

# Comprehensive analysis of the circadian nuclear and cytoplasmic transcriptome in mouse liver

Clémence Hurni<sup>1</sup>, Benjamin Weger<sup>1</sup>, Cédric Gobet<sup>1</sup>, Felix Naef<sup>1</sup>

1: The Institute of Bioengineering (IBI), School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

## Abstract

In eukaryotes, RNA is synthesised in the nucleus, spliced, and exported to the cytoplasm where it is translated and finally degraded. Any of these steps could be subject to temporal regulation during the circadian cycle, resulting in daily fluctuations of RNA accumulation and affecting the distribution of transcripts in different subcellular compartments. Our study analysed the nuclear and cytoplasmic, poly(A) and total transcriptomes of mouse livers collected over the course of a day. These data provide a genome-wide temporal inventory of enrichment in subcellular RNA, and revealed specific signatures of splicing, nuclear export and cytoplasmic mRNA stability related to transcript and gene lengths. Combined with a mathematical model describing rhythmic RNA profiles, we could test the rhythmicity of export rates and cytoplasmic degradation rates of approximately 1400 genes. With nuclear export times usually much shorter than cytoplasmic half-lives, we found that nuclear export contributes to the modulation and generation of rhythmic profiles of 10% of the cycling nuclear mRNAs. This study contributes to a better understanding of the dynamic regulation of the transcriptome during the day-night cycle.

## Introduction

Every step of the RNA life cycle is tightly regulated to ensure proper cellular function, including transcription, co- and post-transcriptional splicing, nuclear export, and cytoplasmic degradation. Collectively, the balance of production and elimination rates determines the abundance of RNA species (e.g. pre-mRNA, polyadenylated pre-mRNA, mRNA) in a given subcellular compartment (e.g. nucleus, cytoplasm) (Chin & Lécuyer, 2017). Studies based on snapshot images of subcellular RNA distributions, either obtained by single-molecule RNA-FISH (smFISH) (Xia et al., 2019), or by RNA-seq of fractionated cells, allowed to characterise nuclear and cytoplasmic transcriptomes (Bahar Halpern et al., 2015; Bhatt et al., 2012; Chen & van Steensel, 2017; Djebali et al., 2012), revealing diverse patterns of subcellular localisation for different RNA subclasses. Interestingly, a study in the mouse liver showed that approximately 10% of protein coding transcripts were retained in the nucleus, efficiently buffering cytoplasmic noise associated with stochastic transcription (Bahar Halpern et al., 2015).

RNA processing rates, and particularly mRNA degradation rates have been estimated in cells either from transcriptional blockage experiments (Lam et al., 2001) or a pulse-chase strategy to label nascent transcripts during a short period of time (Schwanhausser et al., 2011). To estimate life times, RNA levels are typically fitted to exponential functions. Alternatively, mathematical modelling using ordinary differential equations to jointly describe pre-mRNA and mRNA levels can indirectly infer kinetic parameters of RNA processing steps in a label-free and genome-wide manner (Zeisel et al., 2011); however, dynamic datasets reflecting different cellular states are needed to identify the parameters, for instance during cell cycle, cell differentiation (Rabani et al., 2011), or at different circadian time-points (Lück et al., 2014; Wang et al., 2018). Our laboratory previously developed a mathematical model describing the temporal profiles of pre-mRNA and mRNA from total RNA-seq, which uses the periodicity of circadian cycles and the relationship between peak times and relative amplitudes to infer the production and the decay terms; this study allowed the identification of hundreds of post-transcriptionally generated rhythmic mRNA profiles (Wang et al., 2018).

Among mouse tissues, the liver has the highest number of cycling mRNA (Zhang et al., 2014), and the rhythmic patterns are driven by the endogenous circadian clock as well as by the fasting/feeding cycle (Damiola et al., 2000; Vollmers et al., 2009; Weger et al., 2021). The circadian clock is a self-sustained transcription-translation feedback oscillator (Lowrey & Takahashi, 2000; Shearman, 2000). In mammals, the network consists of the CLOCK/BMAL1 transcription factor dimer which promote the expression of its own repressors Cryptochromes (CRY1 and CRY2) and Periods (PER1 and PER2), generating cycling negative feedback loops with a period of approximately 24 hours. A second feedback loop is composed of the complex CLOCK-BMAL1 activating the expression of target genes such as *Reverba, $\beta$* , which repress the expression of BMAL1. The molecular clock also targets thousands of downstream *clock-controlled genes* to couple circadian cycles and physiological pathways.

In the circadian context, rhythmic transcription often underlies rhythmic gene expression, and has been extensively studied in mouse liver (Atger et al., 2015; Koike et al., 2012; Menet et al., 2012). Interestingly, rhythmic transcription does not account for the rhythmic accumulation of all mRNAs, implying regulation at the post-transcriptional level. Indeed, the rates regulating the RNA processing steps often adapt to external inputs such as temperature shifts (Gotic et al., 2016), nutrient levels (Bahar Halpern et al., 2015), inflammation (Bhatt et al., 2012), or time of the day (Lück et al., 2014; Wang et al., 2018). In principle, every step during the RNA life-cycle could be regulated temporally, contributing to rhythmic gene expression. The extent of which post-transcriptional regulation contributes to circadian gene expression is debated, varying from 15% (Atger et al., 2015; Martelot et al., 2012) to 70% (Koike et al., 2012). Conversely, many rhythmically transcribed genes do not have corresponding cycling mRNA (Koike et al., 2012) or protein (Wang et al., 2017), suggesting that the oscillatory amplitudes are damped as rhythm propagates from early to later steps of gene expression (Mermet et al., 2017). Mathematical models showed that temporal regulation of mRNA stability (the inverse of the degradation rate) offers flexibility to boost amplitudes and fine-tune phases (Lück et al., 2014; Martelot et al., 2012; So & Rosbash, 1997; Wang et al., 2018). Rhythmic mRNA stability can be modulated through the activity of plethora of RNA-binding proteins (Benegiamo et al., 2018; Lee

et al., 2012; Morf et al., 2012), through silencing by miRNA (Du et al., 2014), by regulating poly(A) tail length (Kojima et al., 2012), and regulating deadenylase activity (Baggs & Green, 2003). In the nucleus, processing rates, including export rates, can be modulated by N6-methyladenosine methylation (Fustin et al., 2013), or through rhythmic retention in nuclear paraspeckles (Benegiamo et al., 2018; Torres et al., 2016).

This study aims at understanding the circadian dynamics of RNA regulatory steps in mouse liver. More specifically, we focus on nuclear export, cytoplasmic degradation, and to some extent on RNA splicing. We performed RNA-sequencing of polyadenylated and ribo-depleted total RNA purified from nuclear and cytoplasmic fractions of mouse liver samples collected every 4 hours along a full day cycle. This dataset provides a genome-wide and temporal inventory of RNA levels in two subcellular compartments. At the static (non-temporal) level, we found a relationship between the transcript length and the localisation, with long transcripts being enriched in the nucleus. Additionally, gene length influences the proportion of co- and post-transcriptional splicing. By modeling the rhythmic profiles (amplitudes and phases) of nuclear pre-mRNA, nuclear mRNA, and cytoplasmic mRNA, we estimated the export rates and cytoplasmic degradation rates of approximately 1400 genes. We additionally uncovered rhythmic patterns in the nucleus that most likely originate from a rhythmic nuclear export. Moreover, this post-transcriptional regulation step not only contributes to amplifying the amplitude of highly rhythmic genes but also generates *de novo* rhythms, and is predicted to influence the temporal profiles of approximately 10% of the rhythmic nuclear transcriptome.

## Results

### **Global characteristics of Poly(A) and Total RNA populations in hepatic nuclear and cytoplasmic fractions**

To characterise the circadian kinetics of RNA regulatory processes occurring in different cellular compartments, we isolated RNA from nuclear (Nuc / N) and cytoplasmic (Cyt / C) fractions from mouse livers sampled every 4 hours during two consecutive days (Fig 1A). We sequenced

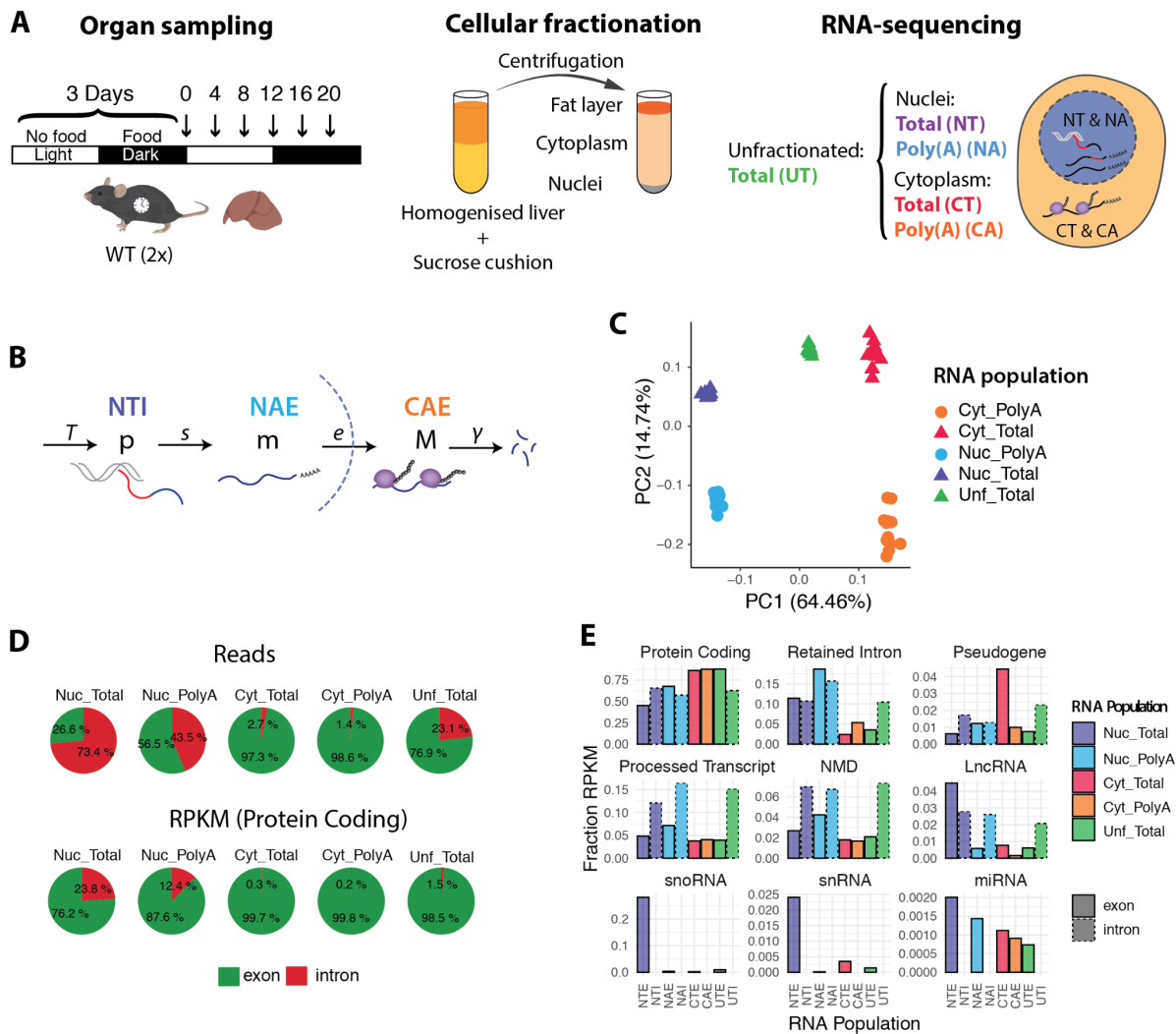
ribosome-depleted total RNA (T) and polyadenylated RNA (A). As a control, we also extracted RNA from unfractionated samples (Unf / U). We quantified the transcript levels using the pseudo-alignment algorithm implemented in Kallisto (Bray et al., 2016), by providing as a reference the annotation of transcript sequences (isoform) of both pre-mRNA (intronic and exonic regions, labelled I) and mRNA (exonic regions only, labelled E). Nuclear Total pre-mRNA (NTI), Nuclear Poly(A) mRNA (NAE), and Cytoplasmic Poly(A) mRNA (CAE) were used in our simplified model describing the sequential events occurring along the RNA lifecycle (Fig 1B). This includes the steps of splicing and polyadenylation of pre-mRNA, export of nuclear mRNA, and degradation of cytoplasmic mRNA. We reduced the complexity of the RNA-seq quantification performed at the transcript-level by summing the counts of all the isoforms of a gene belonging to the same biotype (Table S1, defined in Ensembl, see Methods). Additionally, we summed the counts assigned to all isoforms of the gene (Table S1). The counts estimated by Kallisto were further normalised by the gene length to obtain a quantification akin to reads per kilobase per million (RPKM) (Table S2 and S3, see Methods).

