-value <0.01.

A similar approach was used on 38 genes whose mRNA per cell distributions were measured using an smRNA-FISH approach in mESC (Singer et al. 2014). These distributions were used to infer the burst size and frequency of these genes. The large amount of functional genomic datasets available for this cells-line not only permitted a comparison between bursting parameters and histone acetylation but also with other histone marks and DNaseI hypersensitive sites. Interestingly, the strongest correlations were observed between burst frequency and H3K27ac in regions close to the promoter (**Figure 2.46**). Thus, the link between the presence of acetylated histones around the promoter and the frequency of the transcriptional bursts was broadly observed in many unrelated genes, and did not only occur in circadian genes. On the other side, using the same dataset, no correlation could be found between the burst size and any selected molecular marker (**Figure 2.46**), or between PolII proximal pausing and specific aspects of bursting (**Figure 2.47**).

# Chapter 3    Discussion

In this chapter, I will recapitulate relevant results gathered during the study and comment on their importance. I will notably come back to the different technical approaches used to estimate the bursting parameters of *Bmal1* and other genes and how they participate in precisely determining their bursting parameters. I will also propose models to clarify the biological meaning of the correlations identified in this work. In particular, I will propose explanations for the role of histone acetylation in modulating the burst frequency. I will also mention specific aspects that could be improved in this work, as well as future directions that deserve a deeper look in the field of transcriptional bursting. Finally, I will conclude on the importance of characterizing the bursting signature of genes in various contexts.

Some aspects of the discussion, notably extrapolations on the impact of molecular mechanisms on transcriptional bursting or comments on technical approaches commonly used to monitor transcriptional bursting are partially inspired from my review (Nicolas et al. 2017) that addressed similar themes.

## 3.1    Short-lived luciferase and smRNA-FISH approaches to study transcriptional bursting

In this section, I will come back on the tools developed for the study of *Bmal1* transcriptional bursting. Notably, I will mention the modular properties of the short-lived luciferase expression vector, its use during this study and how it could be further used in future experiments. I also compare the results obtained using real-time short-lived reporter monitoring and smRNA-FISH, the two different approaches used to measure transcriptional bursting. Finally, I also justify the use of a reporter vector to study the transcriptional bursting properties of a promoter endogenously present in the cells.

### 3.1.1    Development of a highly modular short-lived luciferase expression vector

To study the transcriptional bursting properties of the *Bmal1* promoter, a new version of short-lived luciferase expression vector was designed. This improved version is greatly inspired from a previous reporter (Suter et al. 2011a). Since the short-lived luciferase coding region remained unchanged, the mRNA and protein half-lives are equivalent to the ones previously measured (Suter et al. 2011a) (**Figure 2.5**). The upstream region of the expression vector was however adapted to the study of genes of interest. Indeed, the former version was mainly designed to monitor endogenous genes, or to integrate specific promoters using conventional restriction

enzymes approaches. By facilitating the integration of a cloned promoter of choice and permitting the single-copy integration of the reporter in the genome, we generated a modular system, whose multiple features could advantageously be used in this comparative study:

- The Gateway cloning cassette upstream of the short-lived CDS greatly facilitated the insertion of any selected promoter to drive the reporter expression. In addition to avoiding the laborious use of restriction enzymes notably while cloning a high number of promoters with distinct sequences, this cassette is compatible with existing Gateway promoter banks available for various species (Hope 2004; Marquès-Bueno et al. 2016). Since promoter regions can have different structures, we adapted the vector to be compatible with promoters either containing or lacking an endogenous START codon. The only requirements for successful protein expression are the presence of a TSS in the cloned promoter region, and that this region terminates in an exonic region. Although the expression capacities of the vector were verified with multiple promoters (**Figure 2.4**), the versatility of the cloning cassette was only partially exploited in this work, since it was mainly use to study the *Bmal1* promoter and a mutated mRORE version.
- The short-lived expression vector could be successfully used in transiently transfected cells (**Figure 2.4**). However, the presence of an FRT-cassette also allows the optional integration of the reporter into a single genomic site in FRT-compatible cells. The advantages of this strategy are multiple. First, in contrast to transient transfection, expression properties can be measured for prolonged periods without loosing the reporter through dilution or ejection. Second, it guaranties the presence of a single copy of the reporter per cell, which is often crucial to comprehend single-cell resolution data. Finally, it enables the integration of the reporter always at the same genomic region. Consequently, the expression properties of different promoters or promoter variants (such as transcription factor binding sites mutants) can be directly compared without suffering disparities arising from the integration site.
- In parallel to the short-lived luciferase expression vector, several FRT-compatible NIH-3T3 cell lines were generated. In addition to allowing the comparison of transcriptional bursting properties at the various sites, these cells can be used to obtain a suitable expression level. Indeed, for theoretically any FRT-compatible expression vector, controlled expression levels can be achieved by selecting the most appropriate cell line among the available transcription levels.

Thus, although all features were not fully exploited in this work, this version of the short-lived luciferase expression vector could easily be adapted to other experimental purposes, both at single-cell or population resolution. In addition to the study of transcriptional bursting, the reduced half-live of the reporter is also suitable to precisely measure variation of expression in various contexts such as the circadian cycle or responses to stimuli.

### 3.1.2 Combining two experimental approaches to monitor transcriptional bursting

Measuring transcriptional bursting properties of a promoter is challenging. Indeed, it requires precise quantification of gene products, ideally with a single-molecule resolution. Although several technical approaches have successfully been used, they all comprise advantages and drawbacks (Raj and van Oudenaarden 2009; Larson et al. 2009). Although most studies in the field of transcriptional bursting focus on a single technique, combining several monitoring systems in a single study allows overcoming major approach-specific drawbacks. Combined technical approaches were notably used to measure expression noise both at the mRNA and protein level (Dey et al. 2015; Dar et al. 2016), but also to obtain complementary method-specific information. In this way, real-time MS2-GFP measurements have already been combined with the quantitative properties of smRNA-FISH to precisely count the number of detected transcripts produced at transcription sites (Kalo et al. 2015).

Combining approaches is all the more pertinent that some genes display bursting at different time-scales, influenced either by the pausing of PolII convoys or long-term promoter active states (Tantale et al. 2016). In this context, focusing on the precise monitoring of transcription during short periods (typically with the MS2-GFP approach) or obtaining longer time-scale recordings yet less sensitive to signal fluctuations (typically protein reporters) could result in incomplete information.

To gain in measurements precision and to strengthen the confidence in observations, the transcriptional properties of the *Bmal1* promoter was assessed using two complementary approaches. Indeed, real-time monitoring using a short-lived reporter allows the dynamic monitoring of expression, but the readout corresponded to the protein product of the gene. Thus, transcription properties had to be indirectly inferred from the data using modeling approach that may not take into account the whole complexity of gene expression. On the other hand, smRNA-FISH allowed precise quantification of mRNA, the direct product of transcription. Unfortunately, the approach requires fixed cells, preventing dynamic measurement of transcription dynamics.

Fortunately, both technical approaches converged to essentially similar results. Indeed, in both cases, the measured or estimated amount of reporter transcripts with smRNA-FISH and short-lived reporter respectively were equivalent (**Figure 2.28**). Additionally, the circadian time and the integration site similarly affected the transcriptional bursting parameters with both approaches, although differences in absolute values were observed (**Figures 2.18 and 2.30**). Indeed, both approach highlighted that the burst frequency predominantly modulates the temporal expression variations of *Bmal1* over the circadian cycle, while the integration site mainly modulated the burst size (**Figure 3.1**). These findings confirmed that the burst frequency and burst size are uncoupled, and that cells can modulate both aspects separately to regulate expression levels (Suter et al. 2011a; Dar et al. 2012; Senecal et al. 2014; Dey et al. 2015).

Since both approaches resulted in similar conclusions, we considered that they both were sufficiently robust to be independently used to estimate bursting properties in complementary investigations. Thus, some experiments were only performed with a one type of approach particularity suitable for a specific condition. Indeed, real-time reporter was privileged to study

circadian rhythmicity (such as analysis on the mutated *Bmal1* mutated promoter (**Figure 2.37**)), and smRNA-FISH to assess the bursting properties of endogenous gene (**Figure 2.42**).
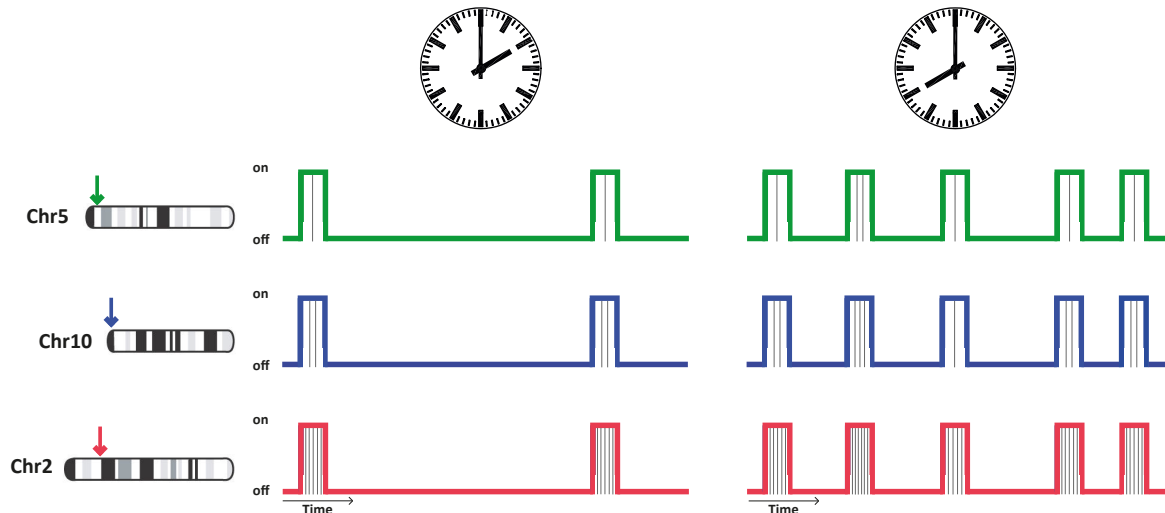


**Figure 3.1 Impact of the circadian phase and genomic position on *Bmal1-sLuc2* bursting properties**
Summary of the conditions tested in this work to modulate the transcriptional bursting parameters of the *Bmal1-sLuc2* reporter. The circadian time mainly modulates the frequency of the bursts, with the highest burst frequency corresponding to CT 16 and the lowest at CT 4. The reporter integration site (represented in different colors: clone L in green, clone M in blue and clone H in red) predominantly influences the size of the bursts (represented with grey bars during "on" states corresponding to the production of transcripts).

### 3.1.3    A reporter to study the transcriptional characteristics of an endogenous promoter

An essential objective of this project was to characterize the transcriptional bursting behavior of the *Bmal1* promoter, notably along the circadian cycle. Despite the possibility of tagging an endogenous allele with a short-lived reporter using, for example, a CRISPR-Cas9 genome editing approach (Sander and Joung 2014), we opted for a transgenic system. Indeed, the circadian molecular clock consists in a highly interconnected transcriptional translational feedback loop in which *Bmal1* has a crucial role (Takahashi 2016). In NIH-3T3 cells as well as in other systems, miss-regulation of *Bmal1* can have dramatic consequences on the proper functioning of the clock (Sasaki et al. 2009; Baggs et al. 2009; Ramanathan et al. 2014). Consequently, any form of editing on an endogenous *Bmal1* allele (notably mutations of regulatory elements) could have led to undesirable consequences on the robustness of the molecular clock and could bias the measured phenotype. Thus, although an integrated reporter may not precisely reflect the full complexity of regulatory mechanism taking place at the endogenous locus, it presented a less hazardous approach and permitted more drastic expression modulations without suffering any uncontrollable consequences on the circadian clock. Additionally, our reporter system allowed comparison between various integration sites, which proved to be a crucial aspect for the study of molecular mechanisms influencing the burst size.

Another possibility to measure the bursting properties of the endogenous *Bmal1* was to use smRNA-FISH on endogenous transcripts. This method is less invasive but would also limit the editing possibilities on endogenous *Bmal1* alleles. Also, this approach cannot differentiate

between the four endogenous *Bmal1* alleles (in tetraploid NIH-3T3 cells), which compromises the understanding of their individual behavior. Additionally, limited perturbations can be applied using this system. However, the approach was still used to compare the bursting characteristics of the *Bmal1-sLuc2* reporter with those of the endogenous gene. Satisfyingly, the two promoters shared many similarities in their transcriptional behavior. In addition to displaying similar mRNA circadian expression (expression peaks at CT16 for both the reporter and the endogenous *Bmal1* gene) and fold-change between expression peaks and troughs (2- and 2.5-folds for the reporter and endogenous *Bmal1* respectively) (**Figures 2.25 and 2.30**), both the reporter and the endogenous *Bmal1* genes showed a marked tendency to modulate the burst frequency rather than the size over the circadian phase (**Figures 2.30 and 2.42**). Additionally, the three integrations sites used in this study seem to exemplify well the expression levels of the endogenous *Bmal1* gene, since the mean number of *Bmal1* transcripts in unsynchronized cells for all four endogenous loci is 41, compared to an average of 12 and 5 mRNA transcribed from a single copy of the reporter in the H and M clones respectively (**Figures 2.30 and 2.41**). Thus, endogenous *Bmal1* genes displayed expression levels between those the H and L reporters. Similarly, H3K27ac signal in reporter-specific regions is reduced by 5 times compared to portions of the *Bmal1* promoter common to the reporter plus the four endogenous alleles. Thus, the reporter acetylation level is also comparable to that of endogenous *Bmal1* alleles.

Surprisingly, a previous analysis on a *Bmal1* reporter highlighted the presence of a refractory period preventing the direct transcriptional reactivation of the gene following an "on" to "off" promoter transition (Zoller et al. 2015) that we did not observed in our system (**Figure 2.19**). Indeed, such phenomenon was absent in the H, M and L clones, suggesting the absence of rate-limiting reactivation steps in *Bmal1-sLuc2* expression. This mode of transcription with a single "on" state is normally mainly observed in TATA-box genes with elevated noise levels (Zoller et al. 2015) and may arise from the presence of a nearby TATA-box in a selection promoter of the FRT cassette. Although this observation could reflect the presence of small regulatory differences between the reporter and the endogenous *Bmal1*, it does not question all transcriptional bursting similarities shared by the two promoters, all the more that the existence of such refractory state was never assessed in the endogenous *Bmal1* gene.


## 3.2    Correlating transcriptional bursting behavior with molecular markers

By focusing on the transcriptional bursting properties of the *Bmal1* promoter, we observed that the integration site and the circadian phase both modulated different aspects of bursting since they predominantly affected the burst size and frequency respectively. Thus, modulation of *Bmal1* expression in these two conditions is likely to arise from different molecular mechanisms. Here, I try to understand which molecular mechanisms specifically modulate the burst size and frequency by first recapitulating the observed correlation between the presence of molecular markers and specific aspects of transcriptional bursting, and by speculating on the impact of such markers on the gene transcriptional regulation.

### 3.2.1 Enigmatic molecular origins of the burst size

Although only a limited number of FRT-compatible cell lines were used in this study, the link between the reporter integration site and the burst frequency is undeniable. Indeed, both in real-time luminescence recording and in smRNA-FISH, the burst frequency is the only aspect of bursting differing between clones H, M and L (**Figure 2.18 and 2.30**). However, the molecular specificities of these integration sites remained to be clarified. Despite both ChIP-qPCR experiments aiming at characterizing the molecular context of the inserted promoters and the identification of the genomic integration site of the reporter to analyze the surrounding chromatin landscape, no molecular marker displayed clear correlation with the burst size (**Figures 2.32 and 2.33**). Similarity, no significant correlations were observed between histone modifications at the promoter of 38 endogenous mESC genes and their burst sizes despite a broad range of values ranging from 4 to 400 transcripts per burst (**Figure 2.46**).

Both in the mESC dataset and between the *Bmal1-sLuc2* integration sites, only weak anti-correlation was observed between the burst size and markers of heterochromatin such as H3K9me3 or H3K27me3, (**Figures 2.33 and 2.46**) (Lehnertz et al. 2003; Francis et al. 2004). More surprising, such weak anti-correlation was also found with H3K4me1/3 and H3K36me3, three marker often associated with active transcription (Hon et al. 2009). However, the role of H3K36me3 in coding region of actively transcribed genes is thought to prevent intragenic transcription arising from cryptic promoters upon PolII passage, and it thus comprises repressive properties (Carrozza et al. 2005; Keogh et al. 2005). Moreover, some H3K36me3 marks observed outside of the gene body could also be associated with expression repression (Strahl et al. 2002; Landry et al. 2003). Concerning the methylated from of H3K4, it can also be associated with features of transcription repression in some contexts (Bernstein et al. 2006; Zhou and Zhou 2011; Cheng et al. 2014). This would indicate that the burst size is slightly sensitive to the chromatin state around the promoter, and especially to repressive marks. This observation is compatible with other studies showing that the general chromatin environment primarily influences the burst size (Singh et al. 2010; Skupsky et al. 2010; Dar et al. 2012; Dey et al. 2015). Indeed, local chromatin state may influence the transcriptional yield by influencing the PolII proximal pausing as observed in yeast (Churchman and Weissman 2011), or the elongation by notably determining the fraction of PolII falling of the template before completing transcription. A denser chromatin state could also reduce the transcription factor residency time on DNA, and thus limit the loading of PolII by influencing the PIC stability. Similarly, DNA regulatory elements with lower binding affinity were found to mainly decrease the burst size (Raj et al. 2006; Suter et al. 2011a; Senecal et al. 2014; Corrigan et al. 2016).

Since the chromatin marks at the integration site only marginally explain the size of the bursts, this bursting parameter is likely influenced by multiple factors, or by molecular phenomena that cannot be captured with ChIP data. Notably, a growing number of evidences point toward the existence of gene clusters with enhanced transcriptional properties. Indeed, PolII is known to form dynamic short-lived aggregates implying existence of transient, actively transcribing genomic domains (Cisse et al. 2013). Although transcription can also occur with isolated PolII, the presence and lifespan of such clusters clearly influences the transcriptional output (Cho et al. 2016). These

transient domains could consist in microenvironments, possibly phase-separated, that favor the sharing of transcriptional machinery between clusters of actively transcribing genes (Hnisz et al. 2017). In line with this theory, a recent study assessing the causes of position effect in the *Drosophila* genome observed that the presence of physical contacts between integrated transgenes and promoters or terminators of active gene were the best predictors of the transcriptional output (Corrales-Berjano et al. 2017). Thus, the recycling of rate-limiting transcriptional complexes is likely to occur in these spatially clustered genes to favor their expression. Such complexes could possibly be PolII, as its reduced presence specifically decrease the burst size (Padovan-Merhar et al. 2015). Interestingly, genes belonging to the same chromatin conformation clusters where found to share expression noise properties (Kar et al. 2017). Given the impact of burst size on cell-to-cell variability (Munsky et al. 2012), this aspect of transcriptional bursting may play a preponderant role in the expression regulation within such clusters. Indeed, while the burst frequency describes the capacity of a promoter to become active, the burst size may rather be influenced by the amount of transcriptional machinery that can efficiently complete transcription during these active time-windows. This could include PolII loading efficiency onto the promoter, its ability to escape proximal pausing and its elongation efficiency without falling from the template. However, the burst size did not correlated with estimated levels of paused PolII and could not be well explained by the chromatin state. Thus, it seems plausible that the formation of active gene clusters favoring the rapid loading of functional transcriptional machinery could largely influence the burst size. Hi-C datasets could be used to identify contacts between active genes and assess their role in defining the burst size using regression approaches.

Finally, in this thesis work, the burst size was curiously not found to correlate with elongating PolII signal. This is surprising, since it is likely to reflect the amount of PolII units transcribing when the promoter is active. However, if we consider that transcription mainly occurs under the form of PolII convoys where multiple enzymes closely follow each other (Tantale et al. 2016; Fujita et al. 2016), the distance between the elongating PolII is likely smaller than the fragments of sonicated DNA. Indeed, ~40 bp of DNA are covered per PolII units (Saeki and Svejstrup 2009). Thus, if the burst size reflects the dimension of the PolII convoys, small burst sizes could still contain a sufficient amount of PolII to account for the immunprecipitation of all corresponding DNA fragments. In turn, higher burst sizes would increase the amount of PolII per DNA fragment without necessarily improving the precipitation efficiency.

### 3.2.2   A link between burst frequency and histone acetylation

In both short-lived reporter and smRNA-FISH analyzes, temporal modulation of *Bmal1* expression levels underlined rhythmic variations in burst frequencies (**Figure 2.18 and 2.30**). Indeed, during the inactive circadian phase, time spent in the "off" state was doubled compared to the expression peak (**Figure 2.19**). Consistently, mutating the ROREs driving *Bmal1* rhythmic transcription (mRORE) led to both an upregulation of *Bmal1* expression during circadian troughs, and a constant burst frequency over the circadian cycle (**Figure 2.36 and 2.37**). Interestingly, mRORE arrhythmic and elevated expression phenotype suggests that circadian oscillations in *Bmal1* expression are mainly driven by RORE-mediated repression. Consistent with this

observation, the ROREs activators RORα and RORγ are weakly expressed in NIH-3T3 cells (Hughes et al. 2009) and do not display rhythmicity at the mRNA level (Liu et al. 2008). This link between the presence of ROREs and variations in burst frequency was consequently probably related to the recruitment of REV-ERBs, the nuclear receptors responsible for RORE-mediated transcriptional repression (Harding and Lazar 1995). Interestingly, the availability of a transcription factor and its ability to bind the promoter was already shown to predominantly affect the burst frequency (Larson et al. 2013; Senecal et al. 2014; Bahar Halpern et al. 2015b; Xu et al. 2015). However, if higher burst frequencies were known to result from increased activators availability, the decrease in burst frequency upon recruitment of a repressor had never been assessed despite the known importance of gene downregulation in generating noise, at least during *Dictyostelium* differentiation (Antolović et al. 2017).

REV-ERBs repression mechanism involves co-repressors that eventually lead to the deacetylation of histones around *Bmal1* promoter (Yin and Lazar 2005). Thus, the molecular mechanisms causing variations in burst frequencies could be linked to the changes in chromatin acetylation following the recruitment of transcription factors. To test this, we analyzed the transcriptional bursting characteristic of endogenous circadian genes (**Figure 2.42**). In the endogenous *Bmal1*, *Cry1* and *Dbp*, the circadian phase also predominantly impacted the burst frequency. If *Bmal1* and *Dbp* promoters do not share any transcription factor binding site and are regulated though different pathways (Jolley et al. 2014; Papazyan et al. 2016), they are both rhythmically acetated over the circadian cycle (**Figure 2.43**). Thus, the burst frequency may directly be influenced by the histone acetylation state of the promoter independently of the upstream regulatory mechanisms. This correlation between burst frequency and promoter acetylation state was further assessed in larger datasets. If a non-significant trend was distinguishable in 8 randomly selected endogenous NIH-3T3 genes with precisely measured bursting parameters (**Figure 2.45**, $R^2$=0.193), a significant correlation was observed in a collection of 38 endogenous mESC genes whose parameters were estimated from smRNA-FISH distributions (**Figure 2.46**, $R^2$=0.345), Interestingly, no correlation was ever observed between histone acetylation state and burst size in any of these condition. If the association between active transcription and histone acetylation is known for a long time (Struhl 1998), its specific link to burst frequency is becoming clearer. In a set of 11 representative yeast prompters, HATs and HDACs involved in early steps of transcription regulation (before elongation) where also found to influence the frequency of the bursts rather than their sizes (Weinberger et al. 2012).

A molecular explanation for this correlation could reside in the chromatin loosing properties of histone acetylation. Although burst frequency is influenced by transcription factor concentration (Nicolas et al. 2017), the frequency of specific transcription factor binding to a DNA motif (in the order of minutes (Cherstvy et al. 2008; Chen and Larson 2016)) is smaller than that of burst frequency (in the order of hour (Lionnet and Singer 2012; Yao 2017)). Thus, burst frequency likely reflects the formation of larger complexes involved in transcription initiation such as the PIC recruitment to the promoter. By participating in the establishment of a transcriptionally permissive chromatin, promoter acetylation state is likely to considerably contribute to the recruitment and stabilization of transcription regulators and machinery. Indeed, by neutralizing

lysine's positive charges, acetylation weakens the histone-histone and histone-DNA interaction (Kurdistani and Grunstein 2003; Verdone et al. 2006). The resulting relaxed chromatin structure facilitates the binding of transcription factors to DNA. If individual transcription factors bind more readily to DNA in a permissive hyperacetylated chromatin environment, the reduced time required for transcription machinery assembly on the promoter is likely to results in shorter lagging time between bursts. Consistently with the role of chromatin flexibility in influencing burst frequency, nucleosome density around the TSS was also associated with burst frequencies in yeast and mammals (Brown et al. 2013; Dey et al. 2015) as well as with expression noise (Newman et al. 2006; Tirosh and Barkai 2008; Small et al. 2014), itself largely influenced by the burst frequency (Weinberger et al. 2012; Hornung et al. 2012; Wu et al. 2017).

Although associations between the promoter acetylation state and the burst frequency were systematically observed in this study, the correlations were relatively weak. This molecular marker is thus insufficient to precisely predict the bursting feature of the gene, and other phenomena likely participate in determining its burst frequency. Among the putative additional contributing factors, the concentration and activity of specific transcription factors was already shown to influence burst frequency (So et al. 2011; Larson et al. 2013; Molina et al. 2013; Kalo et al. 2015; Kafri et al. 2016). Their impact on burst frequency may directly be linked to histone acetylation, since a considerable number of transcription activators are directly or indirectly involved in the recruitment of HATs (Sterner and Berger 2000; Deckert and Struhl 2001). However, the amount of transcription factors recruited to the promoter may have an impact on burst frequency beyond the histone acetylation state, since frequent binding to their DNA motifs is likely to shorten the time required to assemble the PIC on the promoter. However, the ability of transcription factors to influence the burst frequency of hyperacetylated promoters was never tested.