Principal Component Analysis (PCA) on the mRNA measurements shows that the first two components readily split the types of samples: indeed, subcellular localisation explains most of the variation between samples (64%, Fig 1C). Unfractionated samples lie between nuclear and cytoplasmic samples on the first principal component (PC1), although closer to cytoplasmic RNA. PC2 accounts for 17% of the variability, which is mainly related to differences in the type of RNA (Poly(A) versus Total). In comparison the temporal variation is significantly smaller.

The Nuclear Total RNA population (NT) includes all the transcripts found in the nucleus, from nascent transcripts to fully transcribed polyadenylated mature RNAs. As expected, the proportion of intron versus exon reads was the highest in the NT sample, and lowest in the cytoplasmic fraction (Fig 1D). In terms of reads (reflecting the mass of the corresponding RNA), approximately 75% of the reads align to pre-mRNA sequences. These proportions are similar to what has been previously described in mouse liver for nuclear total RNA (Benegiamo et al., 2018). Nuclear Poly(A) RNA population (NA) represents the population of fully-transcribed and polyadenylated transcripts. Interestingly, 45% of these reads map on pre-mRNA (Fig 1D). When reads are normalised by the gene

length (RPKM), this showed that about 1 in 8 polyadenylated transcripts in the nucleus were still not fully spliced. Note that in Fig 1D, we only reported fractions of protein coding transcripts to avoid biases associated with the different compositions of RNA biotypes in different fractions (Fig 1E). Indeed, even if most of the detected RNA are protein coding transcripts, making up to 85% of mRNA molecules in the cytoplasm and in the whole cell, nuclear RNA populations are more diverse. The nucleus is particularly enriched for functional non-coding RNAs such as snoRNA, and for isoforms with a retained intron (10 to 20% of the nuclear RNA fraction). Finally, as expected, pre-mRNA represents less than 3% of the cytoplasmic RNA fraction.

In light of our results, we conclude that our method of fractionating liver cells performed as expected, and that our analysis of RNA distribution in the nucleus and cytoplasm supports previous findings (Bahar Halpern et al., 2015; Bhatt et al., 2012; Chen & van Steensel, 2017; Djebali et al., 2012; Fazal et al., 2019).



**Fig 1. Global characteristics of Poly(A) and Total RNA populations in nuclear and cytoplasmic fractions.**

**A:** Experimental design: 12 WT mice were kept in a 12:12 light-dark cycle. Food was given only during the dark phase (active phase). The night-restricted feeding regimen started 3 days prior to the sampling. Livers were collected every 4 hours ( $n = 2$ ), homogenised, and centrifuged in a sucrose cushion as described in (Menet et al., 2012). RNA from cytoplasmic and nuclear fraction were extracted. rRNA-depleted Total RNA (T) and Polyadenylated RNA (A) from nuclear and cytoplasmic fractions were sequenced. Total RNA from unfractionated liver tissues was also sequenced as an internal reference set. **B:** Representation of each RNA processing step modeled in this study: starting with the transcription by RNA Polymerase II (at a rate  $T$ ), splicing of the introns (red segment) and addition of a poly(A) tail at the 3' end ( $s$ ), export to the cytoplasm ( $e$ ), and degradation ( $\gamma$ ).  $p$  = pre-mRNA,  $m$  = nuclear polyadenylated mRNA,  $M$  = cytoplasmic mRNA. **C:** Principal Component Analysis using mRNA (signal from exonic regions only). PC1 separates nuclear from cytoplasmic samples, and PC2 separates Total from Poly(A) samples. **D:** Percentage of reads mapping on pre-mRNA or mRNA transcripts in each condition (up). Proportion of RPKM (reads normalised by gene length) of protein coding transcripts mapping on

pre-mRNA and mRNA (bottom). **E**: Fraction of each biotype in each RNA sample. Dotted lines represent pre-mRNA. Pre-mRNA from cytoplasmic fractions is not shown.

## **Static relationships between levels of successive RNA species and dependencies on cytoplasmic RNA stability, nuclear export and splicing rates**

Before analysing temporal aspects of these datasets, we first studied the relationships between the temporal averages of the various RNA species, in particular properties related to cytoplasmic RNA stability, nuclear export, and splicing.

First, we compared nuclear (NAE) and cytoplasmic (CAE) polyadenylated mRNAs. As noted in a previous study (Bahar Halpern et al., 2015), in the simplest scenario where there is no degradation in the nucleus, the steady-state ratio of NAE over CAE reflects the ratio of two parameters: the nuclear export rate and the cytoplasmic degradation rate (Fig 1B). A fast export from the nucleus to the cytoplasm would deplete the amount of RNA in the nucleus, and a long cytoplasmic half-life in the cytoplasm would lead to an accumulation of mRNA in the cytoplasm. On the contrary, RNAs with a long export time (or long nuclear retention) and short half-life in the cytoplasm would increase the nuclear to cytoplasmic ratio, resulting in an apparent nuclear accumulation.

Across all genes, the expression levels of NAE span  $8.8 \log_2$  RPKM ( $10^{2.6}$  fold). CAE has a wider distribution of expression level of  $12.5 \log_2$  RPKM (about  $10^4$  fold, Fig 2A). The most abundant transcripts code for secreted proteins such as Murine Urinary Proteins, Albumin, and Apolipoproteins, and are enriched in the cytoplasm. The NAE/CAE ratio varies by a factor of almost 2500, with a median  $\log_2(\text{NAE}/\text{CAE})$  around -2.3, indicating that RNA transcript counts are more abundant in the cytoplasm than in the nucleus. The nuclear-enriched RNA population includes the long non-coding RNAs *Malat1* and *Neat1*, which form the backbones of nuclear speckles and nuclear paraspeckles (Clemson et al., 2009; Hutchinson et al., 2007). Moreover, a fraction of transcripts are also found enriched in the nucleus, consistently with a previous study (Bahar Halpern et al., 2015) showing that in the mouse liver, protein coding genes such as *Nlrp6* (here  $\log_2$ -ratio = 1.6), and glucose



metabolism-related genes *Mlxipl* ( $\log_2(\text{NAE}/\text{CAE}) = 3.4$ ), *Gcgr* ( $\log_2(\text{NAE}/\text{CAE}) = -0.1$ ), and *Gck* ( $\log_2(\text{NAE}/\text{CAE}) = -1.2$ ) were indeed enriched in the nucleus.

In our simplified model, the steady-state  $\log_2(\text{NAE}) = \log_2(\text{Transcription rate}) + \log_2(\text{export time})$ , and  $\log_2(\text{CAE}) = \log_2(\text{Transcription rate}) + \log_2(\text{half-life})$ . The fact that  $\log_2(\text{NAE})$  and  $\log_2(\text{CAE})$  strongly co-vary, and that the transcription rate is added to both, implies that this process carries the largest variation, as expected. We next aimed to determine which molecular process co-varied most with the NAE/CAE ratio. Fig 2B shows the  $\log_2$ -ratio of NAE/CAE as a function of either  $\log_2(\text{NAE})$  or  $\log_2(\text{CAE})$ , and the correlation coefficient is higher for CAE than for NAE ( $r = -0.67$  for CAE,  $r = 0.06$  for NAE). The strong correlation coefficient with NAE and absence of correlation with CAE suggests that the cytoplasmic half-life has a higher influence on the ratio than the export time. We verified the subcellular localisation of a known nuclear-biased transcript (*Mlxipl*) by smFISH, and its nuclear accumulation was evident, although not precisely quantifiable due to its high abundance and crowdedness of mRNA molecules (Fig 2C). On the opposite side, *Actb* is a typical example of a cytoplasmically enriched transcript ( $\log_2(\text{NAE}/\text{CAE}) = -3.64$ ), which was verified by smFISH, showing a  $\log_2(\text{NAE}/\text{CAE})$  that varied between time points, ranging from -1.42 to -3.8, with a median of -2.7.

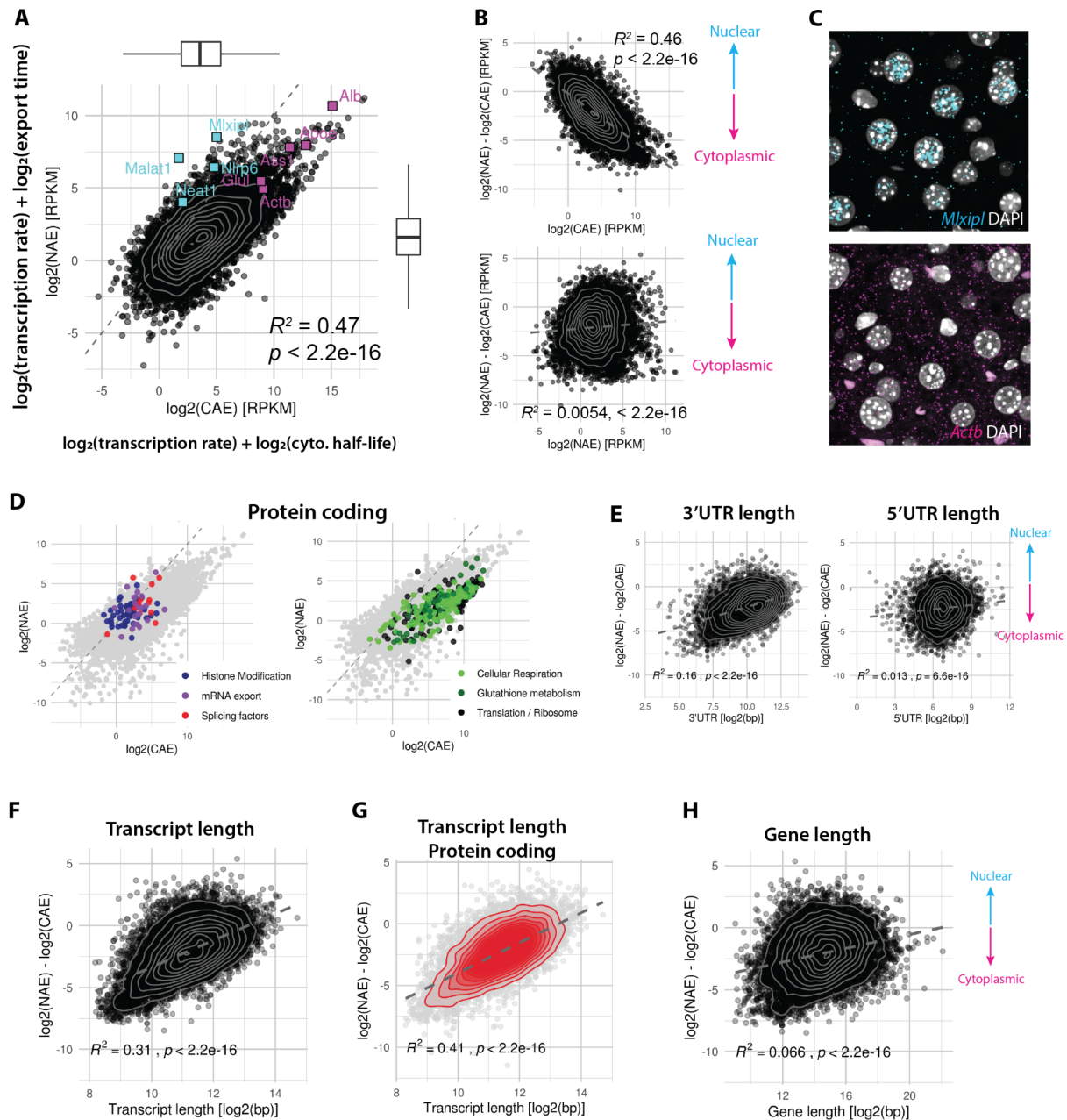
To identify the biological functions of nuclear and cytoplasmic-enriched RNA populations, we performed a Gene Ontology analysis on nuclear and cytoplasm-biased protein coding transcripts. This showed a previously reported concordance between protein and RNA localisation (Fazal et al., 2019). Here, we found that nuclear transcripts coded for epigenetic modifiers, mRNA processing (splicing factors), and export, while proteins encoded by cytoplasmic RNA were involved in translation, oxidative phosphorylation, fatty acid  $\beta$ -oxidation and Phase-II detoxification process (Fig 2D, Table S4).

We next examined several genomic features that may modulate kinetics (transcription rate, export, decay) and as a consequence mRNA localisation, such as gene length, transcript length, number of exons, 3'UTR and 5'UTR length (Chen & van Steensel, 2017; Fazal et al., 2019; Zuckerman &

Ulitsky, 2019). The 3'UTR length (but not the 5'UTR length) showed a positive correlation with the nuclear to cytoplasmic ratio ( $r = 0.39$ ), indicating that transcripts with longer 3'UTR sequences are enriched in the nucleus (Fig 2E). The length of the 3'UTR is often proposed as a predictor of mRNA stability, because of the presence of many regulatory elements. However, both stabilising and destabilising RNA-Binding Proteins and miRNAs often compete for the same mRNA substrate (Garneau et al., 2007), which in the end results in a conflicting or undetectable effect of the 3'UTR length on processing rates (Chen & van Steensel, 2017). Strikingly, we found that spliced transcript length (corresponding to the total exon length) has a significant positive correlation with nuclear localisation ( $r = 0.5$ , Fig 2F), and the correlation is improved when focussing only on protein coding transcripts (Fig 2G,  $r = 0.65$ ). As transcripts of hundreds kilobases take several hours to be fully transcribed (Jackson et al., 2000), this could explain the longer residence time in the nucleus. However, in that case, we would expect a correlation with gene length as well, which was not observed (Fig 2H), suggesting that the structure of the mature mRNA rather than transcription plays a role in the differential localisation.

We next performed an analogous analysis to compare nuclear pre-mRNAs and nuclear polyadenylated mRNAs, estimated from NTI, and NAE, respectively (Fig S1). We found a strong correlation between NTI and NAE (Fig S1A,  $r = 0.77$ ), suggesting that transcription rate, which is the common factor to both NTI and NAE ( $\log_2(\text{NTI}) = \log_2(\text{Transcription rate}) + \log_2(\text{splicing time})$ ), has a larger effect on NTI and NAE levels than export and splicing rates. The fact that the correlation is stronger here than for NAE/CAE ( $r = 0.77$  vs.  $r = 0.69$ ) suggests that the variability (range across all genes) in splicing times is smaller than that of cytoplasmic life times. Here, we found that the  $\log_2(\text{NTI}/\text{NAE})$  is more correlated with the export time (NAE) than with the splicing time (NTI) (Fig S1B,  $r = -0.53$  for NAE,  $r = 0.12$  for NTI). This observation suggests that, on average, the abundance of spliced nuclear mRNA (NAE) is more strongly determined by the variation of nuclear export time rather than splicing time of the pre-mRNA. However, transcription, splicing and export are not always independent processes and can be coupled and thus, the interpretation of the correlations may be more complicated (Hocine et al., 2010; Müller-McNicoll & Neugebauer, 2013; Palazzo & Lee, 2018). Therefore, the anti-correlation of

$\log_2(\text{NAE})$  (defined as  $\log_2(\text{transcription rate}) + \log_2(\text{export time})$ ) with  $\log_2(\text{NTI} / \text{NAE})$  (defined as  $\log_2(\text{splicing time}) - \log_2(\text{export time})$ ) could also be interpreted as a high transcription rate promoting rapid splicing.



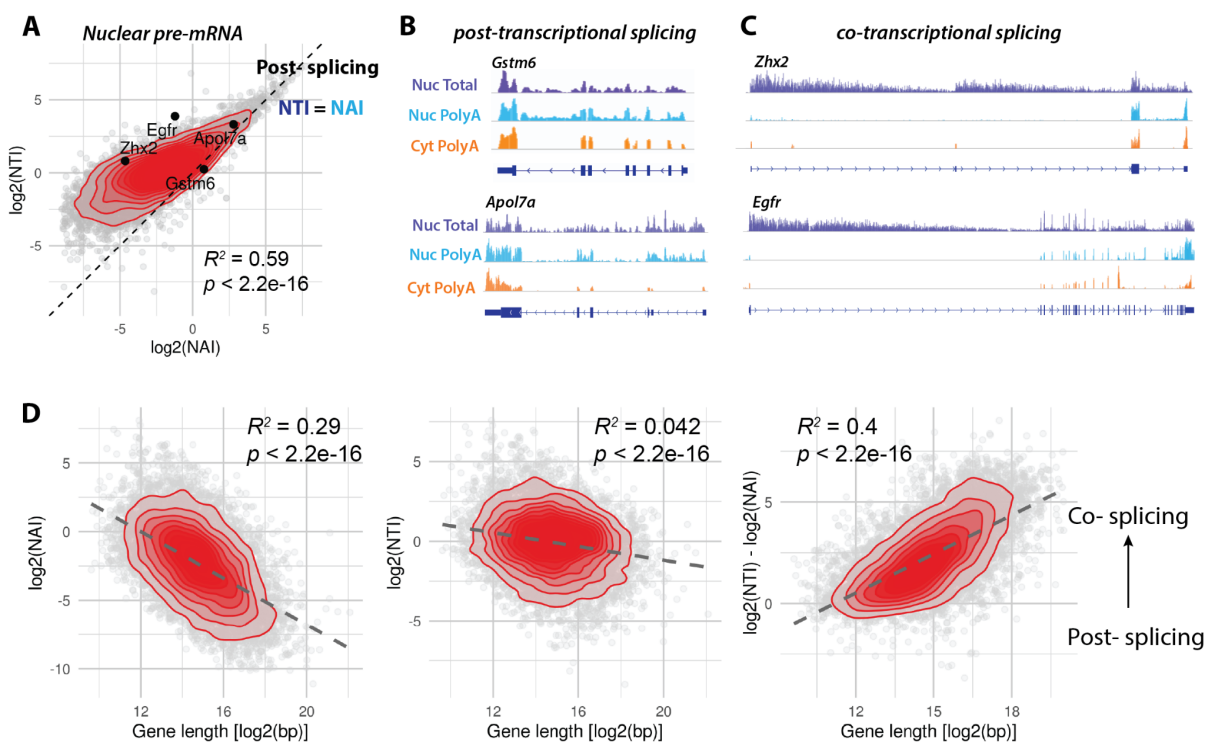
**Fig 2. The ratio of nuclear and cytoplasmic mRNA abundance is related to transcript length.** **A:** Expression level of nuclear mRNA (NAE) versus cytoplasmic mRNA (CAE) in  $\log_2(\text{RPKM})$ , averaged over 6 time-points and 12 animals. In our model, the level of NAE corresponds to the transcription rate plus the nuclear half-life, and the level of CAE to the transcription rate plus cytoplasmic half-life (all in  $\log_2$ -scale). Cyan squares represent known nuclear genes and magenta squares cytoplasmic genes (Bahar Halpern et al., 2015). Dashed line indicates the

1:1 ratio. Note here that the diagonal line has been scaled with the dataset (see Methods). Grey contour lines: 2D kernel density estimation. Number of genes = 10914. **B**: NAE / CAE ratio in  $\log_2$ -scale against the average expression in the cytoplasm (up) or in the nucleus (bottom). **C**: smFISH of the *Mlxip1* and *Actb* transcripts in liver FFPE sections, at ZT12. Nuclei are stained with DAPI (grey). **D**: Transcripts coding for proteins involved in histone modifications (transcripts of the following gene families are shown as dots KDMs, KMTs, KMKs, OGTs, PRDMs, SETs), mRNA export (TREX, THO/TREX complex, NXF1, CPSF complex), and splicing (SR family members) preferentially localise in the nucleus. Transcripts coding for proteins involved in house-keeping functions such as cellular respiration, translation (RPS, RPL, EIF) and detoxification (GSTs, GPxs) are enriched in the cytoplasm. **E, F, G, H**:  $\log_2(\text{NAE}/\text{CAE})$  ratio against the following genomic features: 5'UTR length and 3'UTR length (E), transcript length (spliced isoform, F, G) and gene length (exon + intron, H). These features are transcript-specific, and were averaged based on the relative expression levels of each transcript to obtain a value per gene (E,F,H) or per biotype (G). Grey dashed line is the regression line.

### **The extent of co- versus post-transcriptional splicing increases with transcript length**

In case of complete co-transcriptional splicing, Nuclear Total pre-mRNA (NTI) represents the population of nascent transcripts (Furlan et al., 2020; Gaidatzis et al., 2015; Koike et al., 2012). However, in Fig 1D, we observed that at the gene-level, approximately 12% of the nuclear polyadenylated transcriptome encoding for proteins are pre-mRNA (captured in NAI samples), suggesting that a fraction of pre-mRNA is spliced post-transcriptionally. To estimate the amount of co- and post-transcriptional splicing, we compared the levels of NTI to NAI (Fig 3A). If the expression level of NAI is close to NTI, then, most of the nuclear pre-mRNA are fully transcribed and polyadenylated before being spliced. Conversely, a higher level of NTI compared to NAI means that nascent pre-mRNAs make up most of NTI and introns are spliced co-transcriptionally. Note that technically, we cannot determine if the introns were removed while transcription was still ongoing, or removed on an extremely fast time-scale post-transcriptionally. However, co-transcriptional splicing creates a typical sawtooth pattern of decreasing density of reads mapping on intron, with higher signal toward the 5'-end, described in nascent-seq datasets (Ameur et al., 2011; Menet et al., 2012). We indeed visually observed the sawtooth patterns on genes such as *Egfr* or *Zhx2* (Fig 3B), while these were absent on transcripts on the identity line (NAI = NTI), such as *Gstm6* or *Apol7a* (Fig 3C). Using the NAI/NTI ratio as proxy of co- versus post-transcriptional splicing, we observed a striking

correlation of the ratio with the gene length, such that transcripts originating from a long gene sequence are more likely co-transcriptionally spliced, while short genes are more post-transcriptionally spliced. Note that here we do not consider gene structure or distinguish the order of introns. This length-dependency may reflect the situation where transcription and 3' end processing is already terminated by the time the spliceosome assembles and splices newly transcribed introns. If splicing and transcription are two processes occurring concurrently but without further dependency, the longer the gene, and particularly the longer the downstream exon, the higher the probability for the upstream intron to be detected and spliced. NTI alone, representing the total pool of pre-mRNA in the nucleus, is not influenced by the gene length (Fig 3D). The comparison of the expression level of nuclear pre-mRNA in both Total and Poly(A) conditions is therefore informative about the extent of co- and post-transcriptional splicing. The splicing regime is thus strongly influenced by the gene length, or more precisely by the relation of splicing and elongation time.



**Fig 3. The extent of co- versus post-transcriptional splicing increases with transcript length.** **A:** Comparison of pre-mRNA expression levels in Poly(A) et Total nuclear RNA populations:  $\log_2(NTI)$  against  $\log_2(NAI)$ . Transcripts on the identity line ( $NAI = NTI$ , dashed line) are not spliced co-transcriptionally, while those away from the line are. Genes whose genome track are shown in B and C are shown on the plot. Red: 2D kernel

density. Only protein coding transcripts are shown. **B**: Genome track view of short genes that are mainly post-transcriptionally spliced. Gene length of *Gstm6*: 4.9kb, *Apol7a*: 11kb. **C**: Genome track view of long genes that are mainly co-transcriptionally spliced. Gene length of *Zhx2*: 145kb, *Egfr*: 166 kb. Genome tracks are not to scale. **D**: Correlations of the gene length (exon + intron) with  $\log_2(\text{NAI})$ ,  $\log_2(\text{NTI})$  and  $\log_2(\text{NTI}/\text{NAI})$ . The latter is related to the fraction of co- and post-transcriptional splicing. Overall, long genes are more frequently co-transcriptionally spliced than short genes.