Also, in this study, we only focused on the histone acetylation state around the promoter region. However, active distal enhancers can also be acetylated (Creyghton et al. 2010; Shlyueva et al. 2014), and this marker may contribute to the final burst frequency outcome. Indeed, the formation of long-range chromatin contacts also participate in determining the burst frequency (Bartman et al. 2016; Fukaya et al. 2016; Chen et al. 2017). Since associations between distal enhancers and promoters are currently hard to predict (He et al. 2014; Whalen et al. 2016; Singh et al. 2016), this aspect of histone acetylation, although putatively informative to predict burst frequency, was neglected in this study.

Apart from acetylation, other histone marks may also be informative to predict burst frequency. Although not identified in our study, histone marks involved in PolII elongation such as H3K36me3, H3K79me2 or H4K20me1 were associated with the burst frequency of endogenous hESC genes estimated from single-cell RNA-seq data (Wu et al. 2017). However, the putative link between these marks and the burst frequency remains to be determined.

In conclusion, burst frequency and more generally gene-specific bursting features are likely to arise from multiple molecular factors. Consequently, even if individual makers can be generally linked to specific aspects of transcriptional bursting, as it is the case with histone acetylation and burst frequency, they do not allow precise prediction of the bursting outcome. Considering associations of markers could possibly lead to more robust associations.

Potentially linked to putative participation of additional factors in shaping the burst frequency, the global correlation observed in this study between burst frequency and histone acetylation levels was not systematic. Indeed, acetylation seemed to predominantly influence burst frequency in some genes, while the impact on other genes was limited. Notably, in NIH-3T3 cells, the low burst frequency measured for the *Glutaminase* and *Ctgf* genes did not correspond to the hyperacetylated state of their promoters (**Figures 2.44 and 2.45**). In contrast, the promoters of the mESCs genes *Tbp*, *Blimp*, *Tet1* and *Rex1* were lowly acetylated compared to their elevated burst frequencies (**Table 2.3**). This gene-specific differential impact of histone acetylation on burst frequency could explain the various outcomes of TSA treatment observed on the bursting profile of various genes. If, as it was the case for *Bmal1-sLuc2*, HDAC inhibition led to increased burst frequency with a CMV promoter (Viñuelas et al. 2013), the treatment did not affect the bursting properties of *NcKap1* (Suter et al. 2011a). In the *hPRL* and *prl2c2* genes, it was even predominantly affecting the burst size (Harper et al. 2011; Suter et al. 2011a). Alternatively, these various effects on transcriptional bursting upon TSA treatment could be consequences of the massive cellular perturbation induced by the drug.

Also, this specific correlation between histone acetylation and burst frequency was not observed in another study correlating enrichment of histone marks at the promoter of hESC genes with bursting features, as burst size was found to equally correlate with the presence acetylated histones (Wu et al. 2017). Even within a single promoter, variations in the acetylation state of the different Sp1 binding-sites could display specific impacts on the gene expression noise (Burnett et al. 2009). Thus, although observed at large scales in this study, the link between histone acetylation and burst frequency does not unanimously characterize all genes. It seems however that the trend is observed in a large fraction of genes.

## 3.3 Future perspectives

With this study, we contributed to the understanding of the molecular mechanisms influencing transcriptional bursting. However, this work also raises new questions and comments, either specific to this study or broadly applying to the transcription field. Notably, our understanding of the aspects of transcription influencing the burst size remains very limited. Also the exact mechanisms trough which histone acetylation increases the burst frequency remain to be defined. In this section, I will enumerate unanswered aspects of this study, and propose experiments or future directions that could help in clarifying these open questions. I will also mention specific aspects of this study that could be improved.

### 3.3.1 Gene-specific modulation of the acetylation state

An important finding of this study is the correlation between histone acetylation and the burst frequency in a large fraction of genes. However, causality was not assessed. Indeed, we never tried to experimentally modulate the acetylation state of a gene (typically *Bmal1-sLuc2*) to examine its direct impact on transcriptional bursting. Such approach could be easily implemented, as the TSA HDAC inhibitor was found to efficiently affect *Bmal1-sLuc2* expression levels (**Figure**

**3.38**). If the expression level during the circadian peaks in presence of 0.5µM of TSA resembled that of the untreated condition, the drug efficiently increased the expression level during the circadian troughs. Thus, it could be interesting to verify that this effect is caused by TSA-mediated variations in burst frequency, as this would reveal the direct influence of histone acetylation on burst frequency. If this approach is straight forward, TSA is however not the optimal tool to experimentally alter acetylation levels. Its major drawback is the broad spectrum of its targets, which comprises all class I and II HDACs (Vanhaecke et al. 2004). Consequently, the impact of this drug on gene regulation is extensive and a large fraction of genes get simultaneously missregulated upon drug application (Lopez-Atalaya et al. 2013). Although we could assess the impact on transcriptional bursting shortly after the drug treatment, the missregulation of other genes could impact the transcription properties of the *Bmal1* reporter. Consequently, alternative approaches specifically targeting the reporter promoter for acetylation would be more appropriated. Recently, tools based on the CRISPR-Cas9 systems were developed to tether HATs to specific genomic locations defined by the expression of a gRNA (Hilton et al. 2015). However, this strategy currently suffers limited impact on the expression of the target gene compared to other dCas9-activators fusion proteins (notably viral transactivation domain) (Chavez et al. 2016). However, considering the rapid progresses in the RNA-guided gene activation approaches using CRISPR-Cas9 derivatives, proper tool for site-directed acetylation or deacetylation may be available soon.

### 3.3.2   Identification of additional mechanisms influencing specific aspects of transcriptional bursting

By detecting the link between acetylation and burst frequency, we proved in this study that simple regression analyses were valid approaches to identify molecular mechanisms involved in transcriptional bursting. However, a major limitation for such correlation approaches is the limited amount of publicly availably datasets fulfilling all requirements. Indeed, such datasets require (i) bursting parameters ideally measured in a large subset of genes, (ii) the corresponding mRNA half-lives if the bursting parameters need to be estimated from static distributions (such as smRNA-FISH or single-cell RNA-seq), and (iii) signal quantification for specific markers participating in transcription regulation around the genes and their promoters (typically ChIP-seq datasets). Problematically, all of these requirements should come from comparable experimental backgrounds, i.e. the same cells in the same culture conditions. The availability of such datasets is currently limited. However, in this study, we highlighted the robustness of smRNA-FISH to infer transitional bursting parameters by comparing it with short-lived reporter experimental approach. smRNA-FISH can more readily be applied to large sets of genes than real-time approaches, and in the past years, a growing number of studies simultaneously quantified the absolute number of many genes (Battich et al. 2013; Singer et al. 2014; Chen et al. 2015b). Thus, suitable datasets may soon be available to robustly identify the functional genomic markers that, in addition to histone acetylation, influence specific aspects of transcriptional bursting. Also, the correlation should be done with an increased number of markers participating in gene regulation such as various transcription factors, co-factors, different PolII CTD phosphorylation states and large collections of

histone marks. Notably, the presence of specific HATs and HDACs could be correlated to the burst frequency in place of acetylation levels. Indeed, most of these enzymes can broadly acetylate/deacetylate several histone tail lysine residues (Roth et al. 2001; Seto and Yoshida 2014), but the consequences on transcription could depend on the type of enzyme recruited. This could explain why certain genes display little correlation between histone acetylation and burst frequency as they may recruit different HATs and HDACs than genes with a more marked correlation. Importantly, associations of genomic markers should also be taken into account, as they are likely to work synergistically to modulate transcription and consequently transcriptional bursting.

Concerning the burst size, we were unable in this study to identify molecular mechanisms that reliably influenced it. However, as mentioned above, the size of the bursts could notably be influenced by the sharing of transcription regulators during the formation of chromatin contacts between active transcription sites. To clarify the role of the three-dimensional genomic organization in modulating the burst size, one could use Hi-C maps to determine the number of contacts between the gene of interest and other active genes. Lately, a similar approach revealed that long-range contacts could explain up to 25% of the expression rate of randomly integrated reporters (Corrales-Berjano et al. 2017). Thus, the number and frequency of contacts between the gene locus and other regions of interest such as active promoters, active terminators or distal enhancers should be added to the genomic features to be correlated with burst size and frequency. Indeed, some mechanisms regulating transcriptional bursting may not be directly linked to the presence of ChIP-seq markers at the promoter.

Ultimately, such Hi-C maps could serve to identify regions with high of low number of predicted contacts with active genes. Using targeted genome editing, a reporter designed for the monitoring of transcriptional bursting (such as *Bmal1-sLuc2*) could be integrated in these regions. Comparing the bursting signature of the same reporter at these loci differing by the number and frequency of their long-range interaction with active genes could confirm the predicted role of these latter in modulating specific aspects of transcriptional bursting.

Thus, after showing that correlating bursting features with molecular markers is a valid approach to understand the mechanisms shaping bursting, the method could be greatly improved by increasing the amount of genes, the type of genomic datasets used for correlations or the amount of ChIP-seq markers. These improvements should increase the robustness of the predictions, and possibly reveal subtle contributions of markers (or associations of markers) in tuning either the burst size or the burst frequency.

### 3.3.3   Mechanistic insight into burst frequency

In this study, we separated the expression level into two transcriptional bursting parameters, the burst size and burst frequency. These transcriptional bursting kinetics are approximations that satisfyingly permit a quantitative description of the stochastic transcription process, but they are poorly informative regarding the mechanisms ongoing at the promoter level. Indeed, the identification of a correlation between the histone acetylation state at the promoter and the burst frequency does not clearly define how specific transcription factors, PolII, or any other

transcriptional regulator behave in presence or absence of histone acetylation. In a hyperacetylated permissive chromatin state, an increase burst frequency could notably arise from facilitated recognition of DNA binding sites, strengthened interactions between DNA and transcription regulators or between regulators, or increased residence time on DNA. Also, these histone acetylation-mediated changes in kinetics could apply to specific regulators, or in contrast broadly affect most participants to the gene regulation. The acetylation state could even tether specific regulatory factors to the chromatin that cannot access in another acetylation contexts.

In the recent years, great improvements were made to image eukaryote transcription factor dynamics at the single molecule level (Gebhardt et al. 2013; Chen et al. 2014; Morisaki et al. 2014). Also, transcription factor binding kinetics can also be inferred using Fluorescence recovery after photobleaching (FRAP) (McNally et al. 2000; Yao et al. 2006; Sprouse et al. 2008). Such approaches could be used to clarify the impact histone acetylation has on the transcription regulators, and that eventually results in burst frequency variations. Indeed, by following individual factors, it could be possible to precisely identify the aspect of their dynamics fluctuating with the histone acetylation level.

## 3.4    Concluding remarks

Although transcriptional bursting is a recent research field, it has been intensively investigated over the past decade. Notably, researchers show particular interest in clarifying the associated molecular mechanisms, characterizing the bursting features of genes of interest or elucidating the advantages or inconvenient of such a mode of expression on individual cells or populations. In this final section, I will insist on the importance of studying transcriptional bursting and recapitulate the perspectives offered by a better understanding of this aspect of gene expression.

### 3.4.1    Focus on bursting to understand transcription mechanisms

Transcriptional bursting allows a more precise description of the transcriptional process notably by separating the expression level into a combination of burst size and frequency. This distinction clarifies the impact that specific regulatory mechanisms on have in different aspects of transcription. Typically, an element influencing the burst frequency is more likely to affect the early steps of transcription (for example the initiation), whereas an effect on the burst size likely reflects an implication at later steps such as elongation or termination. Thus, studying gene expression at the single-cell level will likely greatly contribute in a near future to clarify how specific processes such as transcription factor binding, epigenetic state, DNA topology or nucleosome occupancy precisely contribute to transcription (Chen and Larson 2016; Nicolas et al. 2017).

Additionally, the study of transcriptional bursting is likely to elucidate regulatory properties of specific genes as well as their cellular function. Indeed, resemblances in transcriptional bursting signatures are observed between similarly regulated genes (Bengtsson et al. 2005; Newman et al. 2006; Stewart-Ornstein et al. 2012), but also between genes active in the same cellular pathway or with related cellular roles (Padovan-Merhar et al. 2015; Owens et al. 2016).

With detailed information on the transcriptional bursting signature of a gene of interest combined with consequent knowledge on the associated molecular mechanisms, it will also become possible to very precisely tune its expression. For example, increased expression level can be achieved by specifically targeting molecular mechanisms affecting the burst size or the burst frequency, with a substantially different transcriptional output in both cases. Typically, transiently targeting mechanisms responsible for tuning the burst size could result in a massive expression enhancement over a limited period of time. On the other hand, specifically strengthening the molecular aspects influencing the burst frequency would result in a more constant expression with reduced noise levels. Thus, the different facets of transcriptional bursting could be used to adapt overexpression mechanisms to specific experimental requirements.

### 3.4.2 The importance of cell-to-cell variability

Despite the participation of multiple factors in gene expression noise such as post-transcriptional regulation (Albayrak et al. 2016; Dacheux et al. 2017) or extrinsic factors (Raser and O'Shea 2004; Rosenfeld et al. 2005), transcriptional bursting is a major determinant of cell-to-cell variability in isogenic populations (Elowitz et al. 2002; Blake et al. 2003). Because it arises from biophysical phenomena such as low molecular concentrations, diffusion or transcription factor dynamics (Paulsson 2004; Pedraza and Paulsson 2008; Schoech and Zabet 2014), randomness in transcription is unavoidable. Thus, cells adapted mechanisms to either favor or reduce gene expression noise in function of the genes.

In some cases, promoting noise in the expression of certain genes may be beneficial as it generates non-genetic variability in populations of genetically identical cells. In unicellular organisms, elevate noise level in the expression of some genes is essential to anticipate a response to adverse environmental conditions by sustaining sub-population capable or resisting rapidly to various stresses (Balaban et al. 2004; Lewis 2007). This strategy confers evolutionary advantages in unstable environments (Bódi et al. 2017), and notably *E. coli* strains are significantly more noisy in nature than in laboratory cultures (Wolf et al. 2015). This non-genetic variability can even favor acquisition of genetic variations in the population, as some bacteria maintain high levels of expression noise in DNA damage genes even in mutagenic conditions to facilitate mutagenesis in some cells (Uphoff et al. 2016).

In multicellular organism, many cell-fate decisions during development or in stem cells are determined stochastically. Indeed, a large fraction of embryonic genes display high levels of expression noise (Chang et al. 2008; Boettiger and Levine 2009; Singer et al. 2014). In bistable systems, variability in expression of key differentiation genes can lead to stable states though reinforcement mechanisms such as feedback loops (Losick and Desplan 2008). Thus, initially stochastically determined cell-fates can maintain cells in a differentiated state for prolonged periods of time.

In other cases, gene expression noise should be avoided. It is notably the case of the circadian clock that needs to accurately maintain a robust biological time despite the stochasticity inherent to its expression (Gonze et al. 2003). Also, the spatio-temporal expression pattern of some developmental genes is tightly regulated (Boettiger and Levine 2013). Different strategies can be

adopted to reduce the effect of stochasticity. Commonly adopted mechanisms to buffer gene expression noise consist in operating feedback loops (Becker-Weimann et al. 2004; Boettiger and Levine 2013), increasing gene product half-lives (Little et al. 2013) or even retaining transcripts in the nucleus (Battich et al. 2015; Bahar Halpern et al. 2015a). In such genes, suboptimal control of gene expression can have disastrous consequences. Notably, increased noise in the expression of the *NF1* tumor suppressor gene can lead to morphological aberration in neurons (Kemkemer et al. 2002). Noise-increasing mutations in key developmental regulatory genes were also shown to cause incomplete penetrance in *C. elegans* gut formation (Raj et al. 2010).

Cell-to-cell variability also plays an important role in response to treatments. When targeting noisy cellular processes, treatments can lead to response variations between the cells. Notably, the response to TRAIL treatments for apoptosis stimulation highly depends on the uneven presence of Caspases in the target cells (Spencer et al. 2009), and emergence of melanoma cells resistant to drugs treatment depend on the massive co-expression of resistance marker genes in a very small fraction of cells (Shaffer et al. 2017). In patients, expression noise also contributes to resistance of sub-population of cells to drugs and chemotherapy anti-cancer treatment (Sharma et al. 2010). Additionally, the efficiency of anti-HIV treatments is reduced by the latency state adopted by a fraction of viruses, while entering or exiting this quiescent viral state is stochastically determined (Dar et al. 2014).

Thus, whether it is elevated or low, noise in gene expression can have critical impact on key cellular processes, and can notably influence the development of diseases or their treatments. Thus, studying transcriptional bursting is primordial to understand how cells regulate and adapt to this variability in contexts like cell-fate decision, response to environmental variations or regulation and maintain of robustness in biological processes. Therapeutically, a better understanding of transcriptional bursting and its role in gene expression noise could potentially limit the development of diseases and the resistance to treatments. For instance, combining HIV-reactivating drugs with compounds favoring expression noise such as chromatin remodelers largely contribute to the efficiency of anti-viral treatments (Dar et al. 2014). Thus, understanding the implication of gene expression noise in diseases and treatments can also lead to solution to remedy.

# Chapter 4    Material and methods

## 4.1    Cloning

### 4.1.1    Cloning of promoters into the short-lived luciferase expression vector

Promoter regions were cloned into the short-lived luciferase expression vector using a Gateway cloning approach (4-1R version, Invitrogen). First, the promoters were amplified by PCR from genomic mouse DNA (*Car11*, *Per3* and *Bmal1*), a pFRT-H1 vector (Suter et al. 2011a) (H1) or a pGL3-promoter (Promega) (*SV40*) using *att*B flanked primers (**Table 4.1**).

To generate the entry clone containing the promoter flanked with *att*L sequences, the BP Clonase II (Invitrogen, 11789100) was used to recombine the *att*B-PCR products into a pDONR4-1R plasmid (**Figure 4.1**). After transformation of NEB 5-alpha Subcloning Efficiency Competent E. coli (New England Biolabs) with the entry vector, selected colonies were validated by sequencing the insert with a M13 Universal Forward Primer and a reverse T7 Sequencing Primer (**Table 4.1**). Entry clones containing the right sequence were further recombined using LR Clonase II (Invitrogen, 11791100) into the backbone short-lived luciferase expression vector containing a Gateway cassette flancked with *att*B sites (ordered from GeneScript, gene synthesis services, http://www.genscript.com) (**Figure 4.2**). Reaction products were used to transform bacteria, and corresponding clones were amplified to be midipreped using NucleoBond Xtra Midi EF (Macherey Nagel).
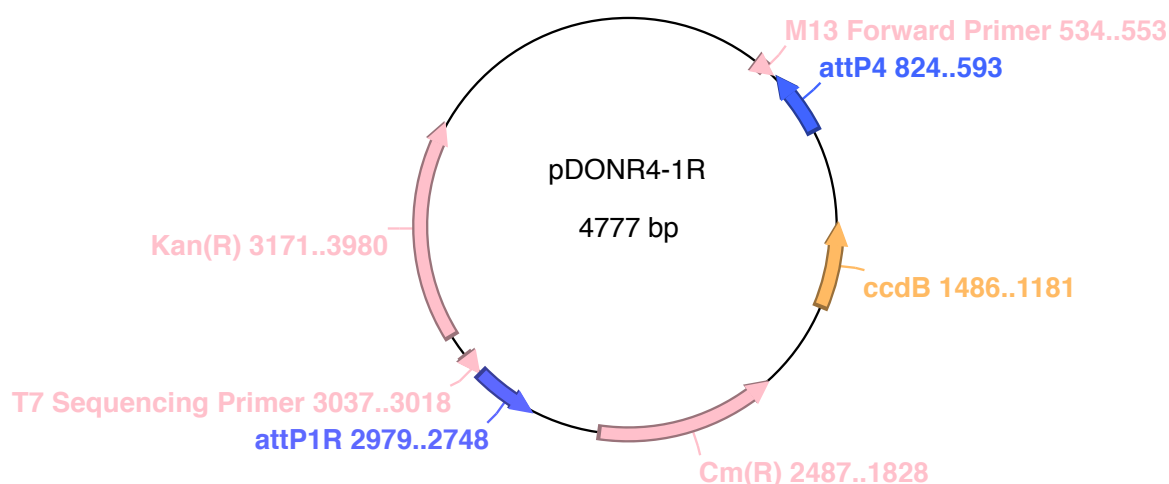


**Figure 4.1 Graphical representation of the pDONR4-1R vector**
Vector used to generate the Gateway ENTRY clone by inserting the Gateway PCR products into the Gateway cassette.

| Name | Sequence 5'->3' | Sens | Size | Useed in |
|------|-----------------|------|------|----------|
| | Gateway cloning | | | |
| attB4F_Car11I3 | ggggacaactttgtatagaaaagttggccaagtgacatggtctca | F | 45 | *Car11* promoter amplification |
| attB1R_Car11I3 | ggggactgcttttttgtacaaacttgaaagtgagttccaggacagc | R | 46 | *Car11* promoter amplification |
| attB4F_Per3_TSS | ggggacaactttgtatagaaaagttgtctgaaaatacggttggattc | F | 47 | *Per3* promoter amplification |
| attB1R_Per3_TSS | ggggactgcttttttgtacaaacttgagaagtatgttcagaagcca | R | 48 | *Per3* promoter amplification |
| attB4F_Bmal1_TSS | ggggacaactttgtatagaaaagttggggctacaacagaacaact | F | 45 | *Bmal1* promoter amplification |
| attB1R_Bmal1_TSS | ggggactgcttttttgtacaaacttgcccgcacccgcact | R | 41 | *Bmal1* promoter amplification |
| attB4F_H1 | ggggacaactttgtatagaaaagttgtagaccgtacgtgattggt | F | 45 | H1 promoter amplification |
| attB1R_H1 | ggggactgcttttttgtacaaacttgactggaagtcgagcttcc | R | 44 | H1 promoter amplification |
| attB4F_SV40 | ggggacaactttgtatagaaaagttgtgcatctcaattagtcagca | F | 46 | SV40 promoter amplification |
| attB1R_SV40 | ggggactgcttttttgtacaaacttgtggctttaccaacagtacc | R | 45 | SV40 promoter amplification |
| | Site-directed mutagenesis | | | |
| Bmal1_mRORE1_F | ttggtcggaaagtactatagtggtgcgacatttag | F | 35 | mutagenesis upstream RORE |
| Bmal1_mRORE1_R | atgtcgcaccactatagtactttccgaccaatccg | R | 35 | mutagenesis upstream RORE |
| Bmal1_mRORE2_F | gaaggcagaaagtatcgaagggacggaggtgcctg | F | 35 | mutagenesis downstream RORE |
| Bmal1_mRORE2_R | ggcacctccgtcccttcgatactttctgccttccc | R | 35 | mutagenesis downstream RORE |
| | Sequencing | | | |
| M13 Forward | gttgtaaaacgacggccagt | F | 20 | Promoter cloning |
| T7 Sequencing | taatacgactcactataggg | R | 20 | Promoter cloning |
| Seq_ATG-linker_R | gtctgtttcaccggaactc | R | 19 | Mutagenesis |
| | Probe amplification | | | |
| Probe_pFRTneo_F | catccgcttacagacaag | F | 18 | Southern blot probe generation |
| Probe_pFRTneo_R | gagaataggaacttcggaat | R | 20 | Southern blot probe generation |
| | Inverse PCR | | | |
| iPCR_pFRTneo_R | gggactatggttgctgac | R | 18 | Integration site identification |
| iPCR_pFRTneo_F | caccacaaaggaaaaagctg | F | 20 | Integration site identification |
| | qPCR primers | | | |
| Qcyc_F | ggagatggcacaggaggaa | F | 19 | RNA expression |
| Qcyc_R | gcccgtagtgcttcagctt | R | 19 | RNA expression |
| QLuc2_F | taaggtggtggacttggaca | F | 20 | RNA expression |
| QLuc2_R | gttgttaacgtagccgctca | R | 20 | RNA expression |
| QBmal1_F | ccaagaaagtatggacacagacaaa | F | 25 | RNA expression |
| QBmal1_R | gcattcttgatccttccttggt | R | 22 | RNA expression |
| QCry1_F | ctggcgtggaagtcatcgt | F | 19 | RNA expression |
| QCry1_R | ctgtccgccattgagttctatg | R | 22 | RNA expression |
| QDBP_F | aagaaggcaaggaaagtcca | F | 20 | RNA expression |
| QDBP_R | tgtacctccggctccagta | R | 19 | RNA expression |
| QGapdh_F | catggccttccgtgttccta | F | 20 | RNA expression |
| QGapdh_R | cctgcttcaccaccttcttga | R | 21 | RNA expression |
| qPCR_preCryl1_F | cttcaaccacgcctaagaca | F | 20 | RNA expression |
| qPCR_preCryl1_F | ggagcttgtttccatccaat | R | 20 | RNA expression |
| qPCR_preDBP_F | tgggacgcctgggtacac | F | 18 | RNA expression |
| qPCR_preDBP_R | gggaatgtgcagcactggtt | R | 20 | RNA expression |
| QChIP_RepRegA_F | gtgcgggccaagtttgta | F | 18 | ChIP-qPCR |
| QChIP_RepRegA_R | gtgggctggatatccaatt | R | 19 | ChIP-qPCR |
| QChIP_RepRegB_F | tgctggtgcccacactat | F | 18 | ChIP-qPCR |
| QChIP_RepRegB_R | ggcgatctcgtgcaagtt | R | 18 | ChIP-qPCR |
| QChIP_Sox2_F | agctcgcagacctacatgaa | F | 20 | ChIP-qPCR |
| QChIP_Sox2_R | tggagtgggaggaagaggta | R | 20 | ChIP-qPCR |
| QChIP_Actb_F | cagccaactttacgcctagc | F | 20 | ChIP-qPCR |
| QChIP_Actb_R | gaccctagtgtgtccccaag | R | 20 | ChIP-qPCR |

**Table 4.1 List of primers used in this study**

List of all primers used in this study. The table comprises the primer names, the 5' to 3' sequence, their sense (either forward F or reverse R), the nucleotide primer size and the context in which the primer was used. The primers are separated by types: Gateway cloning for the cloning of promoter into the short-lived luciferase expression vector, site-directed mutagenesis for the site-directed mutation of *Bmal1* promoter, sequencing, Western blot probe amplification, iPCR for integration site identification and qPCR primer. In Gateway cloning primer, the "attB" gateway recognition sequence is underlined. In the mutagenesis primers, the mutated regions are underlined. mRORE1 refers to the upstream RORE and mRORE2 to the downstream site.
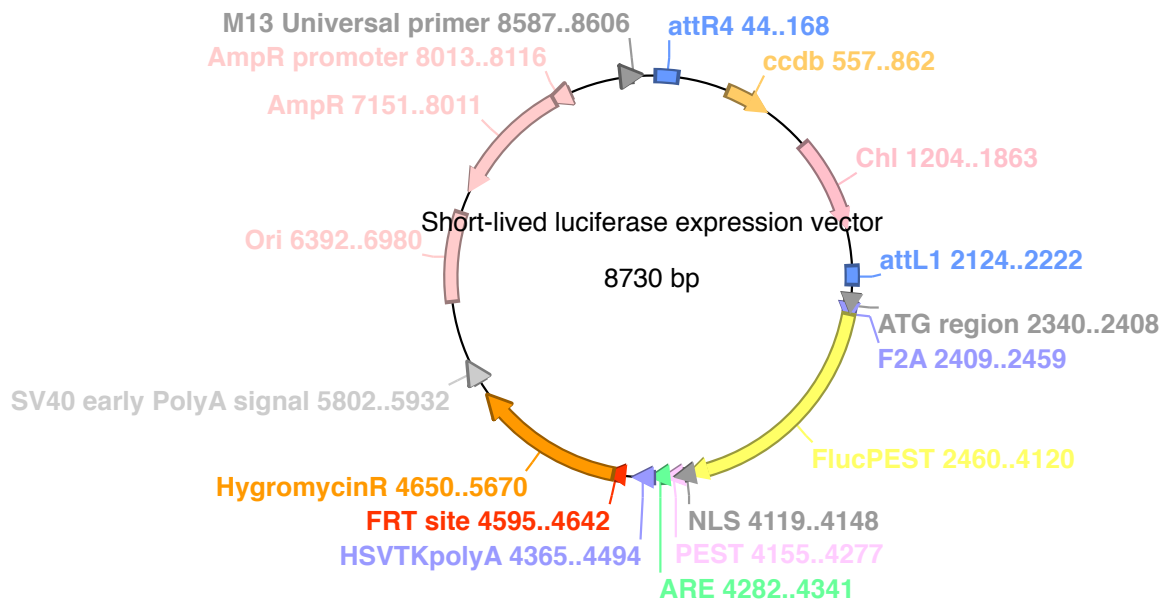


**Figure 4.2 Graphical representation of the short-lived luciferase**
This vector is the final recipient of the promoter regions cloned using the Gateway cloning system.