## **Model-based identification of regulatory steps driving rhythmic cytoplasmic mRNA profiles**

The accumulation of RNA transcripts in any of the subcellular compartments results from the balance between production and elimination. At steady-state (Fig 2, Fig S1), it was generally not possible to identify the two respective contributions. However, as we showed before, it is possible to determine these rates from time-series profiles when genes are rhythmic, using a mathematical modelling approach (Wang et al., 2018). In essence, by modelling the profile shapes (indirectly the time delays and relative amplitude) of causally related and oscillating RNA species, for example intronic and exonic reads from total unfractionated RNA-seq, the approach can infer production and degradation rates. We applied this method to two situations. A joint model of mRNA accumulation in the cytoplasm (CAE) and the nucleus (NAE) allowed us to estimate the cytoplasmic stability and assess rhythmic degradation processes (hereafter referred to as step 2, Fig 4A). Using the fractionated cell data, we now can specifically address cytoplasmic half-life instead of estimate an effective half-life that encompasses all intermediate processes occurring in the nucleus (Wang et al., 2018). Next, we modelled the expression profiles of nuclear mRNA (NAE), which increases due to splicing of pre-mRNA (NTI) and decreases due to nuclear export (step 1). This strategy allowed us to infer the nuclear export rates and to uncover rhythmically regulated exports. Each step is described by an ordinary differential equation, where the first variable (NTI in step 1, NAE in step 2) and the rate of degradation (export  $e$  in step 1, cytoplasmic degradation  $\gamma$  in step 2) are either constant or rhythmic (defined by a cosine function with a phase and relative amplitude). The combination of constant or rhythmic production and degradation terms generates four different kinetic models (M1 to M4, Fig 4B). The optimal model is selected by combining a maximum-likelihood approach with the Bayesian

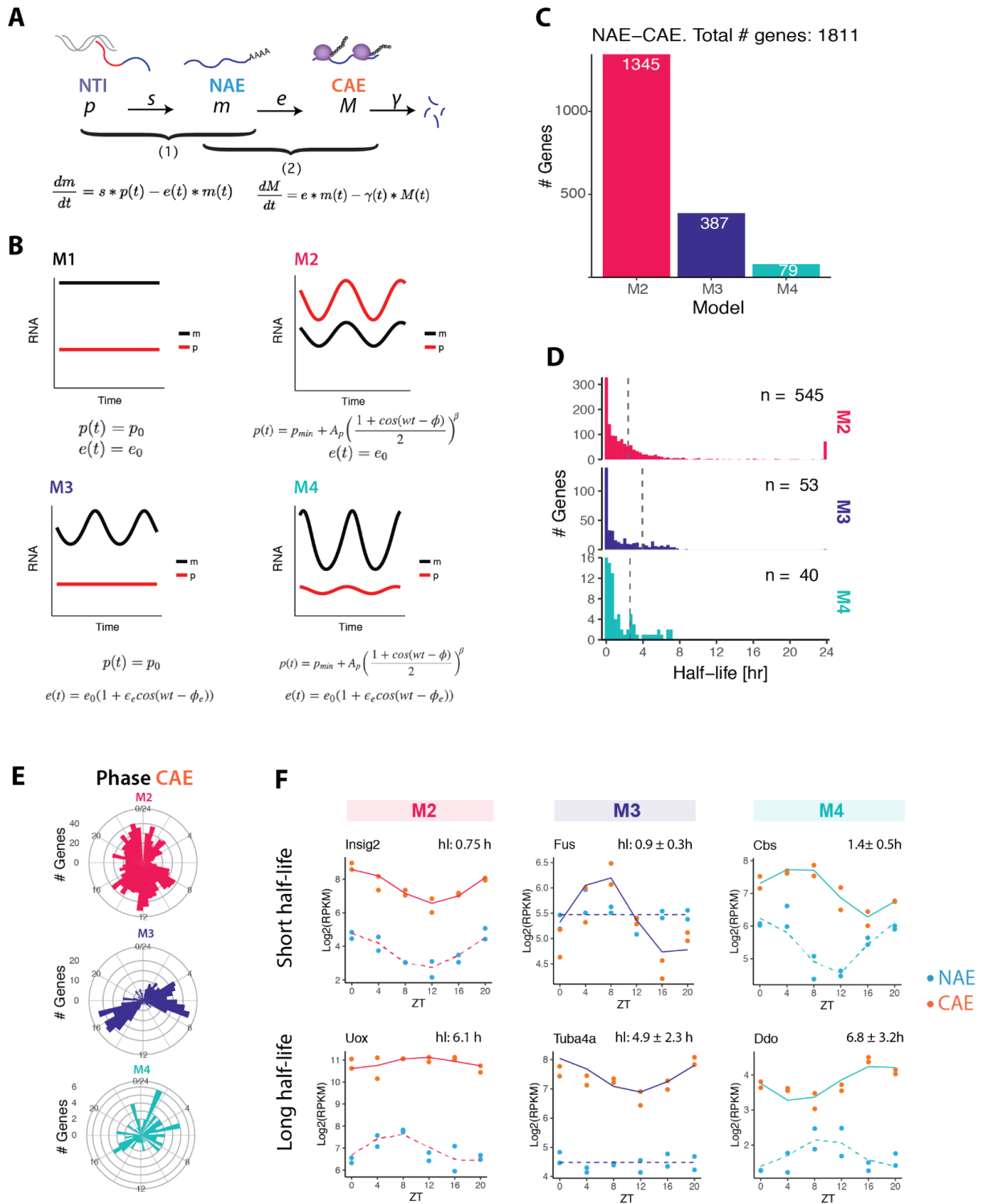
information criterion (BIC) to control for model complexity. For genes with constant levels of pre-mRNA and mRNA (M1), only the ratio between the production and degradation can be determined.

We first applied the mathematical model to step 2 (NAE-CAE) to identify rhythmic cytoplasmic half-lives. Out of the approximately 11,000 genes expressed in the liver, we confidently classified 1,424 genes as having a rhythmic accumulation of nuclear RNA (M2 and M4), of which 79 are additionally rhythmically degraded (M4, Fig 4D). The rhythmic profiles of 386 genes in the cytoplasm are solely due to the temporal regulation of the cytoplasmic half-life (M3). Thus, approximately 15% of expressed genes have a rhythmic temporal profile in at least one cellular compartment. Additionally, 25% are degraded rhythmically, in line with previous studies in the mouse liver (28% in (Wang et al., 2018), 30% in (Lück et al., 2014)). The median half-lives are 2.45 h for M2, 3.88 h for M3 and 2.36 h for M4 (Wang et al., 2018).

The distribution of peak mRNA times in the cytoplasm differs between the models. While M2 peak times were homogeneously distributed throughout the day, M3 exhibited a biphasic phase pattern with distinct associated biological functions. M3 genes peaking between ZT3 and ZT9 are involved in protein degradation such as proteasome components *Psmc3*, *Psmf1*, *Adrm1*. A second group of M3 associated genes peak between ZT14 and ZT18. This group comprises genes related to mRNA processing such as the splicing factors *Srsf1*, *Srsf3*, *Sfpq*, or *Thoc7* and *Hnrnp1*.

In Fig 4D, we show representative genes to illustrate the typical behavior of each model. When degradation is constant (M2), the more stable is the transcript, the more the mRNA oscillations dampen, and the longer the peak time difference between NAE and CAE. Note that in this model, this time delay cannot exceed 6 hours (it reaches 6 hours for infinitely long half-lives, (Martelot et al., 2012)). Transcripts with a short half-life have similar peak times (no delay) and relative amplitudes in the nucleus and in the cytoplasm (Fig 4F). Genes classified in M3 are rhythmic only in the cytoplasm. *Fus* is a well-known example of a gene rhythmically regulated at the post-transcriptional level (Martelot et al., 2012; Wang et al., 2018). Finally, genes in M4 are rhythmically regulated both at the

transcription and degradation level, resulting in more complex patterns, for instance a higher relative amplitude in the cytoplasm (*Ddo*, half-life 7 h) or a longer time delay (*Cbs*, half-life 1.4 h).



**Fig 4. Model-based identification of regulatory steps driving rhythmic cytoplasmic mRNA profiles. A:** Kinetic model describing the temporal accumulation of pre-mRNA and mRNA. RNA processing steps are split in



two steps: first, we describe the temporal accumulation of nuclear mRNA (NAE) in function of splicing rate  $s$  and export rate  $e$  (step 1), and second, the temporal accumulation of cytoplasmic mRNA (CAE) in function of export rate  $e$  and degradation rate  $\gamma$  (step 2). **B**: The combination of constant or rhythmic pre-mRNA ( $p$ ) and constant or rhythmic export in step 1 generates four models. Rhythmic pre-mRNA profile is described by a cosine function with a period of 24 hours, raised to a power  $\beta$  to account for the peaked oscillatory profiles. The same equations apply to step 2, but with  $\gamma$  instead of  $e$ ,  $m$  instead of  $p$ , and  $M$  instead of  $m$ . **C**: Classification of genes in step 2: M2 (rhythmic NAE, constant degradation), M3 (constant NAE, rhythmic degradation), or M4 (rhythmic NAE, rhythmic degradation). **D**: Distribution of estimated cytoplasmic half-lives. To facilitate identification and considering biologically plausible ranges, we set the range of possible half-lives from 10 min to 24 h. During the optimisation,  $\gamma = \log(2)/\text{half-life}$  sometimes reaches our fixed upper limits or lower limits. The median half-life (dotted line) is calculated excluding half-lives reaching the lower or upper boundary. **E**: Circular histogram showing the phases (peak times) of CAE in each model. **F**: Representative temporal profiles (NAE and CAE) of genes in M2, M3 and M4 genes with short half-life (up) or long half-life (bottom).

### **Kinetic model identifies rhythmically exported nuclear mRNAs**

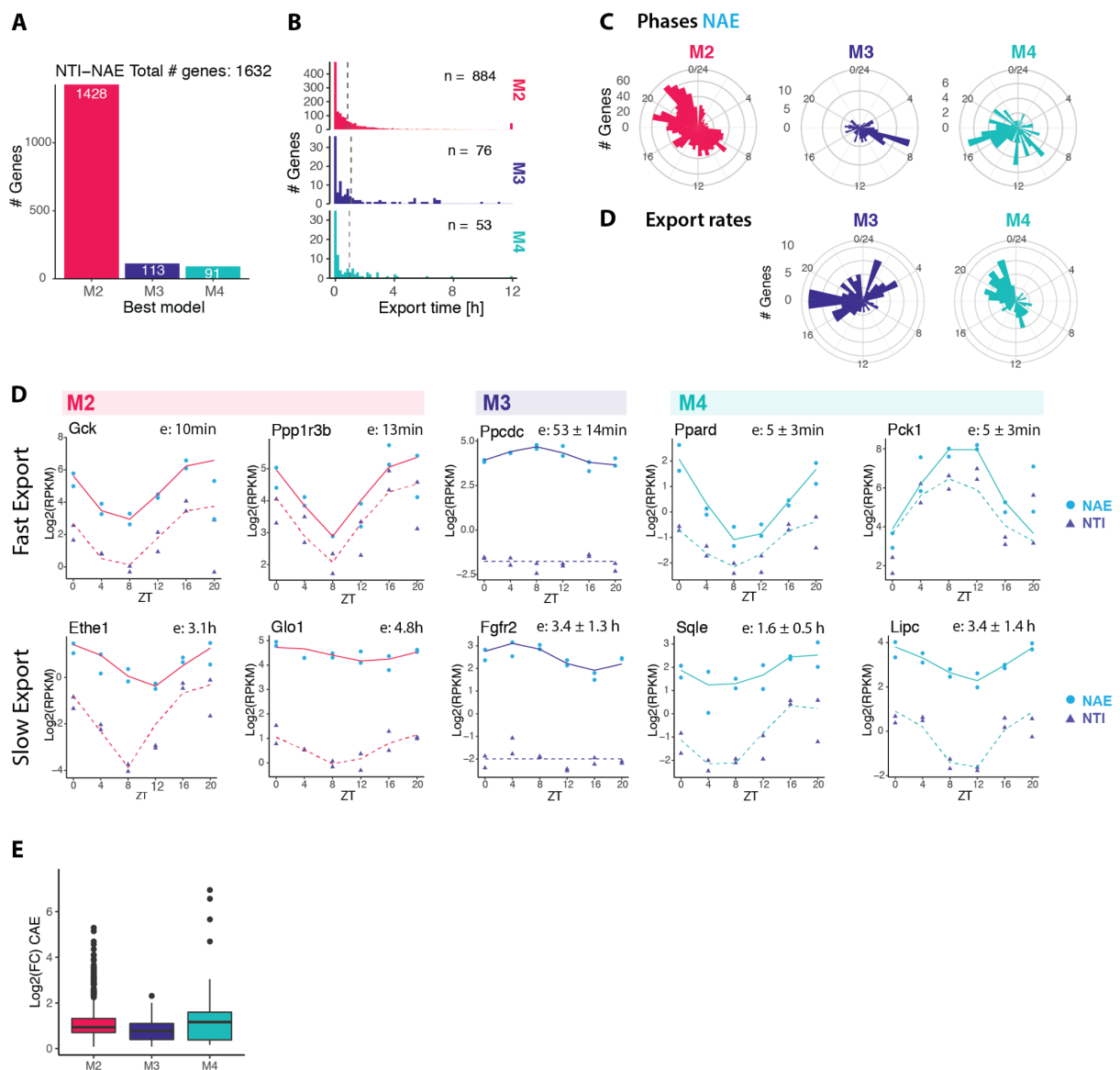
To uncover potentially rhythmic patterns generated by rhythmic nuclear export, we next applied the mathematical model used above to compare the rhythmic profiles of nuclear pre-mRNA (NTI) and nuclear mRNA (NAE). Here, the export time refers to the time spent by a fully transcribed, spliced, and polyadenylated transcript in the nucleus. We confidently classified 1632 genes in one of the three models with rhythmic NAE (Fig 5A). The vast majority were rhythmically transcribed and exported with a constant rate (1428 genes as M2). 13% were exported rhythmically (M3 and M4), suggesting that nuclear export is a post-transcriptional mechanism contributing to modulation and generation of rhythmic profiles. Representative profiles of genes classified as M3, whose rhythmic nuclear mRNA accumulations were only attributed to rhythmic mRNA export, are shown in Fig S2. Proteins encoded by these genes cover a diverse biological functions, such as deacetylation (*Lypla2*, *Sirt5*), nuclear speckles assembly and splicing (*Srrm2*, *Akap8l*), hypermethylation of sn(o)RNA m<sup>7</sup>G cap (*Tgs1*), or NAD(P)H-dependent oxidoreduction (*Rdh13*). Rhythmic export generated nuclear rhythms that propagated to the cytoplasm (Fig S2). However, the median  $\log_2$  fold-change of CAE (cytoplasmic mRNA) is lower than those originating from M2 or M4 genes (Fig 5E), but is still near 1 (2-fold

amplitude). While some oscillations are damped in the cytoplasm because of the long stability of cytoplasmic mRNA, others keep oscillating with relatively large amplitudes.

The distribution of NAE peak times in M3 genes showed an enrichment around ZT8 (Fig 5C, 33% of the M3 genes peak in the time window between ZT6 and ZT10). By contrast, peak times of M2 genes, whose rhythms are generated by rhythmic transcription, exhibited the typical bimodal distribution of phases peaking at the end of the light phase (ZT22) and the dark phase (ZT10), reflecting the two waves of transcriptional activity in the liver (Hughes et al., 2009; Martelot et al., 2012; Sobel et al., 2017). Phases of M4 genes, whose nuclear accumulation is also regulated by rhythmic export, were distributed throughout the day. The time of shortest export time (highest export rate) was enriched in the second part of the night (half of the M4 genes had the highest export rate between ZT18 to ZT 0). The difference between the peak time of export rate and peak of transcription rate modulates the resulting pattern of NAE: rhythmic amplitudes of metabolic transcripts such as *Pck1*, *Lpin1* or *Ppard* are amplified by antiphase rhythmic export, while genes such as *Sqle* or *Lipc* are damped while maintaining a phase coherence because export peak time matches transcription peak time (Fig 5D). Thus, the amplitudes of several key metabolic enzymes were boosted by regulating the export process in a timely manner.

Nuclear residence times are known to be typically shorter compared to cytoplasmic half-lives. Indeed, previous studies estimated that nuclear lifetime ranges from 5 min to less than a couple of hours (maximal estimated nuclear residence time: 40 min in (Mor et al., 2010), 90 min in (Battich et al., 2015), 2 h in (Bahar Halpern et al., 2015)). In our model of step 1, we set the range of possible export times from 5 min to 12 h. More than one third of the genes in our dataset reaches the lower boundary of 5 min (600 genes), suggesting that indeed, rhythmic transcripts are usually exported on a short time scale. The median export time, if we include those genes, is 25 min for M2 and M3 genes, and 12 min for M4 genes. Even if the majority of transcripts were processed and exported within minutes, some mRNA resided for several hours in the nucleus. For example, genes involved in steroid biosynthesis such as *Srd5a1*, *Hsd17b2*, *Dhcr7*, and *Scap*, have an estimated nuclear half-life between 2 h and 3.5 h. However, major regulators of cholesterol synthesis are exported rapidly from the nucleus (*Insig1*,

*Cyp7a1*, *Hmgcr*, export time of 5 min). This suggests that despite being rhythmically transcribed, biologically related genes have distinct dynamic patterns in the nucleus. Genes involved in fructose metabolism were also enriched among long-lived transcripts (*Aldob*: 6,8 h, *Khk*: 4.8 h, *Sord*: 1.4 h, *Fbp1*: 6 h). It is noteworthy that nuclear retained mRNAs include the glucokinase regulatory protein *Gckr* (3.2 h), genes related to thyroid hormone metabolite (*Dio1*: 2.5 h, *Thrsp*: 1.4 h) and related to glyoxylate metabolism (*Agxt*: 1.8 h, *Grhpr*: 1.9 h).

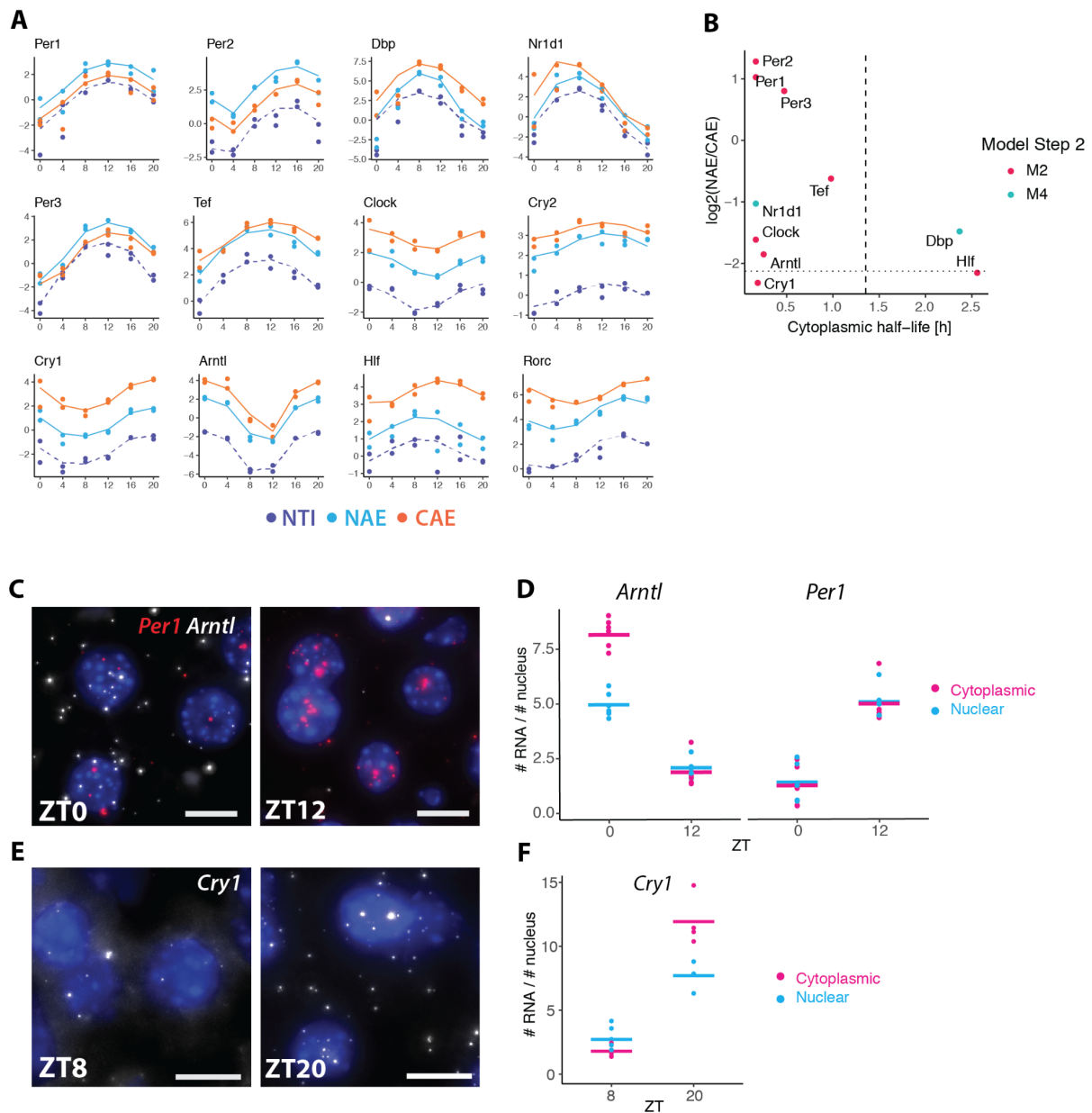


**Fig 5. Kinetic model identifies rhythmically exported nuclear mRNAs.** **A:** Classification of genes in M2 (rhythmic transcription, constant export), M3 genes (constant transcription, rhythmic export), or M4 (rhythmic transcription, rhythmic export). **B:** Distribution of estimated export times of the three models. The median (dotted

line) is calculated excluding export times reaching the lower or upper boundary (5 min and 12 h). **C**: Distribution of peak times of rhythmic nuclear mRNA (NAE). **C**: Distribution of peak time of rhythmic export rates (shortest export time) of M3 and M4 (in M2, export is constant). **D**: Representative temporal profiles of nuclear pre-mRNA (NTI) and nuclear polyadenylated mRNA (NAE) in M2, M3 and M4 genes with short export time (up) or long export time (bottom). Additional examples of M3 genes are shown in Fig S2. **E**: Cytoplasmic mRNA fold-change of genes classified as M2, M3 or M3 in step1.

### **Circadian core clock and PARbZip genes are enriched in the nucleus**

Finally, we focus on the temporal behaviour of core clock and PARbZip family member clock output genes. Cytoplasmic mRNA half-lives of all such genes were below 2.5 h, and nuclear half-lives ranged from 5 min to 1.1 h, allowing an efficient propagation of high-amplitude patterns (Fig 6A, B). Note that *Dbp* and *Rev-Erba* (*Nr1d1*) were classified as both rhythmically exported and rhythmically degraded, which explains their particularly high relative amplitudes. As a consequence of their high cytoplasmic degradation rates, clock genes were more enriched in the nucleus compared to all cycling genes. Of interest, all three *Period* mRNAs (*Per1*, *Per2*, *Per3*) were particularly enriched in the nucleus, despite having similarly short half-lives compared to the other core clock genes, suggesting that these could be potentially slowly exported (Fig 6C). We verified by smFISH the subcellular localisation of *Per1* together with another core clock gene *Bmall* (also named *Arntl*) on the same liver samples used for the RNA-seq (Fig 6C). Quantification of nuclear and cytoplasmic mRNA *in situ* at peak and trough expression confirmed that the nuclear-to-cytoplasmic ratio of *Per1* (1:1 at ZT12) is more nuclear than the ratio of *Bmall* (1.6 times more cytoplasmic mRNA at ZT0). *Cry1*, another core clock gene, showed a similar nuclear-to-cytoplasmic ratio in both RNA-seq and smFISH as *Arntl* (1.5 times more cytoplasmic mRNA at ZT20, Fig 6E, F).



**Fig 6. Circadian core clock and PARbZip genes.** **A:** RNA-seq temporal profiles of clock genes and clock-controlled genes (PARbZip family members), sorted by their nuclear-to-cytoplasmic ratio. **B:** log<sub>2</sub>(NAE/CAE) versus cytoplasmic half-life estimated at step 2. Dashed vertical line represents the median half-life times of all rhythmic mRNAs. Dotted horizontal line represents the median log<sub>2</sub>(NAE/CAE) ratio of all rhythmic mRNAs. **C** and **E:** smFISH of core clock mRNAs *Per1*, *Arntl*, and *Cry1* at their maximal and minimal expression levels in mouse liver tissue. Nuclei stained with DAPI (blue). **D** and **F:** Quantification of mRNA transcripts on smFISH images. 5-6 images were taken per animal, and the average number of mRNA molecules per number of nuclei in each image is represented (dots).

## Discussion

Gene expression is a highly regulated multistep process. In this study, we produced time series RNA measurements from fractionated liver cells to feed a simplified model of the complex network of RNA processing steps, including rhythmically regulated nuclear export and cytoplasmic degradation. Cellular fractionation first allowed us to highlight characteristics and differences of the nuclear and cytoplasmic transcriptomes. Notably, at steady-state, we observed that transcripts enriched in the cytoplasm mainly code for proteins involved in house-keeping functions such as respiration and translation, presumably due to their increased stability (Rabani et al., 2017; Tani et al., 2012). Nuclear enriched mRNAs mainly code for regulatory proteins, and could be unstable as a way to facilitate rapid adaptation of protein production in response to external stimuli (Garneau et al., 2007). Additionally, since nuclear retention acts as a passive filter reducing transcriptional noise (Battich et al., 2015; Stoeger et al., 2016; Xiong et al., 2009), retaining these transcripts could allow a stable production of proteins. Another difference between nuclear and cytoplasmic mRNAs is related to their length. Long transcripts were preferentially located in the nucleus, while short transcripts were more abundant in the cytoplasm. This asymmetric distribution could be again related to the cytoplasmic mRNA stability. Indeed, we found a weak yet significant association of the transcript length with the cytoplasmic degradation rates of rhythmic transcripts, while no correlation was found with the export rates ( $r = -0.21$ ,  $p\text{-value} < 10^{-16}$ , data not shown). A similar link between transcript length and decay rate was already hinted at by a previous study in *Drosophila* Kc167 cells (Chen & van Steensel, 2017). Since mRNA degradation that occurs through the deadenylation of the poly(A) tail and a subsequent exonuclease attack (Garneau et al., 2007) has not been linked with transcript length, the higher instability of long transcripts may reflect an increased probability to be stochastically attacked by endonucleases (Chen & van Steensel, 2017).