## 4.1.2    Site directed mutagenesis

Specific nucleotide substitutions in the promoter region of *Bmal1* were obtained using site directed mutagenesis. 100ng of WT *Bmal1-sLuc2* vector were used as template for a PCR of 20 cycles using the Kapa HiFi HotStart PCR Kit (Kapa Biosystems) and 35 bp mutation-containing primers centered on the region to mutate, and flanked with WT sequences (**Table 4.1**). DpnI restriction enzyme (New England Biolabs) was directly added to the PCR product in a 1/10 volume ratio, and incubated at 37°C for 2 hours. Digested material was directly used to transform of NEB 5-alpha Subcloning Efficiency Competent E. coli (New England Biolabs). Selected colonies were validated by sequencing the downstream region of the *Bmal1* promoter using Seq_ATG-linker_R primer (**Table 4.1**). The double mutant was generated in two rounds, with both RORE mutated separately.

## 4.2    Cell culture and cell lines generation

### 4.2.1    Cell maintenance

NIH-3T3 mouse fibroblasts were grown at 37°C in a humid environment. Unless specified, they were kept in DMEM medium (Gibco, 11965092) complemented with 10% Fetal Bovine Serum (Sigma-Aldrich, F7524) and Penicillin-Streptomycin-Glutamine antibiotics at a final concentration of 0.5 mg/ml (Gibco, 10378016).

### 4.2.2    Counting cells and estimating cell viability

Cells were counted in a LUNA Automated Cell Counter (Logos Biosystems). Suspension cells in DMSO were mixed with an equivalent volume of 0.4% Trypan Blue Solution (15250061, Gibco). 10 µl of the mix was then loaded into a LUNA Cell Counting Slide, and cell concentration was estimated from the software. The number of living cells was calculated as the total number of cells minus the "dark" dead Trypan Blue permeable cells.

### 4.2.3    Generation of homemade NIH-3T3-FRT cells

Homemade NIH-3T3-FRT cells were generated by transfecting WT NIH-3T3 cells with a pFRT-Neo plasmid (kindly provided by Dr. Jürgen Ripperger, University of Fribourg, Switzerland) (**Table 4.2**) The plasmid consisted in a modified version of the original pFRT/*lac*Zeo vector (Invitrogen) in which the *lac*Zeo selection gene was replaced by a neomycin resistance gene (**Figure 4.3**). 48 hours after transfection, cells were transferred into 15cm dishes and selected with G418 (Life Technologies) at a final concentration of 500 µg/ml. After 14 days, individual clones were picked and further expanded while maintaining the antibiotic selection.
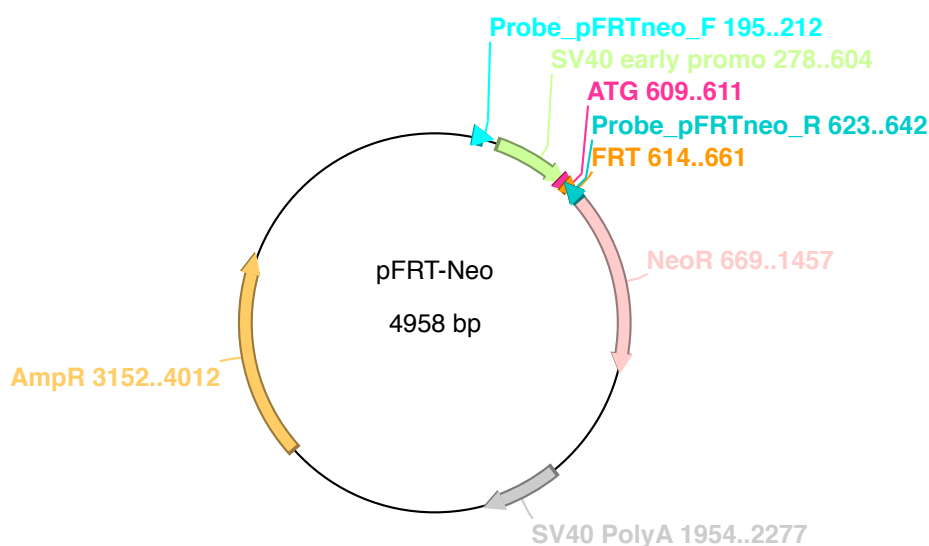


**Figure 4.3 Graphical representation of the pFRT-Neo vector**
This vector was used to generate the homemade NIH-3T3-FRT by stably integrating the vector into random genomic locations using transfection.

### 4.2.4 Transfections

16h prior to transfection, 35000 cells were plated into a 35 mm dishes (Falcon, 353001) to reach 90% confluence at transfection time. Transfection was done using Lipofectamine LTX Reagent with PLUS Reagent (ThermoFisher, 15338100). Briefly, 2 µg of plasmid were mixed with 2 µl of PLUS Reagent in Opti-MEM Reduced Serum Media (Gibco, 31985062). The proportions of the different types of DNA transfected for each experiments are specified in (**Table 4.2**). After addition of 7.5 µl of Lipofectamine LTX Reagent, the transfection reaction was incubated for 10 minutes at room temperature, before being applied onto the cells. Cells were grown in presence of transfection reagents for 36h before being used in an experiment.

| Experiment | Plasmid | Ratio | ng/35mm dish |
|---|---|---|---|
| **Short-lived luciferase transient transfection** | Short-lived luciferase expression vector | 9/10 | 1800 |
| | pMAX-GFP (Lonza) | 1/10 | 200 |
| | | | |
| **Generation of homemade NIH-3T3-FRT cells** | pFRT-Neo | 10/10 | 2000 |
| | | | |
| **Stable reporter integration into genomic FRT site** | pOG44 (Invitrogen) | 8/10 | 1600 |
| | *Bmal1* short-lived expression vector | 2/10 | 400 |

**Table 4.2 Transfection conditions**

List of plasmids transfected into NIH-3T3 cells during this work. The type of experiment is indicated, as well as the combinations of co-transfected plasmids, the volume ratio for each vectors and the corresponding amount transfected vector.

### 4.2.5 Stable integration of the short-lived luciferase expression vector into NIH-3T3-FRT cells

To generate cells stably expressing the *Bmal1-sLuc2* expression vector from a genomic FRT locus, homemade NIH-3T3-FRT clones containing a single insertion of the FRT cassette as well as commercial NIH-3T3-FRT cells (Invitrogen) were co-transfected with a pOG44 Flippase expression plasmid (Invitrogen) and a the *Bmal1-sLuc2* expression vector (**Table 4.2**). 48 hours after transfection, cells were transferred into 15cm dishes and maintained for 24 hours without any selection. Then, medium was supplemented with hygromycin B (InvivoGen) at a final concentration of 200 µg/ml. After 20 days, individual clones arising from the same parental FRT clone were pooled and expanded in selective medium.

## 4.3    Microscopy and real-time cell recording

### 4.3.1    Population real-time luminescence recordings

Real-time recording on populations of cells was done both on freshly transfected cells or on cells stably expressing luciferase from an integrated transgene. For transfected cells, co-transfection with luciferase expression vector and pmaxGFP plasmid (Lonza) used to assess transfection efficiency were done 36 hours prior to beginning of the recording. At recording time, fluorescence levels in the green channel (Filter Set 38 HE, Zeiss) were imaged with an Axio Vert.A1 microscope

(Zeiss) equipped with an HXP 120c light source (Zeiss) and a Retiga CCD camera (QImaging) to evaluate the transfection efficiency in the different conditions.

For real-time luminescence recording in populations of cells stably expressing luciferase, cells were seeded at 500,000 cells per 35 mm 16 h prior to recording. In both cases, for circadian reporters, the cells were synchronized in 100nM Dexamethasone (Sigma Aldrich, stock 200μM in ethanol) for 30 minutes. Medium was then replaced by DMEM medium (Gibco, 11965092) complemented with 10% Fetal Bovine Serum (Sigma-Aldrich, F7524), 1% Penicillin-Streptomycin-Glutamine (Gibco, 10378016) and 100μM D-Luciferin (Sigma Aldrich, L9504, stock at 100mM) and the luminescence was recorded for up to 5 days. The medium replacement was done directly prior to recording for non-circadian reporters. Luminescence was then recorded for several days in a Lumicycle32 (Actimetrics).

Parameters describing the expression pattern of the oscillating reporter (mean expression level, amplitude, period and phase of the expression peak) were inferred from the population luminescence data using the ChronoStar software (Spörl et al. 2011).

In case of recording in presence of specific drugs, the compounds were added directly in the recording medium, together with luciferin. In the negative control condition (no drug), a similar amount of the diluent of the drug (DMSO) was used to allow results comparison. Compounds used and concentrations are listed in (**Table 4.3**).

| Drug name | Target | Effect | Working concentration | Brand | Cat n° | Publication |
|-----------|--------|--------|-----------------------|-------|--------|-------------|
| SR8278 | Rev-Erbs | Antagonist | 1, 5, 10, 50 and 100 μM | Sigma Aldrich | S9576 | (Kojetin et al. 2011) |
| T0901317 | RORa and c | Inverse agonist | 50, 100, 500 and 1000 μM | Sigma Aldrich | T2320 | (Kumar et al. 2010) |
| 7-Ketocholesterol | RORa and c | Ligand | 10, 50, 100, 150 and 200 μM | Sigma Aldrich | C2394 | (Wang et al. 2010b) |
| Trischostatin A | HDAC | Inhibitor | 0.1, 0.5 and 1 μM | Sigma Aldrich | T8552 | (Yoshida et al. 1990) |

**Table 4.3 List of chemical compounds used in population real-time luminescence assays**
List of drugs added to the recording culture medium to affect various aspects of the reporter expression. For all compounds, the stock solution was diluted in DMSO. Names of the drug, published effect and target, brand, catalogue number, original reference mentioning the compound and final concentrations used in this study are indicated, as well as original publication mentioning the impact of the drug on gene expression.

4.3.2    Time-lapse luminescence microscopy

Single-cell luminescence traces were obtained by recording cells in a luminescence microscope. Cells stably expressing short-lived luciferase were diluted 1:50 into non-luminescence NIH-3T3 cells and plated in a 2.3cm glass bottom dish (FluoroDish, World Precision Instruments) to reach a final amount of $8 \times 10^5$ cells per dish 16 hours prior to recording. Cells where synchronized using dexamethasone (Sigma Aldrich) at a final concentration of 100nM for 30 minutes directly before starting the recording. After synchronization, medium was replaced and complemented with D-luciferin (Life Technologies) at a final concentration of 100μM. Luminescence monitoring was performed in a LuminoView LV200 microscope (Olympus) with a cooled C9100-13 EM-CCD camera (Hamamatsu). During recording, cells were maintained at 37°C in a humid environment with 5% $CO_2$. Images were acquired with an exposure time and time resolution of 5 minutes during 48h. Recording of the luminescence signal was optimized using a binning of 4, an EM gain of 150 and a

5x photon imaging mode. Cosmic rays from time-lapse recordings were first removed by minimalizing the signal between two adjacent frames in the ImageJ software, and single-cell were segmented and tracked using the CAST software available at https://git.epfl.ch/repo/cast-supplements.git (Blanchoud et al. 2015). CAST was parameterized with a filter min intensity of 3 and a fixed filter size of 20 pixels, an 'atrous' max size of 3 and threshold of 25, and a tracks filtering interpolation of 1.

### 4.3.3    Single-molecule RNA-FISH (smRNA-FISH)

smRNA-FISH protocols were largely inspired from Stellaris RNA FISH (Biosearch Technologies) approach. Before plating the cells, 18mm diameter round coverslip (Fisher Scientific) were incubated for 30 minutes at room temperature with a solution of PBS and 25 µg/ml Fibronectin (F0895, Sigma-Aldrich). Coverslips were then placed in 12 wells plates, and 0.3 million cells were seeded per well 16 hours prior to DEX synchronization (for condition necessitating circadian synchronization) or medium change (for non-circadian genes). When needed, the synchronization was performed with 100nM DEX for 30 minutes before changing the medium. From the seeding time, cells were always kept in 0% serum culture medium.

At time-points corresponding to specific circadian times (for rhythmically expressed genes) or 10 hours after medium change (for non-circadian genes), medium was aspirated, and cells were washed with 1X PBS. A fixation buffer containing 1X PBS and 3.7% Formaldehyde (F15587, Sigma-Aldrich) was then applied on the cells for 10 minutes at room temperature. Cells were washed twice with PBS, and permeabilized overnight in 70% ethanol at 4°C. The following day, ethanol was removed, and cells were incubated for 5 minutes in Stellaris Wash Buffer A (Biosearch Technologies).

Hybridization was achieved by placing coverslips cell-slide down on a 50 µl drop of Stellaris Hybridization Buffer (Biosearch Technologies) containing a final concentration of 125 nM of specific smRNA-FISH probes (Biosearch Technologies) and dispensed onto Parafilm. The sequences of the far-red probes used in this project are specified in (**Table 4.4**). In addition to reagents present in the original Stellaris protocol, the hybridization buffer was complemented with Ribonucleoside Vanadyl Complex (New England Biolabs) at a final concentration of 5 mM and Yeast tRNA (Ambion) at a final concentration of 250 µg/mL. The hybridization was carried on for 16 hours in a humidified chamber at 37°C in a dark environment.

After hybridization, coverslips were transferred cells-side up to fresh 12-well plate containing Stellaris Wash Buffer A, and incubated for 30 minutes at 37°C in the dark. The wash buffer was then replaced by a the same buffer containing in addition 5 ng/ml of DAPI (62248, Thermo Scientific). In no alternative cell staining method is specified, HCS CellMask Green Stain (H32714, Invitrogen) was used to stain the cytoplasmic region of cells. The compound was added into the same DAPI-containing buffer at a final concentration of 2 µg/ml. Coverslips were incubated in this complemented buffer at 37 °C for 30 minutes in a dark environment. Staining was then washed for 5 minutes at room temperature in Stellaris Wash Buffer B (Biosearch Technologies). Coverslips were finally mounted onto microscopy glass slides using a drop of Vectashield Antifade Mounting Medium (Vector Laboratories). The coverslips were sealed with nail polish, and imaged within 24h

after terminating the protocol.

| Probe n° | Luc2 | Bmal1 | Cry1 | Dbp |
|---|---|---|---|---|
| 1 | tcttcgagtgggtagaatgg | agggaaccggagagtaggtc | acgaacggacacgagctcac | aaaacaccggacgccagag |
| 2 | tagcgcttcatggctttgtg | ctttcctcttgcgattgcag | cagctggcagatgaatggag | agcaaacttctttcgcgga |
| 3 | cgtcggtaaaggcgatggtg | tgttggtaccaaagaagcca | ttccaagaaaacctcctcac | aggagctttgcaatctgca |
| 4 | gtaatgtccacctcgatatg | taacctactttccgaccaat | caggggtcgaggatatagac | ctggctctagccaattttg |
| 5 | cgaacgctcatctcgaagta | tactttctgccttccctaaa | tgatgcccacgttggaagag | ctccttctgtaccaagtgg |
| 6 | atagcgcttcatagcttctg | gttccaaaagcattcactgc | tgaagcaaaaatcgccacct | ctctttgcaactgtgggtc |
| 7 | gatccgatggtttgtattca | ccacatttgaagccttaagt | tggcatcaagatcctcaaga | ggtgaccaaatcaagcagc |
| 8 | aagctattctcgctgcacac | ccaaaatccaatgaaggcct | ttccttgaaaagcctgggaa | ctcctatagtctgggcgag |
| 9 | ccaacacgggcatgaagaac | ttttgtaacttccttccagg | tcccaaaaggctcagaatca | gcaactcaaagactgtccc |
| 10 | acagccacaccgatgaacag | tcaccccaaagtgactttca | ttcttgatagctgcatctcg | ttcaaccagtcagttctgc |
| 11 | ttgtagatgtcgttagctgg | gatgcaaagggccactgtag | aaatgcgcacgatgacttcc | aaatcctaagagcgctgcg |
| 12 | cctttcttgctcacgaatac | ttaagcttcgatccagtgtg | tccaggtcatacagtgtatg | aagcctccaggatcaggac |
| 13 | tgcacgttgaggatcttttg | tcccacatctgaagttacag | gccattgagttctatgatct | agtggttcgggcgaatgtg |
| 14 | ttgtatgatcggttagcttct | gatttgactggggctgtcac | tatatgttagaggtggctgt | atgatatgtcagtcacccg |
| 15 | tggtagtcggtcttgctatc | caaagggctcaaaggtccac | cttgctgacgagagtctgaa | caggttcatgagctcatgg |
| 16 | aggtgtacatgctttggaag | ctaacttcaaaccctggtga | tcatgcactttcctatcaca | ttctttgggcttgctgttt |
| 17 | ggtggcaaatgggaagtcac | gaaggaaggtgcttgcaagg | tcatggtcatcagacagagg | ttttccttcaggagacagc |
| 18 | cacgaagtcgtactcgttga | gaaatgtccattctctggtc | ggaaggaacgccatatttct | tgagggcagagttgccttg |
| 19 | ccactactgttcatgatcag | caaatgtagtgtctacagcg | gacaggccatctgtatcaaa | cggcgcaaaaagactcggg |
| 20 | tgaatcggacacaagcggtg | gatactgcagctgttgccaa | gaggtcgttcaaagtttgcc | aaagtgcgttcccacagca |
| 21 | atgatctggttgccgaagat | aacttcaccattaatgcact | aaattcacgccacaggagtt | tattccacgtccccgaaag |
| 22 | tgaaatggcaccacgctgag | cagtaagcttcacagactgt | tgtttgtggctgctgtataa | cagcaagaaggcgtccagg |
| 23 | agcgtggtgaacatgccgaa | taaaagctgttctctctcta | cttccattttgtcaaagcgt | gggaagaggagctgcaaga |
| 24 | aaagccgcagatcaagtagc | tggcttttatggagtcagta | tgtctggctaaatggtggat | tgtgtccctagatgtcaag |
| 25 | aagcggtacatgagcacgac | acttgtatcaatggctctga | ctgcatcaagcagtaactct | aaaggtcattagcacctcc |
| 26 | aagctgcgcaagaatagctc | tgtctgaagttacacatggt | atccaacttccagcatttat | agatcagcgggatcaggtt |
| 27 | gcagggcagattgaatctta | acactggagcaggtttagtt | aaaggaactgcaggacagcc | ctggaatgcttgacagggc |
| 28 | gaagctaaatagtgtgggca | gcccaaactcaatgatgagg | cagggcagtagcagtgaaaa | ccgagggtcaaaagtctca |
| 29 | tacttgtcgatgagagtgct | taatgcaagctatccacagc | atctgtcctcctaccaaaac | tcctctgagaagcggtgtc |
| 30 | atctcgtgcaagttgcttag | gatgcttctgtgtgcacaatga | aacgcctaatatagtctcca | ttggttgaggcttcagttc |
| 31 | ctggtaggtggaagcgtttg | ccatacatctgaaatgacca | aagcctcttaggacaggtaa | tttccttgccttcttcatg |
| 32 | ttgtttctgtcaggccgtag | aaatatccacatggggggact | gcattccaaggatcgtagat | cagtacttctcatccttct |
| 33 | cttcgggggtgatcagaatg | ttcatgtgctgaacagccat | caaacacttggcaaccttct | gcttcattgttcttgtacc |
| 34 | caccttagcctcgaagaagg | gttggtggcacctctcaaag | ggtttggggtaattaactcc | tcttcttgcatctctcgac |
| 35 | ttaacgtagccgctcatgat | atctgctgccctgagaatta | tcaatattcagtctgcttgc | acagatatctggttctcct |
| 36 | gtcgatgagagcgtttgtag | tcacatcctacgacaaacaa | gctgatagatctgcttcatt | gttttccttctccaggaag |
| 37 | cggtccacgatgaagaagtg | gacaaagaggatcttccctc | accgaggcgagaagacctag | cacggtagtgggacagctc |
| 38 | cttgtatttgatcaggctct | ggatcttgaagacagactcg | tacaactcgggacattctct | ctggtagcgtgaaagcaca |
| 39 | tcgaagatgttggggtgttg | gcatagccgggagggcccag | cacttccttgagagcaattt | gcaaggaacaggactctgc |
| 40 | gtcatggtttttaccgtgttc | ggattgtgcagaagcttttt | cgctttcgtatcagttactg | ggaagtcaggtgtaagtct |
| 41 | acatagtccacgatctcctt | gccccatcttcgtgggtggc | aacttcagttgcgaggactc | cactggccatgggacaagg |
| 42 | agcttcttggcggttgtaac | taaacttatagcaattagtt | attaccacttgtcctatatg | catcatgacgttcttcggg |
| 43 | ctcgtccacgaacacaacac | tgttagctgcgggaaggttg | gcaaatctcttaccaagatc | acaagggaggtgggggaag |
| 44 | ccttaatgagaatctcgcgg | cccctggaatgcctgggac | gtctattctaactgtggctt | gctaagagcacacacagga |
| 45 | gagggaagccgtgagaattg | tgagcctgccctggtaatag | gcctatgcacttgattacat | cggttggatggatgggttc |
| 46 | caagcagcagggtgtctatc | atcatcgatgcctatgtg | gaggctgtaacacagatggc | tcaggattgtgttgatgga |
| 47 | agtttagacgttgatcctgg | ccagccccgcatctgcttcc | cacaatgcactgcgcaagtc | cccgggttctcaagattta |
| 48 | gcgagaagcttgattcacta | | gcggagagacaaaggagctc | cttgagtcagagccagtag |

**Table 4.4 single-molecule RNA FISH probes**
List of the smRNA-FISH probes used to target *Luc2*, *Bmal1*, *Cry1* and *Dbp* transcripts. All probes were combined with a Quasar 670 Dye emitting in far-red. Both *Gapdh* and *Neat1* probes were ordered as DesignReady Probe Sets: Stellaris FISH Probes Mouse Neat1 5' Segment with Quasar 670 Dye (VSMF-3031-5, Biosearch Technologies) and Stellaris FISH Probes Mouse Gapdh with Quasar 670 Dye (VSMF-3015-5, Biosearch Technologies).

In addition to HCS cell-mask Green, alternative compounds were tested to mark the cytoplasm or plasma membrane of cells:

WGA Alexa Fluor 488 conjugate (Invitrogen) was used to stain the cells prior to fixation. The culture medium was changed for Hank's balanced salt solution (HBSS) without phenol red and containing WGA at a final concentration of 5 µg/ml. Cells were incubated for 10 minutes at 37°C in the staining medium, and washed 2 times with PBS. Immediately after the washes, cells were fixed and further treated for following steps of the smRNA-FISH protocol.