Another process that depends on genomic length is splicing. Splicing mainly occurs co-transcriptionally in most organisms (Ameur et al., 2011; Herzog et al., 2018; Khodor et al., 2011, 2012; Tilgner et al., 2012). However, the proportion of introns that are removed concurrently with transcription is still under debate, depending on the organism and experimental methods (Neugebauer,

2019). For example, pre-mRNAs in mouse liver are less frequently spliced co-transcriptionally (Khodor et al., 2012). Here, we investigated the proportion of co- versus post-transcriptionally spliced genes by comparing the expression level of intronic signals in nuclear total pre-mRNA (NTI) and the nuclear polyadenylated pre-mRNA (NAI). We noticed a strong correlation with gene length for protein coding transcripts: with long pre-mRNA inclined to be more co-transcriptionally spliced, while short pre-mRNAs were more post-transcriptionally spliced. These observations support the view that transcription and splicing are two processes occurring in parallel, and that co-transcriptional splicing of introns depends on the transcription time. When the gene is short, transcription is already completed by the time the intron is recognised and spliced. If the gene is long, the spliceosome has more time to recognise splice sites, assemble on the pre-mRNA, and cleave the intron. However, it is also known that the machineries involved in the pre-mRNA processing interact with each other, whether by recruiting splicing and processing factors directly on the site of transcription (Saldi et al., 2016), or because all these factors are concentrated in one location, such as in nuclear speckles and paraspeckles (Gordon et al., 2021). Here, we studied transcript splicing without considering variations in splicing times of introns within a gene body, which differ depending on the length of the downstream exon, or on whether flanking exons are alternatively or constitutively spliced (Braunschweig et al., 2014; Djebali et al., 2012; Tilgner et al., 2012). More sophisticated methods such as long-read sequencing could clarify the splicing regime (Drexler et al., 2020). However, current long-read sequencing technologies are limited in the length of sequenced transcripts (15kb to 30 kb) and are biased toward short length transcripts (Amarasinghe et al., 2020).

We took advantage of the relationship between relative peak times and relative amplitudes describing the propagation of oscillations to estimate the cytoplasmic degradation rates and nuclear export rates of approximately 1400 genes. We found a median cytoplasmic half-life of 2.5 h for cycling transcripts. In general, mRNA half-lives range from minutes to several hours, and the estimation varies depending on the model (cell type, tissue) and method (total RNA-seq, transcription inhibition, metabolic labelling). As a comparison, median half-lives of mRNA was 9 h in NIH3T3 cells (Schwanhausser et al., 2011), 3.9 h in mouse ESCs (Herzog et al., 2017), 4 h in HEK293 cells

(Lugowski et al., 2018), 3.4 h in HELA cells (Tani et al., 2012), and 2.1 h in another study with NIH3T3 fibroblasts (Eisen et al., 2020). In our estimation, we focused only on rhythmic genes, which do indeed need to be relatively short-lived to maintain rhythms.

Nuclear export times are overall shorter and the majority of the rhythmic nuclear mRNA have a half-life of less than 25 min. Note that nuclear degradation is not explicitly included in the model, with the consequence to potentially increase our inferred export rates. Our experimental design reliably detects long export times on the order of our sampling time (in the range of hours), but does not allow a precise quantification of processes happening at extremely short time-scales. Moreover, for a majority of genes, ignoring the nuclear export step as we and others previously did (Lück et al., 2014; Wang et al., 2018) is a valid assumption due to its small contribution to the total mRNA lifetime. However, when nuclear lifetime significantly contributes to the total RNA lifetime, our dataset allows us to separate the processes and provide a more specific interpretation of the kinetic processes. In addition, we were able to uncover rhythmic patterns in the nucleus, which strongly suggests that the nuclear lifetime varies around the day for 12.5 % of the rhythmic nuclear mRNA. As a result, amplitudes of several rhythmically exported mRNA such as *Pck1*, *Por* and *Lpin1* involved in liver metabolism were boosted.

Finally, we provided a comprehensive and temporal inventory of RNA nuclear and cytoplasmic localisation in mouse livers, as well as an identification of the nuclear or cytoplasmic origins of transcript rhythms. In particular, this work suggests that mRNA oscillations can be post-transcriptionally regulated at the level of nuclear export, and contributes to a better understanding of the dynamic regulation of the transcriptome over the 24 hours day.

## **Methods**

### **Animal housing and mouse Liver sampling**

All animal care and handling were approved by the Institutional Animal Care and Use Committee of WIS and by the Canton de Vaud laws for animal protection (authorisation VD.3197). All mice



experiments were performed on male WT mice (C57/BL6J) aged from 8 to 12 weeks old, and housed in 12:12 light:dark cycle. 3 days prior to sacrifice and sampling day, mice were given access to food only during the active phase (night-restricted feeding). Two mice were sacrificed by decapitation every 4 hours (ZT0, 4, 8, 12, 16 and 20). ZT0 is the beginning of the light phase, and ZT12 is the beginning of the dark phase. Pieces of liver were immediately collected after sacrifice for RNA extraction and histological analysis (single-molecule RNA-FISH).

### **Cellular fractionation and RNA extraction**

Nuclear and cytoplasmic liver mRNA were isolated according to the Nascent-seq protocol (Menet et al., 2012) with few modifications. Briefly, approximately 300mg of the big lobe were homogenised with a Potter-elvehjem PTFE pestle together with 3.5ml of ice-cold PBS and 3.5ml of homogenisation buffer (2.2M sucrose, 15mM KCl, EDTA, 10mM Hepes pH 7.6, 0.15 mM spermine, 0.5mM spermidine, 0.5 mM PMSF, 1mM DTT, 20 µg/ml aprotinin, 2.5µg/ml leupeptin, 100µg/ml pepstatin, and 10U/ml RNaseOUT™). The homogenate was mixed with the remaining homogenisation buffer (total volume of homogenisation buffer: 18ml) and layered on top of 7.2ml of a sucrose cushion buffer (2.05 M sucrose, 10% glycerol, 15mM KCl, EDTA, Hepes pH 7.6, 0.15 mM spermine, 0.5 mM spermidine, 20 µg/ml aprotinin, 2.5 µg/ml leupeptin, 100 µg/ml pepstatin, 0.5 mM PMSF, 0.1 M DTT, 10 U/ml RNaseOUT™). The liver homogenate was ultra-centrifuged in a Beckman SW28 rotor at a speed of 103'000 x g, at 4C° for 1 hour. Supernatant (cytoplasmic fraction) was mixed with ethanol (final ethanol concentration of 70%) and stored overnight at -80°C. Cytoplasmic extracts in ethanol were centrifuged at 5000 rpm for 15 min at 4°C to precipitate RNA. Ethanol was discarded, and the pellet was resuspended in 800ul of QIAzol Lysis reagent (QIAGEN). RNA was extracted on spin columns using the miRNeasy Mini Kit (QIAGEN) according to the manufacturer's instructions. After ultracentrifugation, the pellet (nuclei) was resuspended in 3ml of Nuclear RNA extraction buffer (guanidium thiocyanate 50% (w/v), sodium citrate 0.75M pH 7, 200µM sodium acetate 2M pH 4 and 1% β-mercaptoethanol), and was mechanically homogenised through a syringe needle 26G. 3ml of water-saturated phenol (pH 4.5) and 1200µl of chloroform:isoamyl alcohol 24:1 were added to the resuspended pellet, and were well shaken before resting for 15 min at room temperature. The mix was

centrifuged at 4°C at 1500g for 20 min. The supernatant was resuspended in ethanol (final volume 60% (v/v)). RNA is then extracted on spin columns using miRNEasy kit from QIAGEN according to manufacturer's instructions, including a DNase treatment (QIAGEN).

For total liver RNA extraction, pieces from the main liver lobes were frozen in liquid nitrogen. About 20mg were frozen and used for RNA extraction using the miRNEasy kit (QIAGEN).

## **RNA-sequencing**

To sequence polyadenylated nuclear and cytoplasmic RNA, we used the TruSeq stranded mRNA LT kit (Illumina), starting from 650 ng of RNA, according to the manufacturer's protocol. To sequence total RNA (with ribosomal RNA depletion), we used the KAPA RNA HyperPrep Kit with RiboErase prep combined with KAPA Unique Dual-Indexed Adapter Kit (Illumina), starting from 650 ng RNA. Libraries were sequenced on a HiSeq4000. We used a SBS 50 cycles and SR cluster kit, performing a single read sequencing of 65 nucleotides. We sequenced on average 70M reads for Nuclear PolyA and Nuclear Total samples, 25M reads for Cytoplasmic PolyA and Cytoplasmic Total samples, and 50M reads for the Unfractionated samples.

Quantification of pre-mRNA and mRNA expression levels was performed using Kallisto version 0.46.0. A combined index was built using reference fasta files from Ensembl for *mus musculus* (GRCm38). In particular, pre-mRNA (introns and exons) and mRNA (exons) reference sequences were used as an input and named accordingly. Additionally, we had added spike-in mRNA originating from rat liver (male Wistar, donated by the Laboratory for Functional and Metabolic Imaging, EPFL) accounting for 5% of the total mass liver, but these were in the end not used. Therefore, the reference genome used for the mapping includes fasta files from *rattus norvegicus* (Rnor\_6.0). Pseudoalignment and quantification were run on the aforementioned combined kallisto index with parameters "--single --rf-stranded -l 100 -s 30 -s 30". TPM (transcript per million) and estimated counts at (pre-) transcript level were used for further analysis.

We filtered lowly expressed isoforms based on the expression level of the mature mRNA. Specifically, we calculated the Isoform Fraction (IF) for each transcript, which is defined as the TPM

(transcript per million) of the isoform divided by the sum of TPM of all isoforms of a given gene. We defined a gene-specific threshold that depends on the number of annotated isoforms ( $n$ ): the IF had to be greater than  $1/n$  to be kept. Additionally, we set the maximal threshold to 0.2, such that when a gene has only two isoforms, if one contributes to more than 20% of the total expression, it was not filtered out. If a transcript met the condition in at least 6 samples out of the 60 samples (5 RNA populations, 12 time-points), it is kept.

In addition to the isoform-specific quantification, we pooled for each gene the isoforms belonging to the same biotype in order to have a biotype-specific quantification for each gene (sum of the estimated counts of the isoform belonging to the same biotype). As for the gene length used for normalisation, we calculated the weighted effective gene length (estimated by Kallisto), where weights were the expression levels of each isoform in each condition (therefore, gene length may differ between RNA populations). We created 8 biotypes, based on the Ensembl description. Protein coding (PC), Retained Intron (RI), snRNA, and Processed Transcripts are defined by Ensembl. Nonsense mediated decay and non stop decay are grouped as "NMD". We grouped lincRNA, bidirectional promoter lincRNA, sense overlapping, sense intronic, 3prime overlapping ncRNA, lincRNA and in one biotype named "long non-coding RNA" (lncRNA). Small Nucleolar RNA (snoRNA) comprises the biotypes "snoRNA" and "scaRNA". All the biotypes containing the keyword "pseudogenes", whether they are processed, transcribed or not are grouped as "Pseudogene".

Finally, we also summed all the transcripts by gene, independently of their biotype, in order to obtain gene-wise quantification. The gene length was calculated as the weighted average based on the relative expression of all the isoforms in each RNA population.

For the analysis at steady-state (Fig 1, Fig 2 and Table S1), we normalised the estimated counts by Kallisto into RPKM using the R package "DESeq2" (Love et al., 2014). This normalisation procedure was applied on each RNA population separately. We used the combined library size (pre-mRNA + mRNA) per sample to adjust for sequencing depth. The size factor was estimated using mRNA only, such that the median level of mRNA is aligned for all the samples. The same size factor was applied

to the pre-mRNA, such that the relationship between pre-mRNA and mRNA is conserved. Here, we directly implemented a rhythmic analysis during the DESeq2 normalisation process, using the following model:

$$\log_2(\mu_{g,s,t}) = a_{g,s} * \cos(w * t_s) + b_{g,s} * \sin(w * t_s) + \text{intercept}$$

where  $\mu_{g,s,t}$  is the mean count of the negative binomial distribution for gene  $g$  for sample  $s$  at circadian time-point (ZT)  $t_s$ .  $a_{g,s}$  and  $b_{g,s}$  are the coefficients of the cosine and sine functions, and  $w = 2 * \pi / 24$  h. The phase (peak of expression) is defined as  $\tan^{-1}(b/a)$  and the amplitude is (or  $\log_2$  fold change) as  $a^2 + b^2$ . The full model with rhythmic parameters is tested with a Likelihood Ratio test against the reduced model with an intercept only. The intercept fitted by DESeq2 is used as the mean expression level over the 6 time points.

To finally express these expression levels in absolute mRNA counts per hepatocyte (see Fig 6C, E), we used the data in (Bahar Halpern et al., 2015), where the authors converted the counts obtained from sequencing nuclear and cytoplasmic PolyA RNA from mouse liver into number of molecules per cell using smFISH. We applied a linear regression model on the  $\log_2$ -transformed RPKM and added the following  $\log_2$  offsets to the NA, NT, CA and CT datasets: -1.34, -0.15, +1.2 and +0.7.

### Modeling temporal RNA-seq profiles

To describe the accumulation of each RNA species in the nucleus and in the cytoplasm, we assumed that the pre-mRNA  $p$  is spliced and polyadenylated at a rate of  $s$  to produce mature nuclear RNA  $m$ . Then,  $m$  is exported at a rate  $e$  into the cytoplasm and becomes  $M$ , where it is finally degraded at a rate  $\gamma$ . We used NTI to approximate the level of pre-mRNA  $p$ , NAE and CAE to describe the accumulation of mature RNA in the nucleus  $m$  and in the cytoplasm  $M$ . We modeled the accumulation of  $p$  and  $m$  with two independent ordinary differential equations:

$$\frac{dm(t)}{dt} = s p(t) - e(t)m(t) \text{ (Step 1)}$$

$$\frac{dM(t)}{dt} = em(t) - \gamma(t)M(t) \text{ (Step 2)}$$

To distinguish different modes of post-transcriptional regulation, we applied the model-based approach described in (Wang et al., 2018) on the two equations. Briefly,  $e(t)$  and  $\gamma(t)$ , respectively the export and cytoplasmic degradation rates, can be either constant ( $e(t) = e_0$  and  $\gamma(t) = \gamma_0$ ) or rhythmic ( $e(t) = e_0(1 + \varepsilon_e \cos(\omega t - \varphi_e))$  and  $\gamma(t) = \gamma_0(1 + \varepsilon_\gamma \cos(\omega t - \varphi_\gamma))$ ), with  $\omega$  the angular frequency of  $2\pi/24\text{h}^{-1}$ ,  $\varphi$  the phase (peak time) and  $\varepsilon$  the relative amplitude varying from 0 to 1.  $p(t)$  and  $m(t)$  are either constant ( $p(t) = p_0$  and  $m(t) = m_0$ ) or rhythmic ( $p(t) = p_{\min} + A_p((1 + \cos(\omega t - \varphi_p))/2)^\beta$  and  $m(t) = m_{\min} + A_m((1 + \cos(\omega t - \varphi_m))/2)^\beta$ , with  $A$  the absolute amplitude (peak to trough),  $\beta$  ranging from 1 to 2, and  $\varphi$  the phase. Because NTI, NAE, and CAE come from samples processed and sequenced separately, the relative abundance is not defined, therefore, these parameters do not have a biologically interpretable value.

The following procedures and equations describe step 1, but are the same for step 2. To infer the values of the 8 parameters:  $A_p$ ,  $Min_p$ ,  $\varphi_p$ ,  $\beta$ ,  $s$ ,  $e_0$ ,  $\varepsilon_e$  and  $\varphi_e$ , we used the function *optim* from the R package "stats" with the method "L-BFGS-B". We assumed that RNA-seq counts data follows a binomial distribution (Love et al., 2014, p. 201). Thus, the log-likelihood function to be minimised for each gene is:

$$\log L = \sum_t \log NB(n_p(t) | \mu_p(t), \alpha_p(t)) + \log NB(n_m(t) | \mu_m(t), \alpha_m(t))$$

with

$$\mu_p(t) = p(t) S(t) L_p$$

$$\mu_m(t) = m(t) S(t) L_m$$

Here,  $n(t)$  is the read count, the subscripts  $p$  and  $m$  stand for pre-mRNA (NTI) and nuclear mRNA (NAE).  $\alpha$  is the dispersion parameter of the negative binomial distribution, which is specific to each gene and time-point. It is estimated by the R package DESeq2 (fit type: "parametric") (Love et al., 2014).  $\mu(t)$  is the expected mean of counts, which corresponds to the concentration of transcripts ( $p(t)$  and  $m(t)$ ) multiplied by the gene length, and a sample-specific scaling factor  $S$ , and further multiplied

by  $10^9$  to obtain numerically convenient values. The scaling factor  $S$  is first estimated by DESeq2 separately on each RNA population, including only mRNA, such that the median expression levels of mRNA are aligned between samples (Nuclear Total, Nuclear PolyA, and Cytoplasmic PolyA). The parameter  $e$  is bounded to 5 min to 12 h, and  $\gamma$  to 10 min to 24 h.

The combination of rhythmic / constant transcription and export terms generates 4 models (M1 to M4), each of them are fitted to experimental RNA-seq profiles and the most parsimonious model was chosen based on the Bayesian information criterion (BIC). The BIC is defined as:  $-\log + K \log(N)$ , with  $K$  the number of parameters and  $N$  the number of datapoints. In order to discuss only unambiguously classified genes, we selected genes with a BIC probability (Schwarz weights)  $> 0.6$ .

### **Gene enrichment analysis**

We tested for gene set enrichment of protein coding transcripts in nuclear and cytoplasmic fractions using Gene Ontology Terms derived from "Biological Processes". Nuclear and Cytoplasmic enrichments were assessed using DESeq2 (Love et al., 2014). Raw counts of NAE and CAE samples were tested for enrichment using the cellular localisation (Nuclear or Cytoplasmic) as a variable in the design formula. Transcripts with a  $\log_2FC > 2$  and an adjusted p-value  $< 0.01$  were considered as enriched.

We then used Fisher's exact test implemented in the EnrichR package to test for significance (Kuleshov et al., 2016). GO Terms with an adjusted p-value  $< 0.1$  and a Combined Score  $> 30$  were considered as enriched. We grouped together terms that were semantically similar.

### **Single-molecule RNA-FISH**

Dissected liver pieces were immediately fixed in 10% Neutral buffered Formalin (NBF) at  $4C^\circ$  for 24 to 36 hours. Fixed samples were then washed in PBS 1x for 30 min, and dehydrated in standard ethanol series followed by xylene bath, and finally embedded in paraffin (formalin fixed paraffin embedded, FFPE tissues). Sections were  $8\mu m$  thick. Single-molecule RNA Fluorescent *in situ*

Hybridisation (smFISH) experiments were performed using the RNAScope Multiplex Fluorescent Assay V2 (Advanced Cell Diagnostics), according to manufacturer's instructions. Nuclei were counterstained with Dracq5 for 10 min. Sections were mounted with ProLong Gold Antifade Mountant. The *Cry1* smFISH images were taken from (Mermet et al., 2018) from liver cryosection and quantified for nuclear and cytoplasmic abundance. Catalog number of the probes: *Cry1*: 500031, *Per1*: 438751, *Arntl*: 438741, *Actb*: 316741, *Mlxipl*: 558141.

### **Microscopy and image analysis**

Sections were imaged with the Visitron Spinning Disk CSU W1 with a motorized stage. Z-stacks were acquired with a distance of 0.2um between each Z position (~40 images per frame) with an oil-immersion 63x objective. All z-stacks were maximally projected for analysis. The image was performed using ImageJ. mRNA spots are detected by first applying a Gaussian blur, followed by the edge-detector Laplacian filter. The local maxima - corresponding to the spots - are computed and counted. Nuclei are detected by applying a median blur filter, the Otsu method for automatic thresholding, and the watershed algorithm for segmentation. We quantified a density of mRNA per nuclei: we counted all the mRNAs present in one microscopy image and divided by the number of segmented nuclei.

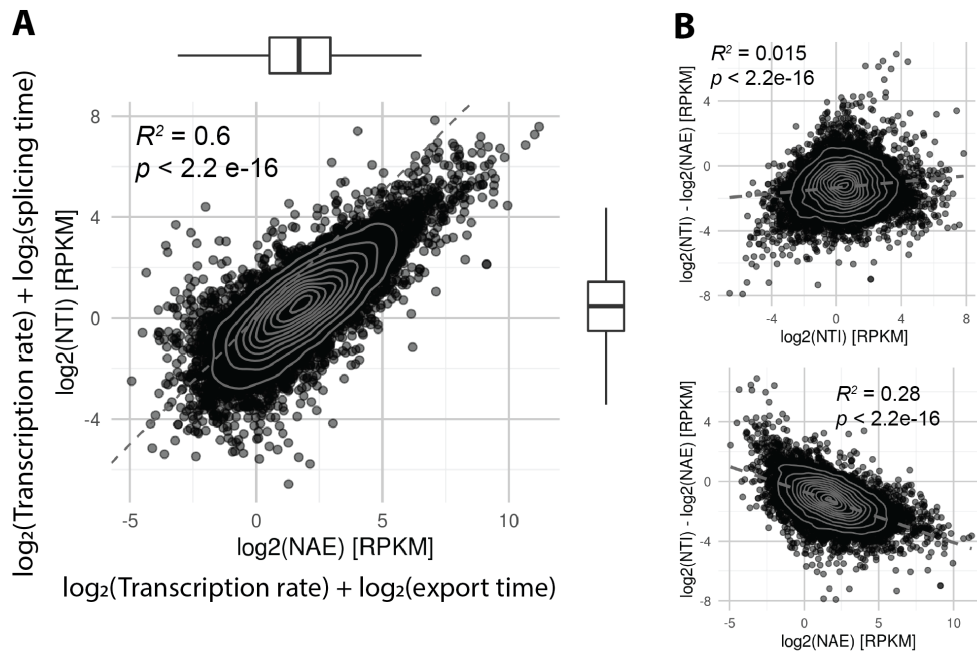
## **Supplementary Information: Comprehensive analysis of the circadian nuclear and cytoplasmic transcriptome in mouse liver**

Clémence Hurni<sup>1</sup>, Benjamin Weger<sup>1</sup>, Cédric Gobet<sup>1</sup>, Felix Naef<sup>1</sup>

1: The Institute of Bioengineering (IBI), School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

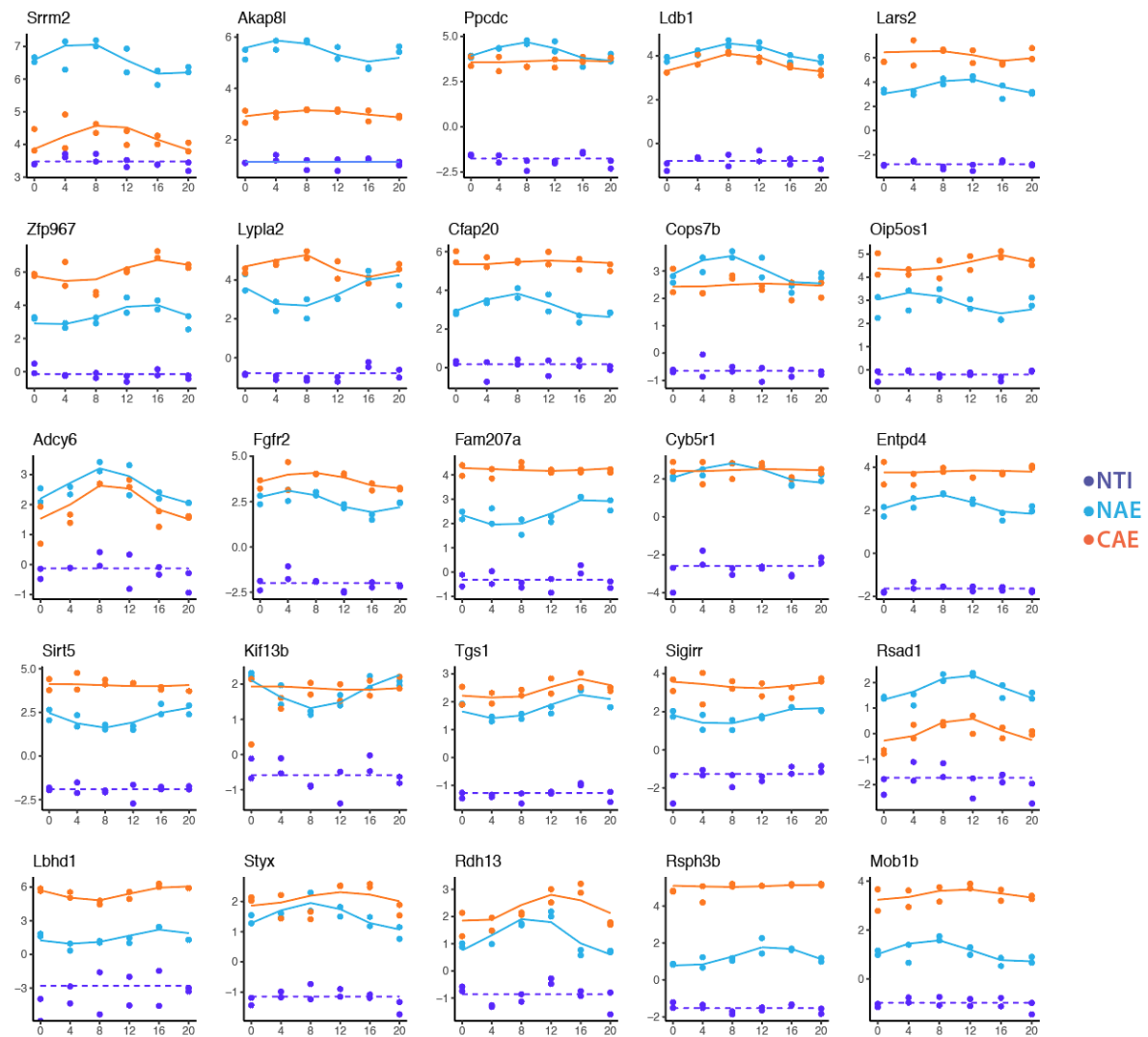


## Supplementary Figure S1



**Fig S1. The ratio of nuclear pre-mRNA and mRNA reflects the relationship between transcription rate, splicing time, and export time.** **A:** NTI versus NAE, in  $\log_2(\text{RPKM})$ , averaged over 6 time points. According to our model,  $\log_2(\text{NTI})$  is defined as  $\log_2(\text{Transcription rate}) + \log_2(\text{splicing time})$  and  $\log_2(\text{NAE})$  is defined as  $\log_2(\text{Transcription rate}) + \log_2(\text{export time})$ . In grey: kernel 2D density. Boxplots on top and on the right show the distribution of NTI and NAE. **B:** NTI / NAE ratio in  $\log_2$ -scale, defined as  $\log_2(\text{splicing time}) - \log_2(\text{export time})$ , against the average expression of NTI (up) or against NAE (bottom). Grey line indicates linear regression.