Carboxyfluorescein succinimidyl ester (CFSE) was used to label intra-cellular space, and was also applied to the cells prior to fixation. The cells were centrifuged for 4 minutes at 1000 rpm. The cellular pellet was then resuspended in PBS containing CellTrace CFSE Cell Proliferation Kit (C34570, Invitrogen) at a final concentration of 5 µM. After 10 minutes incubation at 37°C, the dye was quenched by adding an equal volume of culture medium. After a 5 minutes incubation at 37°C, cells were further centrifuged in similar conditions and resuspended into DMEM culture medium. SmRNA-FISH protocol was continued with the cell fixation step.

Anti-Cadherin antibodies were used in a simultaneous immunofluorescence (IF) and smRNA-FISH approach to stain cellular membranes. Anti-pan Cadhering primary antibodies (ab16505, Abcam) were added at a final concentration of 2 µg/ml directly in the hybridization solution, and incubated with the smRNA-FISH probes for 12 hours at 37°C in a humid chamber. Goat anti-Rabbit IgG Alexa Fluor 488 (A32731, Invitrogen) was used as secondary antibody. It was diluted 1/1000 in the first Wash Buffer A solution following the hybridization step, and left incubated for 30 minutes at 37°C.

Cells were imaged using an upright DM5500B motorized widefield fluorescent microscope (Leica) equipped with a DFC 3000 Black and White CCD Camera. smRNA-FISH was monitored using an HCX PL APO oil objective (magnification of 63X, NA of 1.4). Exposure time varied between the samples for DAPI and HCS-cell mask channels to reach high intensities without reaching saturation, but FISH probes (far-red filter) were always exposed 200ms per stack. For each XY position, at least 30 stacks separated by 0.3 µm were taken, corresponding to a minimum of ~10 micron in Z axis.

To quantify the absolute number of transcript per cell, the open-source CellProfiler software was used (http://cellprofiler.org/) (Carpenter et al. 2006). First, nuclei were identified from the DAPI channel using a shaped-based method. From each identified nucleus, the green channel image was used to identify cytoplasmic regions using a propagation approach. To identify individual transcripts, all stacks of the far-red channel were first projected using maximal intensity. Morph module was then used to homogenate the background, and smRNA-FISH dots were identified using a RobustBackground thresholding approach. Objects were then related to each other to obtain the mean number of transcript per cell, per nucleus and per cytoplasm. Nucleus, cytoplasm and full cell area were also estimated from CellProfiler.

## 4.4    Molecular biology and biochemical assays

### 4.4.1    Blocking protein and mRNA synthesis

Estimation of the protein and transcript half-life of the short-lived luciferase reporter was achieved in transiently transfected cells. After recording luminescence for 22h at the population level, cells were either kept untreated, treated with 25g/ml Cycloheximide (Sigma-Aldrich, C7698), or with 5g/ml Actinomycin D (Sigma-Aldrich, A1410). Protein half-lives were estimated from the slope of the trendline applied to the exponential decay plotted in natural logarithmic scale of the recordings treated with Cycloheximide. Transcript half-live was estimated as done previously (Suter et al. 2011a) by Benjamen Zoller (Naef lab) from the luminescence decay of the condition treated with Actinomycin D (detailed in section 4.5.1).

### 4.4.2    Southern blot

Southern blot assays were performed by Lofstrand Labs Limited (Gaithersburg, Maryland 20879 USA). Briefly, 10 μg of gDNA from isogenic colonies of homemade NIH-3T3-FRT were digested with HindIII restriction enzyme  (New England Biolabs). Digested samples were then loaded onto a 350 ml 0.65% agarose gel with 1 Kb Plus DNA Ladder (Thermo Fisher) loaded on both sides of the gel for 16.2 hours at 60V. The gel was then treated two times with a solution of 0.5M NaOH and 1.5M NaCl for 25 minutes each, followed by two additional treatments in 0.5M Tris pH8 and 1.5M NaCl for 25 minutes each. The membrane was then transferred onto Nytran Supercharge nylon membranes (Whatman) in a TurboBlotter (GE Healthcare Life Sciences) containing 10X SSC for 5 hours. The membrane was then UV-linked and air-dried.

To detect the presence of genomic FRT cassette in digested gDNA, a 32P random priming probe was used. To generate the probe, a 448 DNA fragment was amplified from the pFRT-Neo plasmid by PCR using Probe_pFRT_F and Probe_pFRT_R primers (**Table 4.1**). 25 ng of purified PCR product were then denaturized by heating at 95°C for 5 minutes, and used as template for an amplification step using an Exo-Klenow (DECAprimeTM II Kit, life technologies) for 15 minutes at 37°C in presence of a final concentration of 2mCi/ml [a-$^{32}$P]dCTP. The radioactive probe was then purified using NucAway Spin Columns (Invitrogen). Before hybridization, the membrane was pre-hybridized in 6X SSC, 5X Denhardt's solution and 0.5% SDS at 68°C for three days. The hybridization was carried out at 68°C for 18 hours with the random primed 448bp probe at a concentration of $3.5 \times 10^6$ dpm/μg in the same buffer than for pre-hybridization. The membrane was then washed three times in 2C SSC+0.1% SDS at 68°C for 20 minutes each. Finally, the membrane was autoradiographed for 19h in an intensifier screen at -80°C.

### 4.4.3    Measuring mean luciferase mRNA molecules per cell

To estimate the mean number of mRNA molecules per cell, one million cells stably expressing le *Bmal1-sLuc2* vector from the FRT insert (FRT clones 15, 4 and 7) were plated into 6 well plates 16 hours prior to harvest. Cells were not synchronized during the procedure to assess average mRNA expression levels. Non-luminescent cells were also seeded in parallel in comparable conditions to serve as standard after spike-in with *in vitro* transcribed luciferase RNA. For each of the wells

seeded to harvest mRNA, another well was prepared in an identical way for estimating the number of cells per well.

*In vitro* transcribed short-lived luciferase was synthetized using the MEGAscript for T3 promoter kit (Ambion) from a 2.5 kb fragment corresponding to the coding sequence of the short-lived luciferase of the pKS short-lived luciferase vector (Suter et al. 2011a) linearized using SpeI restriction enzyme (New England Biolabs) (**Figure 4.4**). At RNA harvesting time (16 hours after cell seeding), standard condition were spiked in with 1, 4 16, 64 or 256 million of *in vitro* transcribed luciferase molecules per well (corresponding to 1.3 to 340 picograms per well). RNA in both standard and sample conditions was then harvested using RNeasy Mini Kit (Qiagen), and 500 ng per condition was reverse transcribed using Revertaid first strand cdna synthesis kit (Thermo scientific). cDNA samples corresponding to a starting material of 100 ng of RNA were then used in quantitative PCR (qPCR), using Power SYBR Green Master Mix (Invitrogen) and primers targeting Cyclophilin B transcripts (QcycF and QcycR, **Table 4.1**) and luciferase transcripts (QLuc2F and QLuc2R, **Table 4.1**) in a QuantStudio6 device (Applied Biosystems). DeltaCt of the luciferase molecules was calculated from the mean Ct value of the Cyclophilin B in each condition. The mean number of transcripts per well was evaluated by comparing the DeltaCt of the standard with those of the samples. Finally, the number of transcript per cell was obtained by dividing this value by the estimated number of cells per well.
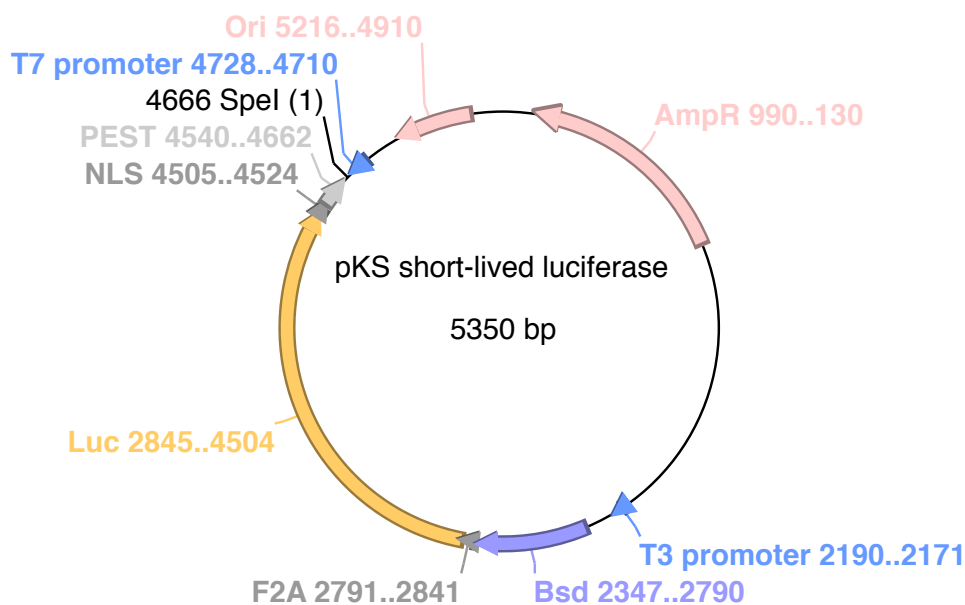


**Figure 4.4 Graphical representation of the pKS short-lived luciferase vector**

The region of this plasmid between the two SpeI restriction sites was used as a template for *in vitro* transcription, initiated at the T3 phage promoter

### 4.4.4  Chromatin ImmunoPrecipitation (ChIP)

The reagents and the protocol used in the following section correspond, unless specified, to the MAGnify Chromatin Immunoprecipitation System (Applied Biosystems). This protocol was applied to both ChIP-qPCR and ChIP-seq samples. The variations in the analysis are specified bellow.

Six million cells were first plated into a 10cm dish and synchronized 16h later using 100nM DEX for 30 minutes. At the desired time-point (CT 4 for *Bmal1* expression trough and CT 16 for peak), cells were washed with PBS and centrifuged for 4 minutes at 1000 rpm. The supernatant was discarded, and the cells crosslinked into a solution of PBS with 1% formaldehyde for 10 minutes. The reaction was stopped by addition of glycine at a final concentration of 0.125 M. After PBS wash at 4°C, cells were lyzed in presence of protease inhibitors for 5 minutes on ice. Cellular debris were centrifuged, and the DNA-containing supernatant was further used. DNA sonication was carried on a E220 Focused Ultrasonicator (Covaris), with 20 minutes sonication time at 140 W and a duty cycle of 5% for 120 cycles per samples. Sonication efficiency was verified on a 0.8% agarose gel, and samples were further sonicated in case of clear enrichment of 200-500 bp fragments.

Prior to being applied on sonicated DNA, antibodies were incubated onto Dynabeads Protein A/G for 1 hour at 4°C with continuous rotations. Antibodies specifications are listed in (**Table 4.5**). The immunoprecipitation was carried on DNA equivalent to 0.5 million cells per condition. A 1/10 sample was kept for each cell condition as input to normalize the precipitated conditions. In the immunoprecipitation samples, beads-antibodies complexes were added to the DNA, and the mixtures were rotated for 2 hours at 4°C. Unbound DNA was washed away on a DynaMag-PCR Magnet (Applied Biosystems) using three time the IP Buffer 1 and two times the IP Buffer 2. The crosslinks were reversed in Reverse Crosslinking Buffer complemented with Proteinase K during a 30 minutes incubation at 55°C followed by a 65°C incubation of 30 minutes.

DNA (including the input samples) were further purified using Purification Magnetic Beads, and eluted in 100 μl of DNA Elution Buffer to each tube. Presence of DNA was tested using Qubit Fluorometric Quantitation (Invitrogen)

For qPCR analysis, 1/10 of eluted sample were used as template. The qPCR was carried on a QuantStudio6 device (Applied Biosystems) using Power SYBR Green Master Mix (Invitrogen) and couples of primers targeting a specific promoter **(Table. 4.1)**.  % of Input was calculated as follow: $100*2\wedge((\text{Input Ct} - 3.32) - \text{Ct IP sample})$, the 3.32 subtraction corresponding to the correction for Input dilution.

For ChIP-seq samples, libraries were prepared following the NEBNext Ultra II DNA Library Prep Kit for Illumina (E7645S, New England Bioloabs) protocol. Briefly, 10 ng of DNA per condition was DNA fragment end-repair, 5′ Phosphorylation and dA-Tailing was carried on by NEBNext Ultra II End Prep Enzyme Mix for 30 minutes at 20°C. The adaptors adaptor ligation and U-loop excision steps were done using a NEBNext Ultra II Ligation Master Mix together with 1/10 diluted NEBNext Adaptor for Illumina (E7335, New England Bioloabs). After a 15 minutes incubation at 20°C, USER Enzyme was added to the ligation mixture, which was further incubated for 15 minutes at 37°C. Unligated material was removed by cleaning-up with AMPure XP Beads (Beckman Coulter). PCR amplification and multiplexing of the library and was done for 8 cycles using NEBNext Ultra II Q5

Master Mix and different Index Primers (E7350, New England Biolabs) for each ChIP condition. The PCR reactions were cleaned individually using AMPure XP Beads, and eluted in TE Buffer.

Sequencing of the ChIP-seq samples was done by multiplexing 12 samples on a unique sequencing lane. Prior to sequencing, samples quality was verified in a Fragment Analyzer (Advanced Analytical), and Qubit Fluorometric quantitation was used to load equi-molar amounts of each samples. High throughput sequencing was performed at the gene expression core facility of EPFL on a NextSeq 500 sequencer (Illumina), using 400mio single-end reads with 75 cycles.

The obtained reads were mapped on the mm10 version of the mouse genome using HTSStation EPFL interface (David et al. 2014). In parallel, reads were mapped on a custom genome corresponding to a 9553bp region of the pFRT-Neo/*Bmal1-sLuc2* expression vector genomic reporter using the STAR algorithm (Dobin et al. 2013). Mapped reads were normalized on the number of reads from the same ChIP sample mapped onto the genome (in RPM).

| Antigen | Organism | Clone | Brand | Reference | µg antibody/cond | Assay |
|---------|----------|-------|-------|-----------|------------------|-------|
| H3K27ac | Rabbit | Polyclonal | Abcam | ab4729 | 3 ug | ChIP-qPCR, ChIP-seq |
| H3K27me3 | Mouse | Monoclonal | Abcam | ab6002 | 7.5 ug | ChIP-qPCR |
| H3K9ac | Rabbit | Polyclonal | Abcam | ab4441 | 15 ug | ChIP-qPCR |
| H3K9me3 | Rabbit | Polyclonal | Abcam | ab8898 | 6 ug | ChIP-qPCR |

**Table 4.5 list and specificities of antibodies used in this study.**

The recognized antigen is specified, as well as the organism producing the antibody, the type of clone (poly- or monoclonal), the brand and reference number, the amount of antibody per condition (one condition corresponded to 0.5 million cells). The type of assay (either ChIP qPCR or ChIP-seq) is specified.

### 4.4.5    Inverse PCR identification of the reporter integration site

Genomic DNA was first extracted from 5 million cells of each parental NIH-3T3-FRT clones using GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich). 100 ng of gDNA was then digested for 2 hours at 37°C using 1000 units of BamHI 6-bp cutter restriction enzyme (New England Biolabs). Digested samples were the ligated in diluted conditions (concentration of 2µg of DNA per µl) using T4 DNA Ligase (New England Biolabs). Ligated samples were cleaned as recommended by QIAquick PCR Purification Kit (Qiagen), and inverse primers were used to amplify the unknown genomic region from the flanking integrated vectors fragments of known sequences. 35 PCR cycles were done using the Kapa HiFi HotStart PCR Kit (Kapa Biosystems) and the two iPCR primers iPCR_pFRTneo_R and iPCR_pFRTneo_R (**Table. 4.1**). PCR products were loaded on a 0.8% agarose gel, and single amplification bands were extracted and purified using QIAquick Gel Extraction Kit (Qiagen). Each isolated band was sequenced two times (with both primers used for amplification) using the Sanger approach (GATC Biotech) to obtain the genomic sequence corresponding to the integration site. Genomic Positions were then determined by blating the resulting sequencing on the mm9 version of the UCSC Genome browser (https://genome.ucsc.edu/).

## 4.5 Modeling and computational analysis

### 4.5.1 mRNA half-life estimation from luminescence decay

Estimations were performed by Dr. Benjamin Zoller (Naef lab) as previously explained in (Suter et al. 2011a). Briefly, mRNA half-life was deduced from the following expression for the time-dependent protein abundance $p(t)$, where $p_0$ is the amount of luciferase protein at the time of Actinomycin D treatment, $\gamma_p$ and $\gamma_m$ the protein and mRNA degradation rates respectively, and $k = m_0 k_p$ the translation rate of the luciferase transcripts present in the cell at the time of Actinomycin D treatment:

$$p(t) = \left( p_0 - \frac{k}{\gamma_{\mathrm{p}} - \gamma_{\mathrm{m}}} \right) e^{-\gamma_{\mathrm{p}} t} + \left( \frac{k}{\gamma_{\mathrm{p}} - \gamma_{\mathrm{m}}} \right) e^{-\gamma_{\mathrm{m}} t} \qquad \text{Equation 4.1}$$

The protein degradation rate $\gamma_p$ = 0.032±0.001 min$^{-1}$ (corresponding to a half-life of 21.66±0.68 min) was taken from translation inhibition experiments (**Table 2.1**).

### 4.5.2 Inferring promoter state from single-cell luminescence time-traces

The promoter states and transcriptional bursting parameters were inferred from single-cell luminescence time-traces as previously described (Zoller et al. 2015). Briefly, a extension of the original two state telegraph model of gene expression (Peccoud and Bernard 1995) including a promoter cycle composed of N inactive states was implemented to account for the refractors state often observed between two transcriptionally periods (Harper et al. 2011; Suter et al. 2011a). The number of inactive states best fitting the data was selected using a reverse jump MCMC allowing jumps from one model to another with a different number of N.

The models were parameterized with N+5 constants: the "on" to "off" transition rate $k_a$, the $k1$, …, $kN$ transition rates between all "off" state and the following state, the translation rate $k_p$ and both the protein and mRNA degradation rates $\gamma_p$ and $\gamma_m$ respectively. The three experimentally measured kinetic rates are listed in (**Table 4.6**).

| Kinetic rates | Measured value |
|---|---|
| mRNA degradation rate | $\gamma_m$ = 0.012±0.003 min$^{-1}$ |
| Protein degradation rate | $\gamma_p$ = 0.032±0.001 min$^{-1}$ |
| Translation rate | $k_p$= 1.708±0.018 min$^{-1}$ |

**Table 4.6 Experimentally measured kinetics rate**
The experimentally measured parameters of the telegraph model are the mRNA and protein degradation rate as well as the translation rate.

The likelihood of the luminescence time-traces to arise from the gene expression model is given by the following formula:

$$\mathcal{L}(D|N,\theta_N) = \sum_{\{\Lambda\}} \prod_{i=1}^{L} P_e(s_i|p_i) P_t(p_i m_i g_i | p_{i-1} m_{i-1} g_{i-1}, N, \theta_N) P_e(s_0|p_0) P_s(p_0 m_0 g_0 | N, \theta_N) \qquad \text{Equation 4.2}$$

In this formula, $\vartheta_N$ is the set of kinetic parameters, the first term $P_e(s_i|p_i)$ is the probability of observing the signal ($s_i$) knowing the protein amount ($p_i$) at time $i$, the second term $P_t(p_i m_i g_i|p_i-1 m_i-1 g_i-1,N,\vartheta_N)$ corresponds to the transition probabilities between time $i$ and the previous time $i-1$ for the protein $p$, the transcript $m$ and the gene $g$ given the parameters of the model, and the last term $P_s(p_0 m_0 g_0|N,\vartheta_N)$ corresponds to the initial condition.

To calculate the transition probabilities in (**Equation 4.2**) we used an approximation consisting in fractioning the full transition probability into two factors: first $P_t(p|p'm')$ the transitions corresponding to the protein dynamics:

$$P_t(p|p'm') = \frac{1}{\sqrt{2\pi\sigma_p^2(p',m',\Delta t)}} \exp\left(-\frac{(p-\mu_p(p',m',\Delta t))^2}{2\sigma_p^2(p',m',\Delta t)}\right) \quad \text{Equation 4.3}$$

and second $P_t(mg|m'g')$ the transition probabilities for the mRNA:

$$P_t(mg|m'g') = \exp\left(L\Delta t\right) \quad \text{Equation 4.4}$$

The likelihood is then computed using the following algorithm following (Zoller et al. 2015):

$$F_{i+1}(pmg) = P_e(s_{i+1}|p) \sum_{p'm'g'} P_t(p|p'm')P_t(mg|m'g')F_i(p'm'g') \quad \text{Equation 4.5}$$

For this project, this algorithm was implemented by Dr. Benjamin Zoller (Naef lab).

### 4.5.3 Inferring transcriptional bursting parameters from smRNA-FISH distributions

The telegraph model can be used to derive the steady state for the mRNA distribution (Raj et al. 2006; Dey et al. 2015):

$$\rho(m) = \frac{\Gamma\left(\frac{\lambda}{\delta}+m\right)}{\Gamma\left(m+1\right)\Gamma\left(\frac{\lambda}{\delta}+\frac{\gamma}{\delta}+m\right)}\frac{\Gamma\left(\frac{\lambda}{\delta}+\frac{\gamma}{\delta}\right)}{\Gamma\left(\frac{\lambda}{\delta}\right)}\left(\frac{\mu}{\delta}\right)^m {}_1F_1\left(\frac{\lambda}{\delta}+m,\frac{\lambda}{\delta}+\frac{\gamma}{\delta}+m,-\frac{\mu}{\delta}\right) \quad \text{Equation 4.6}$$

In a regime where the bursts are short compared to the mRNA half-live (large $\gamma/\delta$), the full mRNA distribution can be approximated with a negative binomial distribution:

$$\rho(k) = \left(1+\frac{\mu}{\gamma}\right)^{-\frac{\lambda}{\delta}}\frac{\Gamma\left(\frac{\lambda}{\delta}+k\right)}{\Gamma\left(\frac{\lambda}{\delta}\right)\Gamma\left(k+1\right)}\left(\frac{\frac{\mu}{\gamma}}{1+\frac{\mu}{\gamma}}\right)^k \quad \text{Equation 4.7}$$

This negative binomial distribution only comprises two parameters: the burst frequency ($\lambda/\delta$, gene "on" rate over mRNA decay rate) and the burst size ($\mu/\gamma$, transcription rate over gene "off" rate). To obtain the parameters maximizing the likelihood given the mRNA counts per cell, we used the Maximum likelihood estimation function for a negative binomial in Matlab. This analysis was implemented by Dr. Nick E Phillips (Naef lab). This strategy was applied to infer the bursting

parameters of *Bmal1-sLuc2* and the endogenous circadian genes *Bmal1*, *Cry1* and *Dbp* in NIH-3T3 fibroblasts.

In addition to the maximum likelihood approach, (**Equation 4.7**) can also be approximated to estimate the transcriptional bursting parameters from the mean and variance of the distribution (Weinberger et al. 2012; Dey et al. 2015):

$$\mu = b \cdot f \cdot \tau_m$$
$$CV^2 = \frac{1}{f \cdot \tau_m} \quad \text{Equation 4.8}$$
$$b = \frac{\mu}{f \cdot \tau_m}$$

Here, $\tau_m$ denotes the transcript half-life and is used to calculate the absolute burst frequency (instead of the burst frequency expressed in transcript half-life units). *CV* is the coefficient of variation. This approach was used to determine the bursting parameters for *Bmal1-sLuc2* with a half-life of 59.75±13 minutes (**Table 2.1**), and for 38 genes in mESC (Singer et al. 2014). The half-lives corresponding to the 38 mESC genes were taken from taken from the MC1 LIF+ condition from (Sharova et al. 2009).

### 4.5.4   Public databases of genomic markers

Enrichment of various genomic features at the integration site and around the TSS of genes tested in (Singer et al. 2014) were determined from public genome-wide databases obtained from the GEO platform (https://www.ncbi.nlm.nih.gov/geo/) (**Table 4.6**).

A regression line was fitted to obtain the $R^2$ correlation coefficient between the signal observed in a given window in RPM (typically 0.5, 5, 50 or 500 kb around the TSS), and any bursting parameter inferred for the corresponding gene. To statistically determine the significance of the correlation, F-statistics were applied.

| Genomic Marker | GEO Accession code | Cell line | Original publication |
| --- | --- | --- | --- |
| DHS | GSM1014177 | NIH-3T3 | (Dunham et al. 2012) |
| H3K9me3 | GSM801505 | NIH-3T3 | (Zhu et al. 2012) |
| H3K4me1 | GSM801534 | NIH-3T3 | (Zhu et al. 2012) |
| H3K27ac | GSM801538 | NIH-3T3 | (Zhu et al. 2012) |
| H4ac | GSM801542 | NIH-3T3 | (Zhu et al. 2012) |
| CTCF | GSM879923 | NIH-3T3 | (Zullo et al. 2012) |
| H3K4me3 | GSM879920 | NIH-3T3 | (Zullo et al. 2012) |
| H3K36me3 | GSM879921 | NIH-3T3 | (Zullo et al. 2012) |
| H3K27me3 | GSM879922 | NIH-3T3 | (Zullo et al. 2012) |
| H3K4me1 | GSM1000121 | ES-E14 | (Dunham et al. 2012) |
| H3K4me3 | GSM1000124 | ES-E14 | (Dunham et al. 2012) |
| H3K9ac | GSM1000123 | ES-E14 | (Dunham et al. 2012) |
| H3K27ac | GSM1000126 | ES-E14 | (Dunham et al. 2012) |
| H3K36me3 | GSM1000125 | ES-E14 | (Dunham et al. 2012) |
| H3K9me3 | GSM1003751 | ES-E14 | (Dunham et al. 2012) |
| DNAse | GSM1014154 | ES-E14 | (Dunham et al. 2012) |
| Pol2 | GSM918749 | ES-Bruce4 | (Dunham et al. 2012) |

**Table 4.6 List of assessed genetic markers at integration site**

Public genome-wide enrichment databases of various markers were used to correlate burst size with specific enrichments at the integration site of the reporter or TSS region of endogenous genes. The databases are accessible from GEO (https://www.ncbi.nlm.nih.gov/geo/). The original publication for which data were generated and the cell type are also indicated.

# Chapter 5    Appendix: contribution to publications

In this appendix, I will list my participations to publications outside of my main thesis work. These projects focus on various aspects of transcription, and my contributions typically consisted in the use of molecular tools or approaches originally developed for the study of *Bmal1* transcriptional bursting signature.

For each of these publications, I will present the abstract, briefly explain the content of the publication and mention more precisely my personal participation.

## 5.1    Blanchoud et al. Methods 2015

**CAST: An automated segmentation and tracking tool for the analysis of transcriptional kinetics from single-cell time-lapse recordings**

Simon Blanchoud , <u>Damien Nicolas</u>, Benjamin Zoller, Onur Tidin and Felix Naef

### 5.1.1    Abstract

Fluorescence and bioluminescence time-lapse imaging allows to investigate a vast range of cellular processes at single-cell or even subcellular resolution. In particular, time-lapse imaging can provide uniquely detailed information on the fine kinetics of transcription, as well as on biological oscillations such as the circadian and cell cycles. However, we face a paucity of automated methods to quantify time-lapse imaging data with single-cell precision, notably throughout multiple cell cycles. We developed CAST (Cell Automated Segmentation and Tracking platform) to automatically and robustly detect the position and size of cells or nuclei, quantify the corresponding light signals, while taking into account both cell divisions (lineage tracking) and migration events. We present here how CAST analyzes bioluminescence data from a short-lived transcriptional luciferase reporter. However, our flexible and modular implementation makes it easily adaptable to a wide variety of time-lapse recordings. We exemplify how CAST efficiently quantifies single-cell gene expression over multiple cell cycles using mouse NIH3T3 culture cells with a luminescence expression driven by the *Bmal1* promoter, a central gene of the circadian oscillator. We further illustrate how such data can be used to quantify transcriptional bursting in conditions of lengthened circadian period, revealing thereby remarkably similar bursting signature compared to the endogenous circadian condition despite marked period lengthening. In summary,

we establish CAST as novel tool for the efficient segmentation, signal quantification, and tracking of time-lapse images from mammalian cell culture.

## 5.1.2 Contribution to the publication

This publication describes the development of CAST, a platform for the segmentation and tracking of real-time single-cell luminescence data. Although this tool is modular and can be adapted to various purposes, it is particularly adapted to track destabilized reporters such as the short-lived luciferase we use to monitor transcriptional bursting. Indeed, CAST excels at segmenting cells displaying low signal-to-noise ratio. Also, it can infer the trajectory of cells between two bursts, even when the signal diminishes to background level.
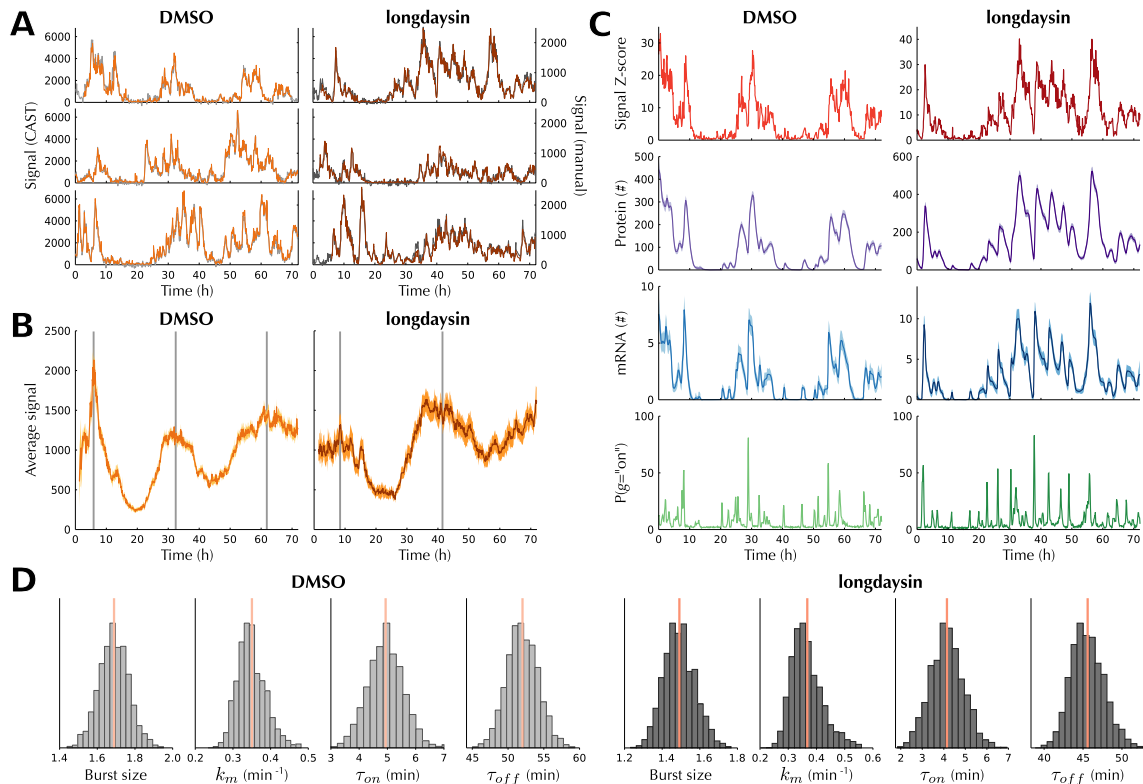


**Figure 5.1 Comparison between the transcriptional bursting pattern of *Bmal1* in control and longdaysin conditions**
(**A**) Superposition of three typical Bmal1 single-cell traces obtained by CAST (brown lines, left axis) and manual tracking (gray lines, right axis) in both endogenous (DMSO, left column) and lengthened circadian period (red, right column). (**B**) Average *Bmal1* signal displayed by individual cells tracked with CAST, both in control condition (orange, n = 127) and in 4 μM longdaysin condition (brown, n = 108), recapitulating the phenotype of a population of cells. Orange areas delimit the standard error and vertical grey bars estimate the position of expression peaks (**C**) Example of deconvolved single-cell traces obtained from our mathematical modeling approach (control condition on the lighter left column, and longdaysin 4 μM on the darker right one). From the raw luminescence signal (first row in panel A, corresponding Z-score in red), we infer the amount of luciferase proteins (purple) and transcripts (blue), as well as the probability of gene activity (green). (**D**) Transcriptional bursting kinetics of *Bmal1* in endogenous (gray) and lengthened (black) circadian conditions. Displayed kinetics are the mean number of transcripts produced per burst (burst size), the transcription rate ($k_m$), as well as the mean time spend in active ($\tau_{on}$) and inactive states ($\tau_{off}$). Figure from (Blanchoud et al. 2015).

To illustrate the capacities of CAST, it was used to segment and track NIH-3T3 luminescent cells I recorded using a luminescence microscope (**Figure 5.1A**). These recordings were generated with a previous version of the *Bmal1* short-lived luciferase reporter (Suter et al. 2011a) monitored in two conditions: endogenous rhythmicity, or lengthened circadian period using longdaysin at a final concentration of 4µM (**Figure 5.1B**) (Hirota et al. 2010b). Individual traces were then used to infer the transcriptional bursting parameters specific to both conditions (**Figure 5.1C**). This analysis was performed by Dr. Benjamin Zoller (Naef lab) using the entire 72-hours time-traces as a single input without taking into account circadian variations in transcription. No major differences in bursting parameters could be detected between the two conditions (**Figure 5.1D**). Thus, longdaysin did not affect the transcriptional bursting characteristics of *Bmal1*, but rather uniformly stretched its rhythmic expression profile.


## 5.2     Zoller et al. Mol Syst Biol 2015

**Structure of silent transcription intervals and noise characteristics of mammalian genes**
Benjamin Zoller, Damien Nicolas, Nacho Molina and Felix Naef

### 5.2.1   Abstract

Mammalian transcription occurs stochastically in short bursts interspersed by silent intervals showing a refractory period. However, the underlying processes and consequences on fluctuations in gene products are poorly understood. Here, we use single allele time-lapse recordings in mouse cells to identify minimal models of promoter cycles, which inform on the number and durations of rate-limiting steps responsible for refractory periods. The structure of promoter cycles is gene specific and independent of genomic location. Typically, five rate-limiting steps underlie the silent periods of endogenous promoters, while minimal synthetic promoters exhibit only one. Strikingly, endogenous or synthetic promoters with TATA boxes show simplified two-state promoter cycles. Since transcriptional bursting constrains intrinsic noise depending on the number of promoter steps, this explains why TATA box genes display increased intrinsic noise genome-wide in mammals, as revealed by single-cell RNA-seq. These findings have implications for basic transcription biology and shed light on interpreting single-cell RNA-counting experiments.

### 5.2.2   Contribution to the publication

Following the identification of a refractory period preventing the reactivation of some promoters subsequent to a burst of transcription, this work reanalyzed luminescence traces from endogenous mouse genes or synthetic promoters (Suter et al. 2011a) to clarify the number of rate-limiting steps composing this refractory period.

In addition to the luminescence traces from the original study, I personally recorded additional conditions to increase the number of available data. Mainly, I generated H1b and H1c, two cell-lines expressing the short-lived luciferase from a H1 synthetic promoter stably integrated into different genomic FRT sites. The two FRT clones used for that purpose were the same than the clones 4 and 12 of the homemade NIH-3T3-FRT clones used to stably integrate the *Bmal1-sLuc2*

reporter (**Figure 2.7**). The "off" state of these two clones was composed of a single step (**Figure 5.2**). Thus, they did not undergo refractoriness after transcriptional events. Thus, the absence of refractory period already observed in the H1a clone was not caused by its integration site. The number and duration of the rate-limiting steps composing the refractory period were more likely determined by the promoter architecture since most genes like H1 containing a TATA-box in their promoter, in addition to displaying elevated expression noise, only had a single "off" state enabling them to retranscribe directly after a subsequent burst.
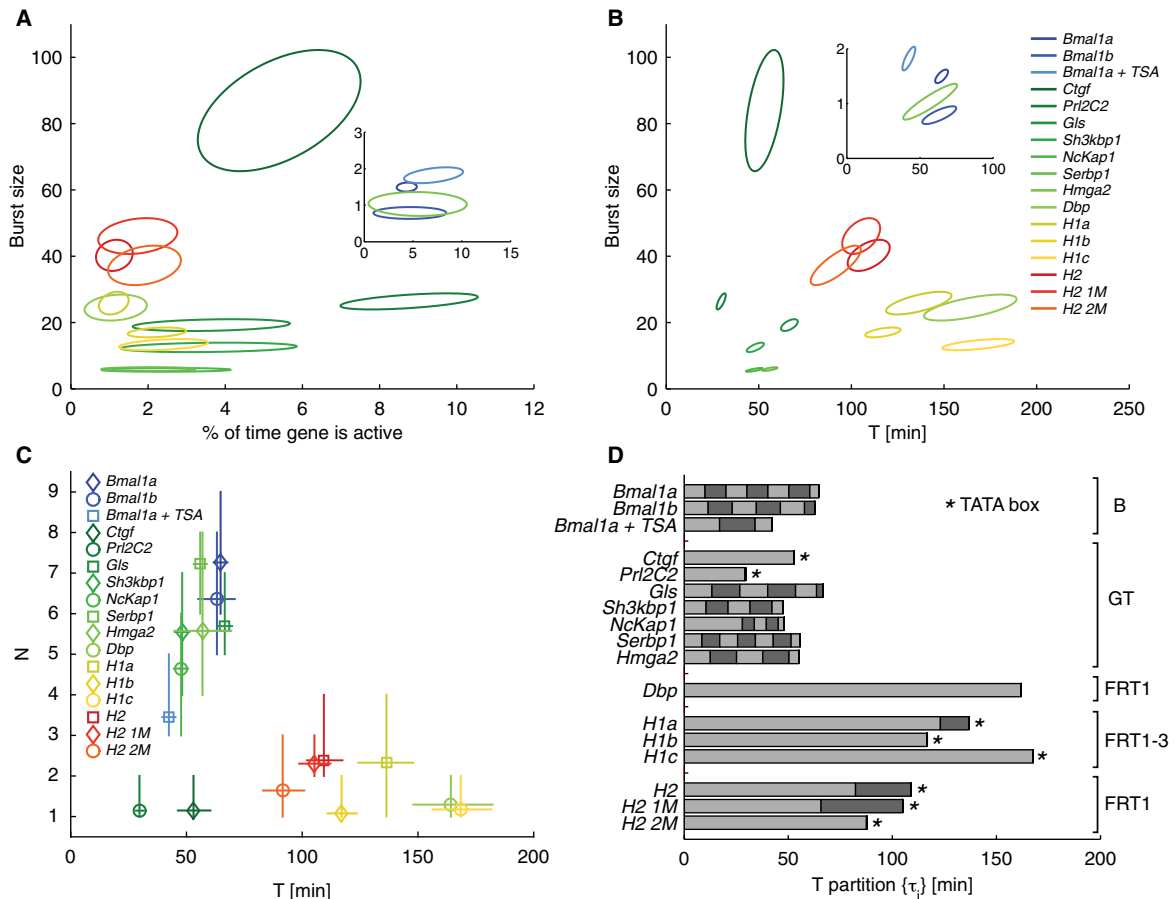


**Structure and kinetics of the promoter cycles for the NIH-3T3 clones**

**(A)** Burst size vs. the fraction of time the gene is active. Each clone is represented by a 95% confidence ellipse from the posterior distribution. All the analyzed clones burst, characterized by small activity fractions. Burst sizes show a large dynamic range across clones (~80-fold). Inset: Magnification of the lower left corner. **(B)** Burst size vs. the total silent period $T$. Elongated confidence ellipses reflect the dependence between those two quantities and the mean mRNA. Although the dynamic range of the silent period (~6-fold) is smaller than for the burst size, it is also gene specific. The synthetic (warm colors) and endogenous (cold colors) promoters cluster in distinct regions. **(C)** Number of inactive states vs. $T$, crosses indicate mean and error bars stand for the 5th and 95th percentiles of the posterior. Endogenous promoters tend to show more inactive steps and shorter cycle times (cluster around $N$~6 and $T$~60 min) compared to synthetic promoters (cluster around $N$~1–2 and $T$~130 min). **(D)** Partitioning of the silent period for the optimal models. The light and dark bars show the mean durations of each sub-step. Partitions in endogenous promoters tend to be more uniform compared to the synthetic promoters. Average inactive times for endogenous promoter are around 10 min, whereas synthetic promoters have average inactive times close to 100 min (~115 min for the first and ~25 min for the subsequent intervals). Figure from (Zoller et al. 2015).

Interestingly, despite little variations in their global expression levels, difference in burst sizes could notably be spotted between the three H1 clones, suggesting that the integration site did affect this bursting parameter.

## 5.3 Nicolas et al. Mol Biosyst 2017

**What shapes eukaryotic transcriptional bursting?**
Damien Nicolas, Nick E. Phillips and Felix Naef

### 5.3.1 Abstract

Isogenic cells in a common environment present a large degree of heterogeneity in gene expression. Part of this variability is attributed to transcriptional bursting: the stochastic activation and inactivation of promoters that leads to the discontinuous production of mRNA. The diversity in bursting patterns displayed by different genes suggests the existence of a connection between bursting and gene regulation. Experimental strategies such as single-molecule RNA FISH, MS2-GFP or short-lived protein reporters allow the quantification of transcriptional bursting and the comparison of bursting kinetics between conditions, allowing therefore the identification of molecular mechanisms modulating transcriptional bursting. In this review we recapitulate the impact on transcriptional bursting of different molecular aspects of transcription such as the chromatin environment, nucleosome occupancy, histone modifications, the number and affinity of regulatory elements, DNA looping and transcription factor availability. More specifically, we examine their role in tuning the burst size or the burst frequency. While some molecular mechanisms involved in transcription such as histone marks can affect every aspect of bursting, others predominantly influence the burst size (e.g. the number and affinity of cis-regulatory elements) or frequency (e.g. transcription factor availability).

### 5.3.2 Contribution to the publication

This review recapitulates the concept of transcriptional bursting and the most common experimental approaches allowing its monitoring. It also lists the recent comparative studies that improved our understanding of how molecular mechanisms influence the burst size or frequency. I wrote most of the core text with the exception of the paragraph on computational approaches commonly used to infer bursting parameters from real-time single-cell traces or RNA distributions, written by Dr. Nick E Phillips (Naef lab). I also designed the figures and indexed the elements of the table.

## 5.4 Mermet et al. submitted 2017

**Clock-dependent chromatin topology modulates circadian transcription and behavior**
Jérôme Mermet, Jake Yeung, Clémence Hurni, Daniel Mauvoisin, Kyle Gustafson, Céline Jouffe, Damien Nicolas, Yann Emmenegger, Cédric Gobet, Paul Franken, Frédéric Gachon, and Felix Naef

## 5.4.1 Abstract

The circadian oscillator provides a powerful model system to study temporal dynamics in transcription regulation. Several studies showed the possible roles of transcription factors and chromatin marks in controlling cyclic gene expression. However, how daily active enhancers modulate rhythmic gene transcription in mammalian tissues remains uncharted. Here, we discovered oscillatory promoter-enhancer interactions along the 24-hour cycle in mouse liver and kidney, and these oscillations depended on the clock transcription factor BMAL1. Deleting a contacted intronic enhancer element in the *Cry1* gene was sufficient to compromise the rhythmic chromatin contacts in tissues. Moreover, the deletion reduced *Cry1* transcriptional burst frequency and shortened the circadian period of locomotor activity rhythms. Our results establish oscillating and clock-controlled promoter-enhancer looping as a new regulatory layer underlying circadian transcription and behavior.

## 5.4.2 Contribution to the publication

This paper recently submitted to Science uses the 4C approach to assess the role of chromatin contacts in regulating circadian genes expression in mouse tissues. This study notably highlighted the existence of intragenic regions in *Cry1* and *Gys2* that dynamically contact the promoter over the circadian period. Presence of contacts is consistent with the genes expression levels at different circadian phases, in different tissues and in absence of circadian clock.
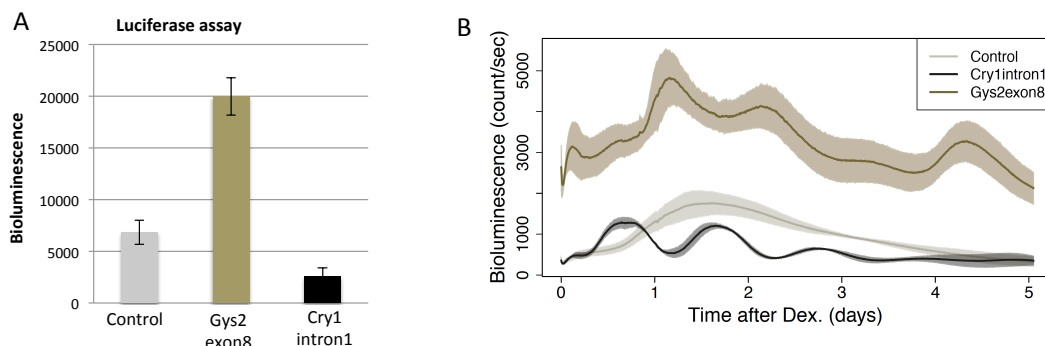


**Figure 5.3 *Cry1* intron 1 and *Gys2* exon 8 elements are circadian transcriptional regulators**
**(A)** Luciferase activity measured 36 hours after transfection of NIH-3T3 cells with a plasmid expressing the luciferase under the control of a minimal promoter (control, grey) supplemented with the *Gys2* exon 8 (brown) or the *Cry1* intron 1 (black) **(B)** Luminescence signal monitored in DEX-synchronized NIH-3T3 expressing the same constructs stably integrated into a genomic FRT site. Results are displayed in mean ± SD over 3 replicates.

To assess the enhancer properties of these intragenic regions, I designed an expression vector composed of a Gateway cloning cassette, a minimal TATA-box promoter, a luciferase CDS and an FRT-site. The two putative enhancers (*Cry1* intron 1 and *Gys2* exon 8) were inserted in the Gateway cassette to regulate the luminescence expression. The arrhythmic minimal promoter drove luciferase expression in absence of enhancers (control). With both transiently transfected and stably integrated reporters, the *Gys2* exon 8 increased the luciferase expression by ~3-folds, while the *Cry1* intron 1 region decreased it by ~2-fold (**Figure 5.3**). The presence of the two

intragenic regions conferred rhythmicity to the luminescence expression. Thus, this *in vitro* approach highlighted the enhancer effect of *Gys2* exon 8 and the silencer effect of *Cry1* intron 1 in NIH-3T3 cells, as well as their role in driving circadian oscillations. I also participated in the design and generation of tools to deplete *Cry1* intronic enhancer using the CRISPR/Cas9 targeted genome editing approach.

# References

Adelman K, Lis JT (2012) Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat Rev Genet 13:720–731. doi: 10.1038/nrg3293

Adelmant G, Bègue A, Stéhelin D, Laudet V (1996) A functional Rev-erb alpha responsive element located in the human Rev-erb alpha promoter mediates a repressing activity. Proc Natl Acad Sci U S A 93:3553–3558.

Aguilar-Arnal L, Hakim O, Patel VR, et al (2013) Cycles in spatial and temporal chromosomal organization driven by the circadian clock. Nat Struct Mol Biol 20:1206–1213. doi: 10.1038/nsmb.2667

Akashi M, Takumi T (2005) The orphan nuclear receptor ROR? regulates circadian transcription of the mammalian core-clock Bmal1. Nat Struct Mol Biol 12:441–448. doi: 10.1038/nsmb925

Albayrak C, Jordi CA, Zechner C, et al (2016) Digital Quantification of Proteins and mRNA in Single Mammalian Cells. Mol Cell 61:914–924. doi: 10.1016/j.molcel.2016.02.030

Albert I, Mavrich TN, Tomsho LP, et al (2007) Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. Nature 446:572–576. doi: 10.1038/nature05632

Allen BL, Taatjes DJ (2015) The Mediator complex: a central integrator of transcription. Nat Rev Mol Cell Biol 16:155–166. doi: 10.1038/nrm3951

André E, Gawlas K, Becker-André M (1998) A novel isoform of the orphan nuclear receptor RORbeta is specifically expressed in pineal gland and retina. Gene 216:277–283.

Andrews FH, Strahl BD, Kutateladze TG (2016) Insights into newly discovered marks and readers of epigenetic information. Nat Chem Biol 12:662–668. doi: 10.1038/nchembio.2149

Antolović V, Miermont A, Corrigan AM, Chubb JR (2017) Generation of Single-Cell Transcript Variability by Repression. Curr Biol 27:1811–1817.e3. doi: 10.1016/j.cub.2017.05.028

Atger F, Gobet C, Marquis J, et al (2015) Circadian and feeding rhythms differentially affect rhythmic mRNA transcription and translation in mouse liver. Proc Natl Acad Sci 112:E6579–E6588. doi: 10.1073/pnas.1515308112

Atkins GB, Hu X, Guenther MG, et al (1999) Coactivators for the Orphan Nuclear Receptor RORα. Mol Endocrinol 13:1550–1557. doi: 10.1210/mend.13.9.0343

Azzaz AM, Vitalini MW, Thomas AS, et al (2014) Human heterochromatin protein 1α promotes nucleosome associations that drive chromatin condensation. J Biol Chem 289:6850–6861. doi: 10.1074/jbc.M113.512137

Baggs JE, Price TS, DiTacchio L, et al (2009) Network features of the mammalian circadian clock. PLoS Biol 7:e52. doi: 10.1371/journal.pbio.1000052

Bahar Halpern K, Caspi I, Lemze D, et al (2015a) Nuclear Retention of mRNA in Mammalian Tissues. Cell Rep 13:2653–2662. doi: 10.1016/j.celrep.2015.11.036

Bahar Halpern K, Tanami S, Landen S, et al (2015b) Bursty gene expression in the intact mammalian liver. Mol Cell 58:147–156. doi: 10.1016/j.molcel.2015.01.027

Balaban NQ, Merrin J, Chait R, et al (2004) Bacterial persistence as a phenotypic switch. Science 305:1622–1625. doi: 10.1126/science.1099390

Balsalobre A, Damiola F, Schibler U (1998) A serum shock induces circadian gene expression in mammalian tissue culture cells. Cell 93:929–937.

Barreau C (2005) AU-rich elements and associated factors: are there unifying principles? Nucleic Acids Res 33:7138–7150. doi: 10.1093/nar/gki1012

Barski A, Cuddapah S, Cui K, et al (2007) High-resolution profiling of histone methylations in the human genome. Cell 129:823–837. doi: 10.1016/j.cell.2007.05.009

Bartman CR, Hsu SC, Hsiung CC-S, et al (2016) Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. Mol Cell 62:237–247. doi: 10.1016/j.molcel.2016.03.007

Batenchuk C, St-Pierre S, Tepliakova L, et al (2011) Chromosomal position effects are linked to sir2-mediated variation in transcriptional burst size. Biophys J 100:L56-58. doi: 10.1016/j.bpj.2011.04.021

Battich N, Stoeger T, Pelkmans L (2015) Control of Transcript Variability in Single Mammalian Cells. Cell 163:1596–1610. doi: 10.1016/j.cell.2015.11.018

Battich N, Stoeger T, Pelkmans L (2013) Image-based transcriptomics in thousands of single human cells at single-molecule resolution. Nat Methods 10:1127–1133. doi: 10.1038/nmeth.2657

Beagrie RA, Pombo A (2016) Gene activation by metazoan enhancers: Diverse mechanisms stimulate distinct steps of transcription. BioEssays 38:881–893. doi: 10.1002/bies.201600032

Becker-Weimann S, Wolf J, Herzel H, Kramer A (2004) Modeling Feedback Loops of the Mammalian Circadian Oscillator. Biophys J 87:3023–3034. doi: 10.1529/biophysj.104.040824

Becskei A, Kaufmann BB, van Oudenaarden A (2005) Contributions of low molecule number and chromosomal positioning to stochastic gene expression. Nat Genet 37:937–944. doi: 10.1038/ng1616

Bell-Pedersen D, Cassone VM, Earnest DJ, et al (2005) Circadian rhythms from multiple oscillators: lessons from diverse organisms. Nat Rev Genet 6:544–556. doi: 10.1038/nrg1633

Bengtsson M, Ståhlberg A, Rorsman P, Kubista M (2005) Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. Genome Res 15:1388–1392. doi: 10.1101/gr.3820805

Berguet G, Hendrickx J, Sabatel C, et al (2014) Automating ChIP-seq Experiments to Generate Epigenetic Profiles on 10,000 HeLa Cells. J Vis Exp. doi: 10.3791/52150

Bernstein BE, Kamal M, Lindblad-Toh K, et al (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. Cell 120:169–181. doi: 10.1016/j.cell.2005.01.001

Bernstein BE, Mikkelsen TS, Xie X, et al (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125:315–326. doi: 10.1016/j.cell.2006.02.041

Bieler J, Cannavo R, Gustafson K, et al (2014) Robust synchronization of coupled circadian and cell cycle oscillators in single mammalian cells. Mol Syst Biol 10:739–739. doi: 10.15252/msb.20145218

Blake WJ, KAErn M, Cantor CR, Collins JJ (2003) Noise in eukaryotic gene expression. Nature 422:633–637. doi: 10.1038/nature01546

Blanchoud S, Nicolas D, Zoller B, et al (2015) CAST: An automated segmentation and tracking tool for the analysis of transcriptional kinetics from single-cell time-lapse recordings. Methods San Diego Calif 85:3–11. doi: 10.1016/j.ymeth.2015.04.023

Bódi Z, Farkas Z, Nevozhay D, et al (2017) Phenotypic heterogeneity promotes adaptive evolution. PLoS Biol 15:e2000644. doi: 10.1371/journal.pbio.2000644

Boettiger AN, Levine M (2009) Synchronous and stochastic patterns of gene activation in the Drosophila embryo. Science 325:471–473. doi: 10.1126/science.1173976

Boettiger AN, Levine M (2013) Rapid transcription fosters coordinate snail expression in the Drosophila embryo. Cell Rep 3:8–15. doi: 10.1016/j.celrep.2012.12.015

Boireau S, Maiuri P, Basyuk E, et al (2007) The transcriptional cycle of HIV-1 in real-time and live cells. J Cell Biol 179:291–304. doi: 10.1083/jcb.200706018

Breaker RR, Joyce GF (2014) The expanding view of RNA and DNA function. Chem Biol 21:1059–1065. doi: 10.1016/j.chembiol.2014.07.008

Bronstein L, Zechner C, Koeppl H (2015) Bayesian inference of reaction kinetics from single-cell recordings across a heterogeneous cell population. Methods San Diego Calif 85:22–35. doi: 10.1016/j.ymeth.2015.05.012

Brown CR, Mao C, Falkovskaia E, et al (2013) Linking stochastic fluctuations in chromatin structure and gene expression. PLoS Biol 11:e1001621. doi: 10.1371/journal.pbio.1001621

Bunger MK, Wilsbacher LD, Moran SM, et al (2000) Mop3 is an essential component of the master circadian pacemaker in mammals. Cell 103:1009–1017.

Burnett JC, Miller-Jensen K, Shah PS, et al (2009) Control of stochastic gene expression by host factors at the HIV promoter. PLoS Pathog 5:e1000260. doi: 10.1371/journal.ppat.1000260

Caro-Maldonado A, Muoz-Pinedo C (2011) Dying for Something to Eat: How Cells Respond to Starvation. Open Cell Signal J 3:42–51. doi: 10.2174/1876390101103010042

Carpenter AE, Jones TR, Lamprecht MR, et al (2006) CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol 7:R100. doi: 10.1186/gb-2006-7-10-r100

Carrozza MJ, Li B, Florens L, et al (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell 123:581–592. doi: 10.1016/j.cell.2005.10.023

Chang HH, Hemberg M, Barahona M, et al (2008) Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature 453:544–547. doi: 10.1038/nature06965

Chavez A, Tuttle M, Pruitt BW, et al (2016) Comparison of Cas9 activators in multiple species. Nat Methods 13:563–567. doi: 10.1038/nmeth.3871

Chen H, Chen J, Muir LA, et al (2015a) Functional organization of the human 4D Nucleome. Proc Natl Acad Sci 112:8002–8007. doi: 10.1073/pnas.1505822112

Chen H, Fujioka M, Jaynes JB, Gregor T (2017) Direct visualization of transcriptional activation by physical enhancer-promoter proximity. doi: 10.1101/099523

Chen H, Larson DR (2016) What have single-molecule studies taught us about gene expression? Genes Dev 30:1796–1810. doi: 10.1101/gad.281725.116

Chen J, Zhang Z, Li L, et al (2014) Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. Cell 156:1274–1285. doi: 10.1016/j.cell.2014.01.062

Chen KH, Boettiger AN, Moffitt JR, et al (2015b) RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science 348:aaa6090. doi: 10.1126/science.aaa6090

Cheng J, Blum R, Bowman C, et al (2014) A role for H3K4 monomethylation in gene repression and partitioning of chromatin readers. Mol Cell 53:979–992. doi: 10.1016/j.molcel.2014.02.032

Cheon S, Park N, Cho S, Kim K (2013) Glucocorticoid-mediated Period2 induction delays the phase of circadian rhythm. Nucleic Acids Res 41:6161–6174. doi: 10.1093/nar/gkt307

Cherstvy AG, Kolomeisky AB, Kornyshev AA (2008) Protein–DNA Interactions: Reaching and Recognizing the Targets. J Phys Chem B 112:4741–4750. doi: 10.1021/jp076432e

Cho H, Zhao X, Hatori M, et al (2012) Regulation of circadian behaviour and metabolism by REV-ERB-$\alpha$ and REV-ERB-$\beta$. Nature 485:123–7. doi: 10.1038/nature11048

Cho W-K, Jayanth N, English BP, et al (2016) RNA Polymerase II cluster dynamics predict mRNA output in living cells. eLife. doi: 10.7554/eLife.13617

Chong S, Chen C, Ge H, Xie XS (2014) Mechanism of transcriptional bursting in bacteria. Cell 158:314–326. doi: 10.1016/j.cell.2014.05.038

Chubb JR, Trcek T, Shenoy SM, Singer RH (2006) Transcriptional pulsing of a developmental gene. Curr Biol CB 16:1018–1025. doi: 10.1016/j.cub.2006.03.092

Churchman LS, Weissman JS (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 469:368–373. doi: 10.1038/nature09652

Cisse II, Izeddin I, Causse SZ, et al (2013) Real-time dynamics of RNA polymerase II clustering in live human cells. Science 341:664–667. doi: 10.1126/science.1239053

Core LJ, Waterfall JJ, Lis JT (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322:1845–1848. doi: 10.1126/science.1162228

Corrales-Berjano M, Rosado A, Cortini R, et al (2017) Clustering of Drosophila housekeeping promoters facilitates their expression. Genome Res. doi: 10.1101/gr.211433.116

Corrigan AM, Tunnacliffe E, Cannon D, Chubb JR (2016) A continuum model of transcriptional bursting. eLife. doi: 10.7554/eLife.13051

Coulon A, Chow CC, Singer RH, Larson DR (2013) Eukaryotic transcriptional dynamics: from single molecules to cell populations. Nat Rev Genet 14:572–584. doi: 10.1038/nrg3484

Creyghton MP, Cheng AW, Welstead GG, et al (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci 107:21931–21936. doi: 10.1073/pnas.1016071107

Crumbley C, Wang Y, Kojetin DJ, Burris TP (2010) Characterization of the core mammalian clock component, NPAS2, as a REV-ERBalpha/RORalpha target gene. J Biol Chem 285:35386–35392. doi: 10.1074/jbc.M110.129288

Dacheux E, Malys N, Meng X, et al (2017) Translation initiation events on structured eukaryotic mRNAs generate gene expression noise. Nucleic Acids Res. doi: 10.1093/nar/gkx430

Dadiani M, van Dijk D, Segal B, et al (2013) Two DNA-encoded strategies for increasing expression with opposing effects on promoter dynamics and transcriptional noise. Genome Res 23:966–976. doi: 10.1101/gr.149096.112

Damiola F, Le Minh N, Preitner N, et al (2000) Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. Genes Dev 14:2950–2961.

Dar RD, Hosmane NN, Arkin MR, et al (2014) Screening for noise in gene expression identifies drug synergies. Science 344:1392–1396. doi: 10.1126/science.1250220

Dar RD, Razooky BS, Singh A, et al (2012) Transcriptional burst frequency and burst size are equally modulated across the human genome. Proc Natl Acad Sci U S A 109:17454–17459. doi: 10.1073/pnas.1213530109

Dar RD, Shaffer SM, Singh A, et al (2016) Transcriptional Bursting Explains the Noise-Versus-Mean Relationship in mRNA and Protein Levels. PloS One 11:e0158298. doi: 10.1371/journal.pone.0158298

David FPA, Delafontaine J, Carat S, et al (2014) HTSstation: A Web Application and Open-Access Libraries for High-Throughput Sequencing Data Analysis. PLoS ONE 9:e85879. doi: 10.1371/journal.pone.0085879

Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. Genes Dev 25:1010–1022. doi: 10.1101/gad.2037511

Deckert J, Struhl K (2001) Histone Acetylation at Promoters Is Differentially Affected by Specific Activators and Repressors. Mol Cell Biol 21:2726–2735. doi: 10.1128/MCB.21.8.2726-2735.2001

Dey SS, Foley JE, Limsirichai P, et al (2015) Orthogonal control of expression mean and variance by epigenetic features at different genomic loci. Mol Syst Biol 11:806.

Di Cerbo V, Mohn F, Ryan DP, et al (2014) Acetylation of histone H3 at lysine 64 regulates nucleosome dynamics and facilitates transcription. eLife. doi: 10.7554/eLife.01632

Di Croce L, Helin K (2013) Transcriptional regulation by Polycomb group proteins. Nat Struct Mol Biol 20:1147–1155. doi: 10.1038/nsmb.2669

Dibner C, Schibler U, Albrecht U (2010) The mammalian circadian timing system: organization and coordination of central and peripheral clocks. Annu Rev Physiol 72:517–549. doi: 10.1146/annurev-physiol-021909-135821

DiTacchio L, Le HD, Vollmers C, et al (2011) Histone lysine demethylase JARID1a activates CLOCK-BMAL1 and influences the circadian clock. Science 333:1881–1885. doi: 10.1126/science.1206022

Dobin A, Davis CA, Schlesinger F, et al (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. doi: 10.1093/bioinformatics/bts635

Dobrzynski M, Bruggeman FJ (2009) Elongation dynamics shape bursty transcription and translation. Proc Natl Acad Sci 106:2583–2588. doi: 10.1073/pnas.0803507106

Doi M, Hirayama J, Sassone-Corsi P (2006) Circadian regulator CLOCK is a histone acetyltransferase. Cell 125:497–508. doi: 10.1016/j.cell.2006.03.033

Du N-H, Arpat AB, De Matos M, Gatfield D (2014) MicroRNAs shape circadian hepatic gene expression on a transcriptome-wide scale. eLife. doi: 10.7554/eLife.02510

Duan J, Shi J, Ge X, et al (2013) Genome-wide survey of interindividual differences of RNA stability in human lymphoblastoid cell lines. Sci Rep 3:1318. doi: 10.1038/srep01318

Dunham I, Kundaje A, Aldred SF, et al (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74. doi: 10.1038/nature11247

Duong HA, Robles MS, Knutti D, Weitz CJ (2011) A molecular mechanism for circadian clock negative feedback. Science 332:1436–1439. doi: 10.1126/science.1196766

Eckel-Mahan KL, Patel VR, Mohney RP, et al (2012) Coordination of the transcriptome and metabolome by the circadian clock. Proc Natl Acad Sci U S A 109:5541–5546. doi: 10.1073/pnas.1118726109

Eide EJ, Woolf MF, Kang H, et al (2005) Control of mammalian circadian rhythm by CKIepsilon-regulated proteasome-mediated PER2 degradation. Mol Cell Biol 25:2795–2807. doi: 10.1128/MCB.25.7.2795-2807.2005

Eldar A, Elowitz MB (2010) Functional roles for noise in genetic circuits. Nature 467:167–173. doi: 10.1038/nature09326

Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. Science 297:1183–1186. doi: 10.1126/science.1070919

Ernst J, Kheradpour P, Mikkelsen TS, et al (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473:43–49. doi: 10.1038/nature09906

Etchegaray J-P, Lee C, Wade PA, Reppert SM (2003) Rhythmic histone acetylation underlies transcription in the mammalian circadian clock. Nature 421:177–182. doi: 10.1038/nature01314

Etchegaray J-P, Yang X, DeBruyne JP, et al (2006) The polycomb group protein EZH2 is required for mammalian circadian clock function. J Biol Chem 281:21209–21215. doi: 10.1074/jbc.M603722200

Everett LJ, Lazar MA (2014) Nuclear receptor Rev-erb?: up, down, and all around. Trends Endocrinol Metab 25:586–592. doi: 10.1016/j.tem.2014.06.011

Ezer D, Moignard V, Göttgens B, Adryan B (2016) Determining Physical Mechanisms of Gene Expression Regulation from Single Cell Gene Expression Data. PLoS Comput Biol 12:e1005072. doi: 10.1371/journal.pcbi.1005072

Fang B, Everett LJ, Jager J, et al (2014) Circadian Enhancers Coordinate Multiple Phases of Rhythmic Gene Transcription In Vivo. Cell 159:1140–1152. doi: 10.1016/j.cell.2014.10.022

Featherstone K, Hey K, Momiji H, et al (2016) Spatially coordinated dynamic gene transcription in living pituitary tissue. eLife 5:e08494. doi: 10.7554/eLife.08494

Femino AM, Fay FS, Fogarty K, Singer RH (1998) Visualization of single RNA transcripts in situ. Science 280:585–590.

Feng D, Liu T, Sun Z, et al (2011) A circadian rhythm orchestrated by histone deacetylase 3 controls hepatic lipid metabolism. Science 331:1315–1319. doi: 10.1126/science.1198125

Filion GJ, van Bemmel JG, Braunschweig U, et al (2010) Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell 143:212–224. doi: 10.1016/j.cell.2010.09.009

Forman BM, Chen J, Blumberg B, et al (1994) Cross-talk among ROR alpha 1 and the Rev-erb family of orphan nuclear receptors. Mol Endocrinol 8:1253–1261. doi: 10.1210/mend.8.9.7838158

Francis NJ, Kingston RE, Woodcock CL (2004) Chromatin compaction by a polycomb group protein complex. Science 306:1574–1577. doi: 10.1126/science.1100576

Fuda NJ, Ardehali MB, Lis JT (2009) Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature 461:186–192. doi: 10.1038/nature08449

Fujita K, Iwaki M, Yanagida T (2016) Transcriptional bursting is intrinsically caused by interplay between RNA polymerases on DNA. Nat Commun 7:13788. doi: 10.1038/ncomms13788

Fukaya T, Lim B, Levine M (2016) Enhancer Control of Transcriptional Bursting. Cell 166:358–368. doi: 10.1016/j.cell.2016.05.025

Fustin JM, O'Neill JS, Hastings MH, et al (2009) Cry1 Circadian Phase in vitro: Wrapped Up with an E-Box. J Biol Rhythms 24:16–24. doi: 10.1177/0748730408329267

García-Alai MM, Gallo M, Salame M, et al (2006) Molecular Basis for Phosphorylation-Dependent, PEST-Mediated Protein Turnover. Structure 14:309–319. doi: 10.1016/j.str.2005.11.012

Gebhardt JCM, Suter DM, Roy R, et al (2013) Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. Nat Methods 10:421–426. doi: 10.1038/nmeth.2411

Gekakis N, Staknis D, Nguyen H, et al (1998) Role of the CLOCK Protein in the Mammalian Circadian Mechanism. Science 280:1564–1569. doi: 10.1126/science.280.5369.1564

Gerber A, Esnault C, Aubert G, et al (2013) Blood-borne circadian signal stimulates daily oscillations in actin dynamics and SRF activity. Cell 152:492–503. doi: 10.1016/j.cell.2012.12.027

Giguère V, McBroom LD, Flock G (1995) Determinants of target gene specificity for ROR alpha 1: monomeric DNA binding by an orphan nuclear receptor. Mol Cell Biol 15:2517–2526.

Golding I, Paulsson J, Zawilski SM, Cox EC (2005) Real-time kinetics of gene activity in individual bacteria. Cell 123:1025–1036. doi: 10.1016/j.cell.2005.09.031

Gonze D, Halloy J, Leloup J-C, Goldbeter A (2003) Stochastic models for circadian rhythms: effect of molecular noise on periodic and chaotic behaviour. C R Biol 326:189–203. doi: 10.1016/S1631-0691(03)00016-7

Guillaumond F, Dardente H, Gigu?re V, Cermakian N (2005) Differential Control of Bmal1 Circadian Transcription by REV-ERB and ROR Nuclear Receptors. J Biol Rhythms 20:391–403. doi: 10.1177/0748730405277232

Gut G, Tadmor MD, Pe'er D, et al (2015) Trajectories of cell-cycle progression from fixed cell populations. Nat Methods 12:951–954. doi: 10.1038/nmeth.3545

Harding HP, Atkins GB, Jaffe AB, et al (1997) Transcriptional activation and repression by RORalpha, an orphan nuclear receptor required for cerebellar development. Mol Endocrinol Baltim Md 11:1737–1746. doi: 10.1210/mend.11.11.0002

Harding HP, Lazar MA (1995) The monomer-binding orphan receptor Rev-Erb represses transcription as a dimer on a novel direct repeat. Mol Cell Biol 15:4791–4802.

Harding HP, Lazar MA (1993) The orphan receptor Rev-ErbA alpha activates transcription via a novel response element. Mol Cell Biol 13:3113–3121.

Harper CV, Finkenstädt B, Woodcock DJ, et al (2011) Dynamic analysis of stochastic transcription cycles. PLoS Biol 9:e1000607. doi: 10.1371/journal.pbio.1000607

Hastings MH, Reddy AB, Maywood ES (2003) A clockwork web: circadian timing in brain and periphery, in health and disease. Nat Rev Neurosci 4:649–661. doi: 10.1038/nrn1177

He B, Chen C, Teng L, Tan K (2014) Global view of enhancer-promoter interactome in human cells. Proc Natl Acad Sci 111:E2191–E2199. doi: 10.1073/pnas.1320308111

He XH, Li JJ, Xie YH, et al (2004) Altered gene expression profiles of NIH3T3 cells regulated by human lung cancer associated gene CT120. Cell Res 14:487–496. doi: 10.1038/sj.cr.7290252

Heintzman ND, Stuart RK, Hon G, et al (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 39:311–318. doi: 10.1038/ng1966

Hey KL, Momiji H, Featherstone K, et al (2015) A stochastic transcriptional switch model for single cell imaging data. Biostat Oxf Engl 16:655–669. doi: 10.1093/biostatistics/kxv010

Hida A, Koike N, Hirose M, et al (2000) The Human and Mouse Period1 Genes: Five Well-Conserved E-Boxes Additively Contribute to the Enhancement of mPer1 Transcription. Genomics 65:224–233. doi: 10.1006/geno.2000.6166

Hilton IB, D'Ippolito AM, Vockley CM, et al (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat Biotechnol 33:510–517. doi: 10.1038/nbt.3199

Hiragami-Hamada K, Soeroes S, Nikolov M, et al (2016) Dynamic and flexible H3K9me3 bridging via HP1β dimerization establishes a plastic state of condensed chromatin. Nat Commun 7:11310. doi: 10.1038/ncomms11310

Hirose T, Smith RJ, Jetten AM (1994) ROR gamma: the third member of ROR/RZR orphan receptor subfamily that is highly expressed in skeletal muscle. Biochem Biophys Res Commun 205:1976–1983. doi: 10.1006/bbrc.1994.2902

Hirota T, Kon N, Itagaki T, et al (2010a) Transcriptional repressor TIEG1 regulates Bmal1 gene through GC box and controls circadian clockwork. Genes Cells 15:111–121. doi: 10.1111/j.1365-2443.2009.01371.x

Hirota T, Lee JW, Lewis WG, et al (2010b) High-Throughput Chemical Screen Identifies a Novel Potent Modulator of Cellular Circadian Rhythms and Reveals CKIα as a Clock Regulatory Kinase. PLoS Biol 8:e1000559. doi: 10.1371/journal.pbio.1000559

Hnisz D, Shrinivas K, Young RA, et al (2017) A Phase Separation Model for Transcriptional Control. Cell 169:13–23. doi: 10.1016/j.cell.2017.02.007

Ho JWK, Jung YL, Liu T, et al (2014) Comparative analysis of metazoan chromatin organization. Nature 512:449–452. doi: 10.1038/nature13415

Hon GC, Hawkins RD, Ren B (2009) Predictive chromatin signatures in the mammalian genome. Hum Mol Genet 18:R195–R201. doi: 10.1093/hmg/ddp409

Honma S, Kawamoto T, Takagi Y, et al (2002) Dec1 and Dec2 are regulators of the mammalian molecular clock. Nature 419:841–844. doi: 10.1038/nature01123

Hope IA (2004) Feasibility of Genome-Scale Construction of Promoter::Reporter Gene Fusions for Expression in Caenorhabditis elegans Using a MultiSite Gateway Recombination System. Genome Res 14:2070–2075. doi: 10.1101/gr.2463804

Hornung G, Bar-Ziv R, Rosin D, et al (2012) Noise-mean relationship in mutated promoters. Genome Res 22:2409–2417. doi: 10.1101/gr.139378.112

Hughes ME, DiTacchio L, Hayes KR, et al (2009) Harmonics of circadian gene transcription in mammals. PLoS Genet 5:e1000442. doi: 10.1371/journal.pgen.1000442

Ignowski JM, Schaffer DV (2004) Kinetic analysis and modeling of firefly luciferase as a quantitative reporter gene in live mammalian cells. Biotechnol Bioeng 86:827–834. doi: 10.1002/bit.20059

Innocentini G da CP, Forger M, Ramos AF, et al (2013) Multimodality and Flexibility of Stochastic Gene Expression. Bull Math Biol 75:2600–2630. doi: 10.1007/s11538-013-9909-3

Itzkovitz S, Lyubimova A, Blat IC, et al (2011) Single-molecule transcript counting of stem-cell markers in the mouse intestine. Nat Cell Biol 14:106–114. doi: 10.1038/ncb2384

Janich P, Arpat AB, Castelo-Szekely V, et al (2015) Ribosome profiling reveals the rhythmic liver translatome and circadian clock regulation by upstream open reading frames. Genome Res 25:1848–1859. doi: 10.1101/gr.195404.115

Ji N, Middelkoop TC, Mentink RA, et al (2013) Feedback control of gene expression variability in the Caenorhabditis elegans Wnt pathway. Cell 155:869–880. doi: 10.1016/j.cell.2013.09.060

Jin X, Shearman LP, Weaver DR, et al (1999) A molecular mechanism regulating rhythmic output from the suprachiasmatic circadian clock. Cell 96:57–68.

Jolley CC, Ukai-Tadenuma M, Perrin D, Ueda HR (2014) A Mammalian Circadian Clock Model Incorporating Daytime Expression Elements. Biophys J 107:1462–1473. doi: 10.1016/j.bpj.2014.07.022

Jong AY, T'ang A, Liu D-P, Huang S-H (2002) Inverse PCR: Genomic DNA Cloning. In: PCR Cloning Protocols. Humana Press, New Jersey, pp 301–307

Jonkers I, Kwak H, Lis JT (2014) Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. eLife 3:e02407.

Juven-Gershon T, Hsu J-Y, Theisen JW, Kadonaga JT (2008) The RNA polymerase II core promoter - the gateway to transcription. Curr Opin Cell Biol 20:253–259. doi: 10.1016/j.ceb.2008.03.003

Kafri P, Hasenson SE, Kanter I, et al (2016) Quantifying β-catenin subcellular dynamics and cyclin D1 mRNA transcription during Wnt signaling in single living cells. eLife. doi: 10.7554/eLife.16748

Kallen JA, Schlaeppi J-M, Bitsch F, et al (2002) X-ray structure of the hRORalpha LBD at 1.63 A: structural and functional data that cholesterol or a cholesterol derivative is the natural ligand of RORalpha. Struct Lond Engl 1993 10:1697–1707.

Kalo A, Kanter I, Shraga A, et al (2015) Cellular Levels of Signaling Factors Are Sensed by β-actin Alleles to Modulate Transcriptional Pulse Intensity. Cell Rep 11:419–432. doi: 10.1016/j.celrep.2015.03.039

Kar G, Kim JK, Kolodziejczyk AA, et al (2017) Flipping between Polycomb repressed and active transcriptional states introduces noise in gene expression. Nat Commun. doi: 10.1038/s41467-017-00052-2

Katada S, Sassone-Corsi P (2010) The histone methyltransferase MLL1 permits the oscillation of circadian gene expression. Nat Struct Mol Biol 17:1414–1421. doi: 10.1038/nsmb.1961

Kemkemer R, Schrank S, Vogel W, et al (2002) Increased noise as an effect of haploinsufficiency of the tumor-suppressor gene neurofibromatosis type 1 in vitro. Proc Natl Acad Sci U S A 99:13783–13788. doi: 10.1073/pnas.212386999

Kempe H, Schwabe A, Crémazy F, et al (2015) The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise. Mol Biol Cell 26:797–804. doi: 10.1091/mbc.E14-08-1296

Keogh M-C, Kurdistani SK, Morris SA, et al (2005) Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell 123:593–605. doi: 10.1016/j.cell.2005.10.025

Kim D-H, Rhee JC, Yeo S, et al (2015) Crucial roles of mixed-lineage leukemia 3 and 4 as epigenetic switches of the hepatic circadian clock controlling bile acid homeostasis in mice. Hepatol Baltim Md 61:1012–1023. doi: 10.1002/hep.27578

Kim JK, Marioni JC (2013) Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data. Genome Biol 14:R7. doi: 10.1186/gb-2013-14-1-r7

Kim JY, Kwak PB, Weitz CJ (2014) Specificity in Circadian Clock Feedback from Targeted Reconstitution of the NuRD Corepressor. Mol Cell 56:738–748. doi: 10.1016/j.molcel.2014.10.017

Koike N, Yoo S-H, Huang H-C, et al (2012) Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. Science 338:349–354. doi: 10.1126/science.1226339

Kojetin D, Wang Y, Kamenecka TM, Burris TP (2011) Identification of SR8278, a Synthetic Antagonist of the Nuclear Heme Receptor REV-ERB. ACS Chem Biol 6:131–134. doi: 10.1021/cb1002575

Kojetin DJ, Burris TP (2014) REV-ERB and ROR nuclear receptors as drug targets. Nat Rev Drug Discov 13:197–216. doi: 10.1038/nrd4100

Kojima S, Sher-Chen EL, Green CB (2012) Circadian control of mRNA polyadenylation dynamics regulates rhythmic protein expression. Genes Dev 26:2724–2736. doi: 10.1101/gad.208306.112

Kouzarides T (2007) Chromatin Modifications and Their Function. Cell 128:693–705. doi: 10.1016/j.cell.2007.02.005

Kumar N, Solt LA, Conkright JJ, et al (2010) The Benzenesulfoamide T0901317 [N-(2,2,2-Trifluoroethyl)-N-[4-[2,2,2-trifluoro-1-hydroxy-1-(trifluoromethyl)ethyl]phenyl]-benzenesulfonamide] Is a Novel Retinoic Acid Receptor-Related Orphan Receptor-?/? Inverse Agonist. Mol Pharmacol 77:228–236. doi: 10.1124/mol.109.060905

Kume K, Zylka MJ, Sriram S, et al (1999) mCRY1 and mCRY2 are essential components of the negative limb of the circadian clock feedback loop. Cell 98:193–205.

Kurdistani SK, Grunstein M (2003) Histone acetylation and deacetylation in yeast. Nat Rev Mol Cell Biol 4:276–284. doi: 10.1038/nrm1075

Landry J, Sutton A, Hesman T, et al (2003) Set2-catalyzed methylation of histone H3 represses basal expression of GAL4 in Saccharomyces cerevisiae. Mol Cell Biol 23:5972–5978.

Landry JJM, Pyl PT, Rausch T, et al (2013) The Genomic and Transcriptomic Landscape of a HeLa Cell Line. G3 GenesGenomesGenetics 3:1213–1224. doi: 10.1534/g3.113.005777

Larson DR (2011) What do expression dynamics tell us about the mechanism of transcription? Curr Opin Genet Dev 21:591–599. doi: 10.1016/j.gde.2011.07.010

Larson DR, Fritzsch C, Sun L, et al (2013) Direct observation of frequency modulated transcription in single cells using light activation. eLife 2:e00750. doi: 10.7554/eLife.00750

Larson DR, Singer RH, Zenklusen D (2009) A single molecule view of gene expression. Trends Cell Biol 19:630–637. doi: 10.1016/j.tcb.2009.08.008

Larson DR, Zenklusen D, Wu B, et al (2011) Real-time observation of transcription initiation and elongation on an endogenous yeast gene. Science 332:475–478. doi: 10.1126/science.1202142

Lawrence M, Daujat S, Schneider R (2016) Lateral Thinking: How Histone Modifications Regulate Gene Expression. Trends Genet TIG 32:42–56. doi: 10.1016/j.tig.2015.10.007

Le Martelot G, Canella D, Symul L, et al (2012) Genome-wide RNA polymerase II profiles and RNA accumulation reveal kinetics of transcription and associated epigenetic changes during diurnal cycles. PLoS Biol 10:e1001442. doi: 10.1371/journal.pbio.1001442

Lécuyer E, Parthasarathy N, Krause HM (2008) Fluorescent in situ hybridization protocols in Drosophila embryos and tissues. Methods Mol Biol Clifton NJ 420:289–302. doi: 10.1007/978-1-59745-583-1_18

Lehnertz B, Ueda Y, Derijck AAHA, et al (2003) Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr Biol CB 13:1192–1200.

Leibiger C, Kosyakova N, Mkrtchyan H, et al (2013) First Molecular Cytogenetic High Resolution Characterization of the NIH 3T3 Cell Line by Murine Multicolor Banding. J Histochem Cytochem 61:306–312. doi: 10.1369/0022155413476868

Lenstra TL, Rodriguez J, Chen H, Larson DR (2016) Transcription Dynamics in Living Cells. Annu Rev Biophys 45:25–47. doi: 10.1146/annurev-biophys-062215-010838

Lewis K (2007) Persister cells, dormancy and infectious disease. Nat Rev Microbiol 5:48–56. doi: 10.1038/nrmicro1557

Li B, Carey M, Workman JL (2007) The Role of Chromatin during Transcription. Cell 128:707–719. doi: 10.1016/j.cell.2007.01.015

Lin Q, Ding H, Zheng Z, et al (2012) Promoter methylation analysis of seven clock genes in Parkinson's disease. Neurosci Lett 507:147–150. doi: 10.1016/j.neulet.2011.12.007

Lionnet T, Czaplinski K, Darzacq X, et al (2011) A transgenic mouse for in vivo detection of endogenous labeled mRNA. Nat Methods 8:165–170. doi: 10.1038/nmeth.1551

Lionnet T, Singer RH (2012) Transcription goes digital. EMBO Rep 13:313–321. doi: 10.1038/embor.2012.31

Little SC, Tikhonov M, Gregor T (2013) Precise developmental gene expression arises from globally stochastic transcriptional activity. Cell 154:789–800. doi: 10.1016/j.cell.2013.07.025

Liu AC, Tran HG, Zhang EE, et al (2008) Redundant function of REV-ERBalpha and beta and non-essential role for Bmal1 cycling in transcriptional regulation of intracellular circadian rhythms. PLoS Genet 4:e1000023. doi: 10.1371/journal.pgen.1000023

Liu C, Li S, Liu T, et al (2007) Transcriptional coactivator PGC-1alpha integrates the mammalian clock and energy metabolism. Nature 447:477–481. doi: 10.1038/nature05767

Longo VD, Mattson MP (2014) Fasting: Molecular Mechanisms and Clinical Applications. Cell Metab 19:181–192. doi: 10.1016/j.cmet.2013.12.008

Lopez-Atalaya JP, Ito S, Valor LM, et al (2013) Genomic targets, and histone acetylation and gene expression profiling of neural HDAC inhibition. Nucleic Acids Res 41:8072–8084. doi: 10.1093/nar/gkt590

Losick R, Desplan C (2008) Stochasticity and Cell Fate. Science 320:65–68. doi: 10.1126/science.1147888

Lück S, Thurley K, Thaben PF, Westermark PO (2014) Rhythmic Degradation Explains and Unifies Circadian Transcriptome and Proteome Data. Cell Rep 9:741–751. doi: 10.1016/j.celrep.2014.09.021

Lyubimova A, Itzkovitz S, Junker JP, et al (2013) Single-molecule mRNA detection and counting in mammalian tissue. Nat Protoc 8:1743–1758. doi: 10.1038/nprot.2013.109

Mao YS, Sunwoo H, Zhang B, Spector DL (2011) Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. Nat Cell Biol 13:95–101. doi: 10.1038/ncb2140

Marguerat S, Baehler J (2012) Coordinating genome expression with cell size. Trends Genet 28:560–565. doi: 10.1016/j.tig.2012.07.003

Marquès-Bueno MM, Morao AK, Cayrel A, et al (2016) A versatile Multisite Gateway-compatible promoter and transgenic line collection for cell type-specific functional genomics in Arabidopsis. Plant J Cell Mol Biol 85:320–333. doi: 10.1111/tpj.13099

Matoulkova E, Michalova E, Vojtesek B, Hrstka R (2012) The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. RNA Biol 9:563–576. doi: 10.4161/rna.20231

Mauer J, Luo X, Blanjoie A, et al (2017) Reversible methylation of m(6)Am in the 5' cap controls mRNA stability. Nature 541:371–375. doi: 10.1038/nature21022

McNally JG, Müller WG, Walker D, et al (2000) The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. Science 287:1262–1265.

Menet JS, Pescatore S, Rosbash M (2014) CLOCK:BMAL1 is a pioneer-like transcription factor. Genes Dev 28:8–13. doi: 10.1101/gad.228536.113

Mermet J, Yeung J, Naef F (2017) Systems Chronobiology: Global Analysis of Gene Regulation in a 24-Hour Periodic World. Cold Spring Harb Perspect Biol. doi: 10.1101/cshperspect.a028720

Mignone F, Gissi C, Liuni S, Pesole G (2002) Untranslated regions of mRNAs. Genome Biol 3:REVIEWS0004.

Miller BH, McDearmon EL, Panda S, et al (2007) Circadian and CLOCK-controlled regulation of the mouse transcriptome and cell proliferation. Proc Natl Acad Sci 104:3342–3347. doi: 10.1073/pnas.0611724104

Mitsui S, Yamaguchi S, Matsuo T, et al (2001) Antagonistic role of E4BP4 and PAR proteins in the circadian oscillatory mechanism. Genes Dev 15:995–1006. doi: 10.1101/gad.873501

Mohawk JA, Green CB, Takahashi JS (2012) Central and Peripheral Circadian Clocks in Mammals. Annu Rev Neurosci 35:445–462. doi: 10.1146/annurev-neuro-060909-153128

Molina N, Suter DM, Cannavo R, et al (2013) Stimulus-induced modulation of transcriptional bursting in a single mammalian gene. Proc Natl Acad Sci U S A 110:20563–20568. doi: 10.1073/pnas.1312310110

Morf J, Rey G, Schneider K, et al (2012) Cold-Inducible RNA-Binding Protein Modulates Circadian Gene Expression Posttranscriptionally. Science 338:379–383. doi: 10.1126/science.1217726

Morisaki T, Müller WG, Golob N, et al (2014) Single-molecule analysis of transcription factor binding at transcription sites in live cells. Nat Commun 5:4456. doi: 10.1038/ncomms5456

Mugler A, Walczak AM, Wiggins CH (2009) Spectral solutions to stochastic models of gene expression with bursts and regulation. Phys Rev E Stat Nonlin Soft Matter Phys 80:41921. doi: 10.1103/PhysRevE.80.041921

Munsky B, Neuert G, van Oudenaarden A (2012) Using gene expression noise to understand gene regulation. Science 336:183–187. doi: 10.1126/science.1216379

Muramoto T, Cannon D, Gierlinski M, et al (2012) Live imaging of nascent RNA dynamics reveals distinct types of transcriptional pulse regulation. Proc Natl Acad Sci U S A 109:7350–7355. doi: 10.1073/pnas.1117603109

Muramoto T, Müller I, Thomas G, et al (2010) Methylation of H3K4 Is required for inheritance of active transcriptional states. Curr Biol CB 20:397–406. doi: 10.1016/j.cub.2010.01.017

Nagoshi E, Brown SA, Dibner C, et al (2005) Circadian gene expression in cultured cells. Methods Enzymol 393:543–557. doi: 10.1016/S0076-6879(05)93028-0

Nagoshi E, Saini C, Bauer C, et al (2004) Circadian Gene Expression in Individual Fibroblasts. Cell 119:693–705. doi: 10.1016/j.cell.2004.11.015

Nakahata Y, Yoshida M, Takano A, et al (2008) A direct repeat of E-box-like elements is required for cell-autonomous circadian rhythm of clock genes. BMC Mol Biol 9:1. doi: 10.1186/1471-2199-9-1

Nakashima A, Kawamoto T, Honda KK, et al (2008) DEC1 Modulates the Circadian Phase of Clock Gene Expression. Mol Cell Biol 28:4080–4092. doi: 10.1128/MCB.02168-07

Neuert G, Munsky B, Tan RZ, et al (2013) Systematic identification of signal-activated stochastic gene regulation. Science 339:584–587. doi: 10.1126/science.1231456

Newman JRS, Ghaemmaghami S, Ihmels J, et al (2006) Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature 441:840–846. doi: 10.1038/nature04785

Nicolas D, Phillips NE, Naef F (2017) What shapes eukaryotic transcriptional bursting? Mol BioSyst. doi: 10.1039/C7MB00154A

Nishida H, Suzuki T, Kondo S, et al (2006) Histone H3 acetylated at lysine 9 in promoter is associated with low nucleosome density in the vicinity of transcription start site in human cell. Chromosome Res 14:203–211. doi: 10.1007/s10577-006-1036-7

Noguchi T, Wang LL, Welsh DK (2013) Fibroblast PER2 Circadian Rhythmicity Depends on Cell Density. J Biol Rhythms 28:183–192. doi: 10.1177/0748730413487494

Norris AJ, Stirland JA, McFerran DW, et al (2003) Dynamic Patterns of Growth Hormone Gene Transcription in Individual Living Pituitary Cells. Mol Endocrinol 17:193–202. doi: 10.1210/me.2002-0201

Ochiai H, Sugawara T, Sakuma T, Yamamoto T (2014) Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells. Sci Rep 4:7125. doi: 10.1038/srep07125

O'Geen H, Echipare L, Farnham PJ (2011) Using ChIP-Seq Technology to Generate High-Resolution Profiles of Histone Modifications. In: Tollefsbol TO (ed) Epigenetics Protocols. Humana Press, Totowa, NJ, pp 265–286

Ohno T, Onishi Y, Ishida N (2006) A novel E4BP4 element drives circadian expression of mPeriod2. Nucleic Acids Res 35:648–655. doi: 10.1093/nar/gkl868

O'Neill JS, Hastings MH (2008) Increased Coherence of Circadian Rhythms in Mature Fibroblast Cultures. J Biol Rhythms 23:483–488. doi: 10.1177/0748730408326682

Onishi Y, Hanai S, Ohno T, et al (2008) Rhythmic SAF-A binding underlies circadian transcription of the Bmal1 gene. Mol Cell Biol 28:3477–3488. doi: 10.1128/MCB.02227-07

Owens NDL, Blitz IL, Lane MA, et al (2016) Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. Cell Rep 14:632–647. doi: 10.1016/j.celrep.2015.12.050

Padovan-Merhar O, Nair GP, Biaesch AG, et al (2015) Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. Mol Cell 58:339–352. doi: 10.1016/j.molcel.2015.03.005

Panda S (2016) Circadian physiology of metabolism. Science 354:1008–1015. doi: 10.1126/science.aah4967

Panda S, Antoch MP, Miller BH, et al (2002) Coordinated transcription of key pathways in the mouse by the circadian clock. Cell 109:307–320.

Papazyan R, Zhang Y, Lazar MA (2016) Genetic and epigenomic mechanisms of mammalian circadian transcription. Nat Struct Mol Biol 23:1045–1052. doi: 10.1038/nsmb.3324

Paré A, Lemons D, Kosman D, et al (2009) Visualization of Individual Scr mRNAs during Drosophila Embryogenesis Yields Evidence for Transcriptional Bursting. Curr Biol 19:2037–2042. doi: 10.1016/j.cub.2009.10.028

Parish CR (1999) Fluorescent dyes for lymphocyte migration and proliferation studies. Immunol Cell Biol 77:499–508. doi: 10.1046/j.1440-1711.1999.00877.x

Partch CL, Green CB, Takahashi JS (2014) Molecular architecture of the mammalian circadian clock. Trends Cell Biol 24:90–99. doi: 10.1016/j.tcb.2013.07.002

Patel DJ, Wang Z (2013) Readout of Epigenetic Modifications. Annu Rev Biochem 82:81–118. doi: 10.1146/annurev-biochem-072711-165700

Paulsson J (2004) Summing up the noise in gene networks. Nature 427:415–418. doi: 10.1038/nature02257

Peccoud J, Bernard Y (1995) Markovian Modeling of Gene-Product Synthesis. Theor Popul Biol 48:222–234.

Peccoud J, Ycart B (1995) Markovian Modeling of Gene-Product Synthesis.

Pedraza JM, Paulsson J (2008) Effects of Molecular Memory and Bursting on Fluctuations in Gene Expression. Science 319:339–343. doi: 10.1126/science.1144331

Phatnani HP, Greenleaf AL (2006) Phosphorylation and functions of the RNA polymerase II CTD. Genes Dev 20:2922–2936. doi: 10.1101/gad.1477006

Piechaczyk M, Blanchard JM, Marty L, et al (1984) Post-transcriptional regulation of glyceraldehyde-3-phosphate-dehydrogenase gene expression in rat tissues. Nucleic Acids Res 12:6951–6963.

Plautz SA, Boanca G, Riethoven J-JM, Pannier AK (2011) Microarray Analysis of Gene Expression Profiles in Cells Transfected With Nonviral Vectors. Mol Ther 19:2144–2151. doi: 10.1038/mt.2011.161

Player AN, Shen LP, Kenny D, et al (2001) Single-copy gene detection using branched DNA (bDNA) in situ hybridization. J Histochem Cytochem Off J Histochem Soc 49:603–612.

Preitner N, Damiola F, Lopez-Molina L, et al (2002) The orphan nuclear receptor REV-ERBalpha controls circadian transcription within the positive limb of the mammalian circadian oscillator. Cell 110:251–260.

Qin JY, Zhang L, Clift KL, et al (2010) Systematic Comparison of Constitutive Promoters and the Doxycycline-Inducible Promoter. PLoS ONE 5:e10611. doi: 10.1371/journal.pone.0010611

Quinodoz M, Gobet C, Naef F, Gustafson KB (2014) Characteristic bimodal profiles of RNA polymerase II at thousands of active mammalian promoters. Genome Biol 15:R85. doi: 10.1186/gb-2014-15-6-r85

Raghuram S, Stayrook KR, Huang P, et al (2007) Identification of heme as the ligand for the orphan nuclear receptors REV-ERB? and REV-ERB? Nat Struct Mol Biol 14:1207–1213. doi: 10.1038/nsmb1344

Raikhel NV, Mishkind ML, Palevitz BA (1984) Characterization of a wheat germ agglutinin-like lectin from adult wheat plants. Planta 162:55–61. doi: 10.1007/BF00397421

Raj A, Peskin CS, Tranchina D, et al (2006) Stochastic mRNA synthesis in mammalian cells. PLoS Biol 4:e309. doi: 10.1371/journal.pbio.0040309

Raj A, Rifkin SA, Andersen E, van Oudenaarden A (2010) Variability in gene expression underlies incomplete penetrance. Nature 463:913–918. doi: 10.1038/nature08781

Raj A, Tyagi S (2010) Detection of individual endogenous RNA transcripts in situ using multiple singly labeled probes. Methods Enzymol 472:365–386. doi: 10.1016/S0076-6879(10)72004-8

Raj A, van den Bogaard P, Rifkin SA, et al (2008) Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods 5:877–879. doi: 10.1038/nmeth.1253

Raj A, van Oudenaarden A (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. Cell 135:216–226. doi: 10.1016/j.cell.2008.09.050

Raj A, van Oudenaarden A (2009) Single-Molecule Approaches to Stochastic Gene Expression. Annu Rev Biophys 38:255–270. doi: 10.1146/annurev.biophys.37.032807.125928

Ramanathan C, Xu H, Khan SK, et al (2014) Cell Type-Specific Functions of Period Genes Revealed by Novel Adipocyte and Hepato-cyte Circadian Clock Models. PLoS Genet 10:e1004244. doi: 10.1371/journal.pgen.1004244

Raser JM, O'Shea EK (2004) Control of stochasticity in eukaryotic gene expression. Science 304:1811–1814. doi: 10.1126/science.1098641

Raspè E, Mautino G, Duval C, et al (2002) Transcriptional regulation of human Rev-erbalpha gene expression by the orphan nuclear receptor retinoic acid-related orphan receptor alpha. J Biol Chem 277:49275–49281. doi: 10.1074/jbc.M206215200

Reddy TE, Gertz J, Crawford GE, et al (2012) The hypersensitive glucocorticoid response specifically regulates period 1 and expression of circadian genes. Mol Cell Biol 32:3756–3767. doi: 10.1128/MCB.00062-12

Reischl S, Vanselow K, Westermark PO, et al (2007) Beta-TrCP1-mediated degradation of PERIOD2 is essential for circadian dynamics. J Biol Rhythms 22:375–386. doi: 10.1177/0748730407303926

Rey G, Cesbron F, Rougemont J, et al (2011) Genome-wide and phase-specific DNA-binding rhythms of BMAL1 control circadian output functions in mouse liver. PLoS Biol 9:e1000595. doi: 10.1371/journal.pbio.1000595

Rinott R, Jaimovich A, Friedman N (2011) Exploring transcription regulation through cell-to-cell variability. Proc Natl Acad Sci U S A 108:6329–6334. doi: 10.1073/pnas.1013148108

Ripperger JA, Schibler U (2006) Rhythmic CLOCK-BMAL1 binding to multiple E-box motifs drives circadian Dbp transcription and chromatin transitions. Nat Genet 38:369–374. doi: 10.1038/ng1738

Ripperger JA, Shearman LP, Reppert SM, Schibler U (2000) CLOCK, an essential pacemaker component, controls expression of the circadian transcription factor DBP. Genes Dev 14:679–689.

Rogers S, Wells R, Rechsteiner M (1986) Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. Science 234:364–368. doi: 10.1126/science.2876518

Rosenfeld N, Perkins TJ, Alon U, et al (2006) A fluctuation method to quantify in vivo fluorescence data. Biophys J 91:759–766. doi: 10.1529/biophysj.105.073098

Rosenfeld N, Young JW, Alon U, et al (2005) Gene Regulation at the Single-Cell Level. Science 307:1962–1965. doi: 10.1126/science.1106914

Rosonina E, Kaneko S, Manley JL (2006) Terminating the transcript: breaking up is hard to do. Genes Dev 20:1050–1056. doi: 10.1101/gad.1431606

Roth SY, Denu JM, Allis CD (2001) Histone Acetyltransferases. Annu Rev Biochem 70:81–120. doi: 10.1146/annurev.biochem.70.1.81

Russell J, Zomerdijk JCBM (2006) The RNA polymerase I transcription machinery. Biochem Soc Symp 203–216.

Ryan MD, Drew J (1994) Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. EMBO J 13:928–33.

Saeki H, Svejstrup JQ (2009) Stability, flexibility, and dynamic interactions of colliding RNA polymerase II elongation complexes. Mol Cell 35:191–205. doi: 10.1016/j.molcel.2009.06.009

Said HM, Hagemann C, Stojic J, et al (2007) GAPDH is not regulated in human glioblastoma under hypoxic conditions. BMC Mol Biol 8:55. doi: 10.1186/1471-2199-8-55

Sainsbury S, Bernecky C, Cramer P (2015) Structural basis of transcription initiation by RNA polymerase II. Nat Rev Mol Cell Biol 16:129–143. doi: 10.1038/nrm3952

Sanchez A, Golding I (2013) Genetic determinants and cellular constraints in noisy gene expression. Science 342:1188–1193. doi: 10.1126/science.1242975

Sander JD, Joung JK (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol 32:347–355. doi: 10.1038/nbt.2842

Sanyal A, Lajoie BR, Jain G, Dekker J (2012) The long-range interaction landscape of gene promoters. Nature 489:109–113. doi: 10.1038/nature11279

Sasaki YTF, Ideue T, Sano M, et al (2009) MEN / noncoding RNAs are essential for structural integrity of nuclear paraspeckles. Proc Natl Acad Sci 106:2525–2530. doi: 10.1073/pnas.0807899106

Sato TK, Panda S, Miraglia LJ, et al (2004) A functional genomics strategy reveals Rora as a component of the mammalian circadian clock. Neuron 43:527–537. doi: 10.1016/j.neuron.2004.07.018

Sato TK, Yamada RG, Ukai H, et al (2006) Feedback repression is required for mammalian circadian clock function. Nat Genet 38:312–319. doi: 10.1038/ng1745

Satou R, Sugihara N, Ishizuka Y, et al (2013) DNA methylation of the BMAL1 promoter. Biochem Biophys Res Commun 440:449–453. doi: 10.1016/j.bbrc.2013.09.124

Schlake T, Bode J (1994) Use of Mutated FLP Recognition Target (FRT) Sites for the Exchange of Expression Cassettes at Defined Chromosomal Loci. Biochemistry (Mosc) 33:12746–12751. doi: 10.1021/bi00209a003

Schmidt EE, Schibler U (1995) Cell size regulation, a mechanism that controls cellular RNA accumulation: consequences on regulation of the ubiquitous transcription factors Oct1 and NF-Y and the liver-enriched transcription factor DBP. J Cell Biol 128:467–483.

Schoech AP, Zabet NR (2014) Facilitated diffusion buffers noise in gene expression. Phys Rev E. doi: 10.1103/PhysRevE.90.032701

Schramm L, Hernandez N (2002) Recruitment of RNA polymerase III to its target promoters. Genes Dev 16:2593–2620. doi: 10.1101/gad.1018902

Senecal A, Munsky B, Proux F, et al (2014) Transcription factors modulate c-Fos transcriptional bursts. Cell Rep 8:75–83. doi: 10.1016/j.celrep.2014.05.053

Seto E, Yoshida M (2014) Erasers of histone acetylation: the histone deacetylase enzymes. Cold Spring Harb Perspect Biol 6:a018713. doi: 10.1101/cshperspect.a018713

Sevier SA, Kessler DA, Levine H (2016) Mechanical bounds to transcriptional noise. Proc Natl Acad Sci 113:13983–13988. doi: 10.1073/pnas.1612651113

Shaffer SM, Dunagin MC, Torborg SR, et al (2017) Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. Nature 546:431–435. doi: 10.1038/nature22794

Shahrezaei V, Swain PS (2008) Analytical distributions for stochastic gene expression. Proc Natl Acad Sci U S A 105:17256–17261. doi: 10.1073/pnas.0803850105

Shandilya J, Roberts SGE (2012) The transcription cycle in eukaryotes: From productive initiation to RNA polymerase II recycling. Biochim Biophys Acta BBA - Gene Regul Mech 1819:391–400. doi: 10.1016/j.bbagrm.2012.01.010

Shao W, Zeitlinger J (2017) Paused RNA polymerase II inhibits new transcriptional initiation. Nat Genet 49:1045–1051. doi: 10.1038/ng.3867

Sharma SV, Lee DY, Li B, et al (2010) A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell 141:69–80. doi: 10.1016/j.cell.2010.02.027

Sharova LV, Sharov AA, Nedorezov T, et al (2009) Database for mRNA Half-Life of 19 977 Genes Obtained by DNA Microarray Analysis of Pluripotent and Differentiating Mouse Embryonic Stem Cells. DNA Res 16:45–58. doi: 10.1093/dnares/dsn030

Shaw G, Kamen R (1986) A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell 46:659–667. doi: 10.1016/0092-8674(86)90341-7

Shlyueva D, Stampfel G, Stark A (2014) Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet 15:272–286. doi: 10.1038/nrg3682

Shogren-Knaak M (2006) Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. Science 311:844–847. doi: 10.1126/science.1124000

Shostak A, Ruppert B, Ha N, et al (2016) MYC/MIZ1-dependent gene repression inversely coordinates the circadian clock with cell cycle and proliferation. Nat Commun 7:11807. doi: 10.1038/ncomms11807

Sierk ML, Zhao Q, Rastinejad F (2001) DNA Deformability as a Recognition Feature in the RevErb Response Element. Biochemistry (Mosc) 40:12833–12843. doi: 10.1021/bi011086r

Siersbæk R, Madsen JGS, Javierre BM, et al (2017) Dynamic Rewiring of Promoter-Anchored Chromatin Loops during Adipocyte Differentiation. Mol Cell 66:420–435.e5. doi: 10.1016/j.molcel.2017.04.010

Simon M, North JA, Shimko JC, et al (2011) Histone fold modifications control nucleosome unwrapping and disassembly. Proc Natl Acad Sci 108:12711–12716. doi: 10.1073/pnas.1106264108

Singer ZS, Yong J, Tischler J, et al (2014) Dynamic heterogeneity and DNA methylation in embryonic stem cells. Mol Cell 55:319–331. doi: 10.1016/j.molcel.2014.06.029

Singh A, Razooky B, Cox CD, et al (2010) Transcriptional bursting from the HIV-1 promoter is a significant source of stochastic noise in HIV-1 gene expression. Biophys J 98:L32-34. doi: 10.1016/j.bpj.2010.03.001

Singh S, Yang Y, Poczos B, Ma J (2016) Predicting Enhancer-Promoter Interaction from Genomic Sequence with Deep Neural Networks. doi: 10.1101/085241

Skinner SO, Xu H, Nagarkar-Jaiswal S, et al (2016) Single-cell analysis of transcription kinetics across the cell cycle. eLife 5:e12175. doi: 10.7554/eLife.12175

Skupsky R, Burnett JC, Foley JE, et al (2010) HIV promoter integration site primarily modulates transcriptional burst size rather than frequency. PLoS Comput Biol. doi: 10.1371/journal.pcbi.1000952

Small EC, Xi L, Wang J-P, et al (2014) Single-cell nucleosome mapping reveals the molecular basis of gene expression heterogeneity. Proc Natl Acad Sci U S A 111:E2462-2471. doi: 10.1073/pnas.1400517111

So AY-L, Bernal TU, Pillsbury ML, et al (2009) Glucocorticoid regulation of the circadian clock modulates glucose homeostasis. Proc Natl Acad Sci 106:17582–17587. doi: 10.1073/pnas.0909733106

So L-H, Ghosh A, Zong C, et al (2011) General properties of transcriptional time series in Escherichia coli. Nat Genet 43:554–560. doi: 10.1038/ng.821

Sobel JA, Krier I, Andersin T, et al (2017) Transcriptional regulatory logic of the diurnal cycle in the mouse liver. PLOS Biol 15:e2001069. doi: 10.1371/journal.pbio.2001069

Spencer SL, Gaudet S, Albeck JG, et al (2009) Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature 459:428–432. doi: 10.1038/nature08012

Spörl F, Schellenberg K, Blatt T, et al (2011) A Circadian Clock in HaCaT Keratinocytes. J Invest Dermatol 131:338–348. doi: 10.1038/jid.2010.315

Sprouse RO, Karpova TS, Mueller F, et al (2008) Regulation of TATA-binding protein dynamics in living yeast cells. Proc Natl Acad Sci U S A 105:13304–13308. doi: 10.1073/pnas.0801901105

St. John PC, Hirota T, Kay SA, Doyle FJ (2014) Spatiotemporal separation of PER and CRY posttranslational regulation in the mammalian circadian clock. Proc Natl Acad Sci 111:2040–2045. doi: 10.1073/pnas.1323618111

Stegle O, Teichmann SA, Marioni JC (2015) Computational and analytical challenges in single-cell transcriptomics. Nat Rev Genet 16:133–145. doi: 10.1038/nrg3833

Sterner DE, Berger SL (2000) Acetylation of histones and transcription-related factors. Microbiol Mol Biol Rev MMBR 64:435–459.

Stewart-Ornstein J, Weissman JS, El-Samad H (2012) Cellular noise regulons underlie fluctuations in Saccharomyces cerevisiae. Mol Cell 45:483–493. doi: 10.1016/j.molcel.2011.11.035

Stokkan KA, Yamazaki S, Tei H, et al (2001) Entrainment of the circadian clock in the liver by feeding. Science 291:490–493. doi: 10.1126/science.291.5503.490

Storch K-F, Lipan O, Leykin I, et al (2002) Extensive and divergent circadian gene expression in liver and heart. Nature 417:78–83. doi: 10.1038/nature744

Strahl BD, Grant PA, Briggs SD, et al (2002) Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. Mol Cell Biol 22:1298–1306.

Stratmann M, Stadler F, Tamanini F, et al (2010) Flexible phase adjustment of circadian albumin D site-binding protein (Dbp) gene expression by CRYPTOCHROME1. Genes Dev 24:1317–1328. doi: 10.1101/gad.578810

Stratmann M, Suter DM, Molina N, et al (2012) Circadian Dbp Transcription Relies on Highly Dynamic BMAL1-CLOCK Interaction with E Boxes and Requires the Proteasome. Mol Cell 48:277–287. doi: 10.1016/j.molcel.2012.08.012

Struhl K (1998) Histone acetylation and transcriptional regulatory mechanisms. Genes Dev 12:599–606.

Sun Z, Feng D, Everett LJ, et al (2011) Circadian epigenomic remodeling and hepatic lipogenesis: lessons from HDAC3. Cold Spring Harb Symp Quant Biol 76:49–55. doi: 10.1101/sqb.2011.76.011494

Sunwoo H, Dinger ME, Wilusz JE, et al (2008) MEN / nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. Genome Res 19:347–359. doi: 10.1101/gr.087775.108

Suter DM, Molina N, Gatfield D, et al (2011a) Mammalian genes are transcribed with widely different bursting kinetics. Science 332:472–474. doi: 10.1126/science.1198817

Suter DM, Molina N, Naef F, Schibler U (2011b) Origins and consequences of transcriptional discontinuity. Curr Opin Cell Biol 23:657–662. doi: 10.1016/j.ceb.2011.09.004

Symmons O, Raj A (2016) What's Luck Got to Do with It: Single Cells, Multiple Fates, and Biological Nondeterminism. Mol Cell 62:788–802. doi: 10.1016/j.molcel.2016.05.023

Takahashi JS (2016) Transcriptional architecture of the mammalian circadian clock. Nat Rev Genet 18:164–179. doi: 10.1038/nrg.2016.150

Takeda Y, Jothi R, Birault V, Jetten AM (2012) RORc directly regulates the circadian expression of clock genes and downstream targets in vivo. Nucleic Acids Res 40:8519–8535. doi: 10.1093/nar/gks630

Takeda Y, Kang HS, Angers M, Jetten AM (2011) Retinoic acid-related orphan receptor γ directly regulates neuronal PAS domain protein 2 transcription in vivo. Nucleic Acids Res 39:4769–4782. doi: 10.1093/nar/gkq1335

Tani H, Mizutani R, Salam KA, et al (2012) Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. Genome Res 22:947–956. doi: 10.1101/gr.130559.111

Taniguchi H, Fernandez AF, Setien F, et al (2009) Epigenetic Inactivation of the Circadian Clock Gene BMAL1 in Hematologic Malignancies. Cancer Res 69:8447–8454. doi: 10.1158/0008-5472.CAN-09-0551

Taniguchi Y, Choi PJ, Li G-W, et al (2010) Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. Science 329:533–538. doi: 10.1126/science.1188308

Tantale K, Mueller F, Kozulic-Pirher A, et al (2016) A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. Nat Commun 7:12248. doi: 10.1038/ncomms12248

Teles J, Pina C, Edén P, et al (2013) Transcriptional regulation of lineage commitment--a stochastic model of cell fate decisions. PLoS Comput Biol 9:e1003197. doi: 10.1371/journal.pcbi.1003197

Tie F, Banerjee R, Stratton CA, et al (2009) CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. Development 136:3131–3141. doi: 10.1242/dev.037127

Tirosh I, Barkai N (2008) Two strategies for gene regulation by promoter nucleosomes. Genome Res 18:1084–1091. doi: 10.1101/gr.076059.108

Travnickova-Bendova Z, Cermakian N, Reppert SM, Sassone-Corsi P (2002) Bimodal regulation of mPeriod promoters by CREB-dependent signaling and CLOCK/BMAL1 activity. Proc Natl Acad Sci U S A 99:7728–7733. doi: 10.1073/pnas.102075599

Ueda HR, Chen W, Adachi A, et al (2002) A transcription factor response element for gene expression during circadian night. Nature 418:534–539. doi: 10.1038/nature00906

Ueda HR, Hayashi S, Chen W, et al (2005) System-level identification of transcriptional circuits underlying mammalian circadian clocks. Nat Genet 37:187–192. doi: 10.1038/ng1504

Ueshima T, Kawamoto T, Honda KK, et al (2012) Identification of a new clock-related element EL-box involved in circadian regulation by BMAL1/CLOCK and HES1. Gene 510:118–125. doi: 10.1016/j.gene.2012.08.022

Ukai-Tadenuma M, Kasukawa T, Ueda HR (2008) Proof-by-synthesis of the transcriptional logic of mammalian circadian clocks. Nat Cell Biol 10:1154–1163. doi: 10.1038/ncb1775

Ukai-Tadenuma M, Yamada RG, Xu H, et al (2011) Delay in Feedback Repression by Cryptochrome 1 Is Required for Circadian Clock Function. Cell 144:268–281. doi: 10.1016/j.cell.2010.12.019

Uphoff S, Lord ND, Okumus B, et al (2016) Stochastic activation of a DNA damage response causes cell-to-cell mutation rate variation. Science 351:1094–1097. doi: 10.1126/science.aac9786

Valekunja UK, Edgar RS, Oklejewicz M, et al (2013) Histone methyltransferase MLL3 contributes to genome-scale circadian transcription. Proc Natl Acad Sci 110:1554–1559. doi: 10.1073/pnas.1214168110

van der Horst GT, Muijtjens M, Kobayashi K, et al (1999) Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms. Nature 398:627–630. doi: 10.1038/19323

Vanhaecke T, Papeleu P, Elaut G, Rogiers V (2004) Trichostatin A - like Hydroxamate Histone Deacetylase Inhibitors as Therapeutic Agents: Toxicological Point of View. Curr Med Chem 11:1629–1643. doi: 10.2174/0929867043365099

Vera M, Biswas J, Senecal A, et al (2016) Single-Cell and Single-Molecule Analysis of Gene Expression Regulation. Annu Rev Genet 50:267–291. doi: 10.1146/annurev-genet-120215-034854

Verdone L, Agricola E, Caserta M, Di Mauro E (2006) Histone acetylation in gene regulation. Brief Funct Genomic Proteomic 5:209–221. doi: 10.1093/bfgp/ell028

Vielhaber E, Eide E, Rivers A, et al (2000) Nuclear entry of the circadian regulator mPER1 is controlled by mammalian casein kinase I epsilon. Mol Cell Biol 20:4888–4899.

Viñuelas J, Kaneko G, Coulon A, et al (2013) Quantifying the contribution of chromatin dynamics to stochastic gene expression reveals long, locus-dependent periods between transcriptional bursts. BMC Biol 11:15. doi: 10.1186/1741-7007-11-15

Vollmers C, Gill S, DiTacchio L, et al (2009) Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression. Proc Natl Acad Sci 106:21453–21458. doi: 10.1073/pnas.0909591106

Vollmers C, Schmitz RJ, Nathanson J, et al (2012) Circadian Oscillations of Protein-Coding and Regulatory RNAs in a Highly Dynamic Mammalian Liver Epigenome. Cell Metab 16:833–845. doi: 10.1016/j.cmet.2012.11.004

Voss TC, Hager GL (2014) Dynamic regulation of transcriptional states by chromatin and transcription factors. Nat Rev Genet 15:69–81. doi: 10.1038/nrg3623

Wagner EJ, Carpenter PB (2012) Understanding the language of Lys36 methylation at histone H3. Nat Rev Mol Cell Biol 13:115–126. doi: 10.1038/nrm3274

Wang Y, Kumar N, Crumbley C, et al (2010a) A second class of nuclear receptors for oxysterols: Regulation of RORa and RORc activity by 24S-hydroxycholesterol (cerebrosterol). Biochim Biophys Acta BBA - Mol Cell Biol Lipids 1801:917–923. doi: 10.1016/j.bbalip.2010.02.012

Wang Y, Kumar N, Solt LA, et al (2010b) Modulation of Retinoic Acid Receptor-related Orphan Receptor α and γ Activity by 7-Oxygenated Sterol Ligands. J Biol Chem 285:5013–5025. doi: 10.1074/jbc.M109.080614

Weinberger L, Voichek Y, Tirosh I, et al (2012) Expression Noise and Acetylation Profiles Distinguish HDAC Functions. Mol Cell 47:193–202. doi: 10.1016/j.molcel.2012.05.008

Welsh DK, Takahashi JS, Kay SA (2010) Suprachiasmatic Nucleus: Cell Autonomy and Network Properties. Annu Rev Physiol 72:551–577. doi: 10.1146/annurev-physiol-021909-135919

Whalen S, Truty RM, Pollard KS (2016) Enhancer–promoter interactions are encoded by complex genomic signatures on looping chromatin. Nat Genet 48:488–496. doi: 10.1038/ng.3539

Wilsbacher LD, Yamazaki S, Herzog ED, et al (2002) Photic and circadian expression of luciferase in mPeriod1-luc transgenic mice in vivo. Proc Natl Acad Sci 99:489–494. doi: 10.1073/pnas.012248599

Wirth D, Hauser H (2004) Flp-Mediated Integration of Expression Cassettes into FRT-Tagged Chromosomal Loci in Mammalian Cells. In: Recombinant Gene Expression. Humana Press, New Jersey, pp 467–476

Wolf L, Silander OK, van Nimwegen E (2015) Expression noise facilitates the evolution of gene regulation. eLife. doi: 10.7554/eLife.05856

Wu S, Li K, Li Y, et al (2017) Independent regulation of gene expression level and noise by histone modifications. PLOS Comput Biol 13:e1005585. doi: 10.1371/journal.pcbi.1005585

Xiao J, Zhou Y, Lai H, et al (2013) Transcription Factor NF-Y Is a Functional Regulator of the Transcription of Core Clock Gene Bmal1. J Biol Chem 288:31930–31936. doi: 10.1074/jbc.M113.507038

Xu H, Sepúlveda LA, Figard L, et al (2015) Combining protein and mRNA quantification to decipher transcriptional regulation. Nat Methods 12:739–742. doi: 10.1038/nmeth.3446

Yamaguchi S, Mitsui S, Yan L, et al (2000) Role of DBP in the circadian oscillatory mechanism. Mol Cell Biol 20:4773–4781.

Yamajuku D, Shibata Y, Kitazawa M, et al (2010) Identification of functional clock-controlled elements involved in differential timing of Per1 and Per2 transcription. Nucleic Acids Res 38:7964–7973. doi: 10.1093/nar/gkq678

Yamamoto T, Nakahata Y, Soma H, et al (2004) Transcriptional oscillation of canonical clock genes in mouse peripheral tissues. BMC Mol Biol 5:18. doi: 10.1186/1471-2199-5-18

Yang F, Inoue I, Kumagai M, et al (2013) Real-time analysis of the circadian oscillation of the Rev-Erb β promoter. J Atheroscler Thromb 20:267–276.

Yao J (2017) Imaging Transcriptional Regulation of Eukaryotic mRNA Genes: Advances and Outlook. J Mol Biol 429:14–31. doi: 10.1016/j.jmb.2016.11.007

Yao J, Munson KM, Webb WW, Lis JT (2006) Dynamics of heat shock factor association with native gene loci in living cells. Nature 442:1050–1053. doi: 10.1038/nature05025

Yin L, Lazar MA (2005) The Orphan Nuclear Receptor Rev-erbα Recruits the N-CoR/Histone Deacetylase 3 Corepressor to Regulate the Circadian *Bmal1* Gene. Mol Endocrinol 19:1452–1459. doi: 10.1210/me.2005-0057

Yin L, Wu N, Curtin JC, et al (2007) Rev-erbalpha, a heme sensor that coordinates metabolic and circadian pathways. Science 318:1786–1789. doi: 10.1126/science.1150179

Yoo S-H, Yamazaki S, Lowrey PL, et al (2004) PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. Proc Natl Acad Sci U S A 101:5339–46. doi: 10.1073/pnas.0308709101

Yoshida M, Kijima M, Akita M, Beppu T (1990) Potent and specific inhibition of mammalian histone deacetylase both in vivo and in vitro by trichostatin A. J Biol Chem 265:17174–17179.

Yuan G-C, Liu Y-J, Dion MF, et al (2005) Genome-scale identification of nucleosome positions in S. cerevisiae. Science 309:626–630. doi: 10.1126/science.1112178

Yunger S, Rosenfeld L, Garini Y, Shav-Tal Y (2013) Quantifying the transcriptional output of single alleles in single living mammalian cells. Nat Protoc 8:393–408.

Yunger S, Rosenfeld L, Garini Y, Shav-Tal Y (2010) Single-allele analysis of transcription kinetics in living mammalian cells. Nat Methods 7:631–633. doi: 10.1038/nmeth.1482

Zamir I, Zhang J, Lazar MA (1997) Stoichiometric and steric principles governing repression by nuclear hormone receptors. Genes Dev 11:835–846.

Zhang R, Lahens NF, Ballance HI, et al (2014) A circadian gene expression atlas in mammals: Implications for biology and medicine. Proc Natl Acad Sci 111:16219–16224. doi: 10.1073/pnas.1408886111

Zhang Y, Fang B, Emmett MJ, et al (2015) GENE REGULATION. Discrete functions of nuclear receptor Rev-erbα couple metabolism to the clock. Science 348:1488–1492. doi: 10.1126/science.aab3021

Zhao H, Sifakis EG, Sumida N, et al (2015) PARP1- and CTCF-Mediated Interactions between Active and Repressed Chromatin at the Lamina Promote Oscillating Transcription. Mol Cell 59:984–997. doi: 10.1016/j.molcel.2015.07.019

Zhou BO, Zhou J-Q (2011) Recent transcription-induced histone H3 lysine 4 (H3K4) methylation inhibits gene reactivation. J Biol Chem 286:34770–34776. doi: 10.1074/jbc.M111.273128

Zhu B, Gates LA, Stashi E, et al (2015) Coactivator-Dependent Oscillation of Chromatin Accessibility Dictates Circadian Gene Amplitude via REV-ERB Loading. Mol Cell 60:769–783. doi: 10.1016/j.molcel.2015.10.024

Zhu X-D, Sadowski PD (1995) Cleavage-dependent Ligation by the FLP Recombinase: CHARACTERIZATION OF A MUTANT FLP PROTEIN WITH AN ALTERATION IN A CATALYTIC AMINO ACID. J Biol Chem 270:23044–23054. doi: 10.1074/jbc.270.39.23044

Zhu Y, van Essen D, Saccani S (2012) Cell-type-specific control of enhancer activity by H3K9 trimethylation. Mol Cell 46:408–423. doi: 10.1016/j.molcel.2012.05.011

Zoller B, Nicolas D, Molina N, Naef F (2015) Structure of silent transcription intervals and noise characteristics of mammalian genes. Mol Syst Biol 11:823.

Zopf CJ, Quinn K, Zeidman J, Maheshri N (2013) Cell-Cycle Dependence of Transcription Dominates Noise in Gene Expression. PLoS Comput Biol 9:e1003161. doi: 10.1371/journal.pcbi.1003161

Zullo JM, Demarco IA, Piqué-Regi R, et al (2012) DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina. Cell 149:1474–1487. doi: 10.1016/j.cell.2012.04.035

# Damien NICOLAS
## Molecular biologist

| | |
|---|---|
| Address: | Rue Pré-du-Marché |
| | 1004 Lausanne |
| | Switzerland |
| E-mail: | damien.nicolas07@gmail.com |
| Phone: | +41 79 512 39 48 |
| Data of birth: | March 7, 1988 |
| Nationality: | Swiss and Italian |

## Skills

- **Molecular biology and biochemistry:**
  Cellular biology (mammalian and prokaryotic cells), cloning, microscopy (fluorescence and luminescence, real-time and fixed cells), PCR (standard and quantitative), Chromatin Immunoprecipitation, CRISPR/Cas9 genome editing, single-molecule RNA-FISH

- **Languages:**

  French:   mother tongue
  English:   fluent
  German:   intermediate

- **Personal skills:**
  Independent, efficient, meticulous, organized, respect of deadlines, experience in work teams

- **IT skills:**
  Microsoft Office, notions of R and Illustrator

## Professional experience

| | | |
|---|---|---|
| 2012 – 2017 | **École Polytechnique Fédérale de Lausanne (EPFL)** Switzerland | **Thesis project** in Laboratory of Computational Systems Biology **Prof. Felix Naef** <br><br> "Rhythmic modulation of transcriptional burst frequency in circadian gene promoters" <br><br> *Characterization of Bmal1 transcriptional bursting properties along the circadian cycle and identification of molecular mechanisms modulating transcriptional bursting* |
| 2011 – 2012 | **Institut de Microbiologie (CHUV)** Switzerland | **Master project** in Laboratory of HIV viral-host interactions **Dr. Angela Ciuffi** <br><br> "Dissection of LEDGF/p75 enhancer" <br><br> *Characterization of the upstream enhancer participating in LEDGF/p75 transcription regulation and identification of the binding transcription factors* |
| 2010 | **Université de Lausanne (UNIL)** Switzerland | **Internship** in Center for Integrative Genomics **Prof. Winship Herr** <br><br> "Evolutionary conservation of the HCF-1 protease O-linked N-Acetylglucosamine transferase" <br><br> *Comparison of sequences to assess evolutionary properties of the HCF-1 enzyme* |

# CV Damien NICOLAS

**Education**    2012 –        **PhD** in Molecular Life Sciences
                               École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

                 2010 – 2012:  **Master** of Science in Molecular Life Sciences (MLS)
                               University of Lausanne (UNIL), Switzerland

                 2007 – 2010   **Bachelor** of Science in Biology
                               University of Lausanne (UNIL), Switzerland

## Scientific achievements

### Publications

Clock-dependent chromatin topology modulates circadian transcription and behavior
Mermet J, Yeung J, Hurni C, Mauvoisin D, Gustafson K, Jouffe C, **Nicolas D**, Emmenegger Y, Gobet C, Franken P, Gachon F, and Naef F
Submitted, 2017

What shapes eukaryotic transcriptional bursting?
**Nicolas D**, Phillips NE and Naef F
Mol Biosyst, 2017

Structure of silent transcription intervals and noise characteristics of mammalian genes
Zoller B, **Nicolas D**, Molina N and Naef F
Mol Syst Biol, 2015

CAST: An automated segmentation and tracking tool for the analysis of transcriptional kinetics from single-cell time-lapse recordings
Blanchoud S**, Nicolas D**, Zoller B, Tidin O and Naef F
Methods, 2015

Comparative genome analysis of Pseudomonas knackmussii B13, the first bacterium known to degrade chloroaromatic compounds
Miyazaki R, Bertelli C, Benaglio P, Canton J, De Coi N, Gharib WH, Gjoksi B, Goesmann A, Greub G, Harshman K, Linke B, Mikulic J, Mueller L, **Nicolas D**, Robinson-Rechavi M, Rivolta C, Roggo C, Roy S, Sentchilo V, Siebenthal AV, Falquet L and van der Meer JR.
Environ Microbiol, 2015

### Conferences

August 28-30 2016:  6th Single Molecule Localization Microscopy Symposium, EPFL, Switzerland
**Poster**: Dissecting the transcriptional bursting of a circadian gene promoter

October 20-23 2014:  2nd International SystemsX.ch Conference on Systems Biology, EPFL, Switzerland
**Poster**: Transcriptional bursting dynamics of circadian genes around the clock

May 4-7 2014:  Stochastic Biology: from Cells to Populations, IST, Klosterneuburg, Austria
**Poster**: Transcriptional bursting dynamics of circadian genes around the clock

## Additional experience

- Teaching: courses assistant at École Polytechnique Fédérale de Lausanne (2012-2016)
- Event organization: co-organizer of SV Happy Hour and SV Running (2014-2017)
- Health care: Red cross health auxiliary at Swiss Armed Forces (2012)