## Supplementary Figure S2



**Fig S2: Temporal profiles of genes classified as M3 in step 1.** 25 profiles of genes classified as M3 (constant transcription, rhythmic export) according to the following criteria: all the genes have a  $\log_2FC > 0.8$ , and all classified as M2 or M4 in the analysis of step 2 (NAE vs CAE). Plots are arranged by the expression level of NAE (highest NAE *Srrm2* in the upper left corner). Light blue dots and solid line: NAE, purple dots and dotted line: NTI.

## References

- Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*, 21(1), 30. <https://doi.org/10.1186/s13059-020-1935-5>
- Ameur, A., Zaghlool, A., Halvardson, J., Wetterbom, A., Gyllensten, U., Cavelier, L., & Feuk, L. (2011). Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nature Structural & Molecular Biology*, 18(12), 1435–1440. <https://doi.org/10.1038/nsmb.2143>
- Atger, F., Gobet, C., Marquis, J., Martin, E., Wang, J., Weger, B., Lefebvre, G., Descombes, P., Naef, F., & Gachon, F. (2015). Circadian and feeding rhythms differentially affect rhythmic mRNA transcription and translation in mouse liver. *Proceedings of the National Academy of Sciences*, 112(47), E6579–E6588. <https://doi.org/10.1073/pnas.1515308112>
- Baggs, J. E., & Green, C. B. (2003). Nocturnin, a Deadenylyase in *Xenopus laevis* Retina: A Mechanism for Posttranscriptional Control of Circadian-Related mRNA. *Current Biology*, 13(3), 189–198. [https://doi.org/10.1016/S0960-9822\(03\)00014-9](https://doi.org/10.1016/S0960-9822(03)00014-9)
- Bahar Halpern, K., Caspi, I., Lemze, D., Levy, M., Landen, S., Elinav, E., Ulitsky, I., & Itzkovitz, S. (2015). Nuclear Retention of mRNA in Mammalian Tissues. *Cell Reports*, 13(12), 2653–2662. <https://doi.org/10.1016/j.celrep.2015.11.036>
- Battich, N., Stoeger, T., & Pelkmans, L. (2015). Control of Transcript Variability in Single Mammalian Cells. *Cell*, 163(7), 1596–1610. <https://doi.org/10.1016/j.cell.2015.11.018>
- Benegiamo, G., Mure, L. S., Erikson, G., Le, H. D., Moriggi, E., Brown, S. A., & Panda, S. (2018). The RNA-Binding Protein NONO Coordinates Hepatic Adaptation to Feeding. *Cell Metabolism*, 27(2), 404–418.e7. <https://doi.org/10.1016/j.cmet.2017.12.010>
- Bhatt, D. M., Pandya-Jones, A., Tong, A.-J., Barozzi, I., Lissner, M. M., Natoli, G., Black, D. L., & Smale, S. T. (2012). Transcript Dynamics of Proinflammatory Genes Revealed by Sequence Analysis of Subcellular RNA Fractions. *Cell*, 150(2), 279–290.

<https://doi.org/10.1016/j.cell.2012.05.043>

- Braunschweig, U., Barbosa-Morais, N. L., Pan, Q., Nachman, E. N., Alipanahi, B., Gonatopoulos-Pournatzis, T., Frey, B., Irimia, M., & Blencowe, B. J. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Research*, 24(11), 1774–1786. <https://doi.org/10.1101/gr.177790.114>
- Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, 34(5), 525–527. <https://doi.org/10.1038/nbt.3519>
- Chen, T., & van Steensel, B. (2017). Comprehensive analysis of nucleocytoplasmic dynamics of mRNA in *Drosophila* cells. *PLoS Genetics*, 13(8), e1006929. <https://doi.org/10.1371/journal.pgen.1006929>
- Chin, A., & Lécuycer, E. (2017). RNA localization: Making its way to the center stage. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1861(11, Part B), 2956–2970. <https://doi.org/10.1016/j.bbagen.2017.06.011>
- Clemson, C. M., Hutchinson, J. N., Sara, S. A., Ensminger, A. W., Fox, A. H., Chess, A., & Lawrence, J. B. (2009). An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Molecular Cell*, 33(6), 717–726. <https://doi.org/10.1016/j.molcel.2009.01.026>
- Damiola, F., Le Minh, N., Preitner, N., Kornmann, B., Fleury-Olela, F., & Schibler, U. (2000). Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. *Genes & Development*, 14(23), 2950–2961.
- Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G. K., Khatun, J., Williams, B. A., Zaleski, C., Rozowsky, J., Röder, M., Kokocinski, F., Abdelhamid, R. F., ... Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature*, 489(7414), 101–108. <https://doi.org/10.1038/nature11233>
- Drexler, H. L., Choquet, K., & Churchman, L. S. (2020). Splicing Kinetics and Coordination

- Revealed by Direct Nascent RNA Sequencing through Nanopores. *Molecular Cell*, 77(5), 985-998.e8. <https://doi.org/10.1016/j.molcel.2019.11.017>
- Du, N.-H., Arpat, A. B., De Matos, M., & Gatfield, D. (2014). MicroRNAs shape circadian hepatic gene expression on a transcriptome-wide scale. *ELife*, 3, e02510. <https://doi.org/10.7554/eLife.02510>
- Eisen, T. J., Eichhorn, S. W., Subtelny, A. O., Lin, K. S., McGeary, S. E., Gupta, S., & Bartel, D. P. (2020). The Dynamics of Cytoplasmic mRNA Metabolism. *Molecular Cell*, 77(4), 786-799.e10. <https://doi.org/10.1016/j.molcel.2019.12.005>
- Fazal, F. M., Han, S., Parker, K. R., Kaewsapsak, P., Xu, J., Boettiger, A. N., Chang, H. Y., & Ting, A. Y. (2019). Atlas of Subcellular RNA Localization Revealed by APEX-Seq. *Cell*, 178(2), 473-490.e26. <https://doi.org/10.1016/j.cell.2019.05.027>
- Furlan, M., Galeota, E., Gaudio, N. D., Dassi, E., Caselle, M., Pretis, S. de, & Pelizzola, M. (2020). Genome-wide dynamics of RNA synthesis, processing, and degradation without RNA metabolic labeling. *Genome Research*. <https://doi.org/10.1101/gr.260984.120>
- Fustin, J.-M., Doi, M., Yamaguchi, Y., Hida, H., Nishimura, S., Yoshida, M., Isagawa, T., Morioka, M. S., Takeya, H., Manabe, I., & Okamura, H. (2013). RNA-Methylation-Dependent RNA Processing Controls the Speed of the Circadian Clock. *Cell*, 155(4), 793–806. <https://doi.org/10.1016/j.cell.2013.10.026>
- Gaidatzis, D., Burger, L., Florescu, M., & Stadler, M. B. (2015). Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. *Nature Biotechnology*, 33(7), 722–729. <https://doi.org/10.1038/nbt.3269>
- Garneau, N. L., Wilusz, J., & Wilusz, C. J. (2007). The highways and byways of mRNA decay. *Nature Reviews Molecular Cell Biology*, 8(2), 113–126. <https://doi.org/10.1038/nrm2104>
- Gordon, J. M., Phizicky, D. V., & Neugebauer, K. M. (2021). Nuclear mechanisms of gene expression control: Pre-mRNA splicing as a life or death decision. *Current Opinion in Genetics & Development*, 67, 67–76. <https://doi.org/10.1016/j.gde.2020.11.002>

- Gotic, I., Omid, S., Fleury-Olela, F., Molina, N., Naef, F., & Schibler, U. (2016). Temperature regulates splicing efficiency of the cold-inducible RNA-binding protein gene *Cirbp*. *Genes & Development*, *30*(17), 2005–2017. <https://doi.org/10.1101/gad.287094.116>
- Herzel, L., Straube, K., & Neugebauer, K. M. (2018). Long-read sequencing of nascent RNA reveals coupling among RNA processing events. *Genome Research*, *28*(7), 1008–1019. <https://doi.org/10.1101/gr.232025.117>
- Herzog, V. A., Reichholf, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T. R., Wlotzka, W., von Haeseler, A., Zuber, J., & Ameres, S. L. (2017). Thiol-linked alkylation of RNA to assess expression dynamics. *Nature Methods*, *14*(12), 1198–1204. <https://doi.org/10.1038/nmeth.4435>
- Hocine, S., Singer, R. H., & Grünwald, D. (2010). RNA Processing and Export. *Cold Spring Harbor Perspectives in Biology*, *2*(12), a000752. <https://doi.org/10.1101/cshperspect.a000752>
- Hughes, M. E., DiTacchio, L., Hayes, K. R., Vollmers, C., Pulivarthy, S., Baggs, J. E., Panda, S., & Hogenesch, J. B. (2009). Harmonics of Circadian Gene Transcription in Mammals. *PLOS Genetics*, *5*(4), e1000442. <https://doi.org/10.1371/journal.pgen.1000442>
- Hutchinson, J. N., Ensminger, A. W., Clemson, C. M., Lynch, C. R., Lawrence, J. B., & Chess, A. (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics*, *8*(1), 39. <https://doi.org/10.1186/1471-2164-8-39>
- Jackson, D. A., Pombo, A., & Iborra, F. (2000). The balance sheet for transcription: An analysis of nuclear RNA metabolism in mammalian cells. *The FASEB Journal*, *14*(2), 242–254. <https://doi.org/10.1096/fasebj.14.2.242>
- Khodor, Y. L., Menet, J. S., Tolan, M., & Rosbash, M. (2012). Cotranscriptional splicing efficiency differs dramatically between *Drosophila* and mouse. *RNA*, *18*(12), 2174–2186. <https://doi.org/10.1261/rna.034090.112>
- Khodor, Y. L., Rodriguez, J., Abruzzi, K. C., Tang, C.-H. A., Marr, M. T., & Rosbash, M.

- (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in *Drosophila*. *Genes & Development*, 25(23), 2502–2512.  
<https://doi.org/10.1101/gad.178962.111>
- Koike, N., Yoo, S.-H., Huang, H.-C., Kumar, V., Lee, C., Kim, T.-K., & Takahashi, J. S. (2012). Transcriptional Architecture and Chromatin Landscape of the Core Circadian Clock in Mammals. *Science (New York, N. Y.)*, 338(6105), 349–354.  
<https://doi.org/10.1126/science.1226339>
- Kojima, S., Sher-Chen, E. L., & Green, C. B. (2012). Circadian control of mRNA polyadenylation dynamics regulates rhythmic protein expression. *Genes & Development*, 26(24), 2724–2736. <https://doi.org/10.1101/gad.208306.112>
- Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W., & Ma'ayan, A. (2016). Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, 44(W1), W90-97. <https://doi.org/10.1093/nar/gkw377>
- Lam, L. T., Pickeral, O. K., Peng, A. C., Rosenwald, A., Hurt, E. M., Giltneane, J. M., Averett, L. M., Zhao, H., Davis, R. E., Sathyamoorthy, M., Wahl, L. M., Harris, E. D., Mikovits, J. A., Monks, A. P., Hollingshead, M. G., Sausville, E. A., & Staudt, L. M. (2001). Genomic-scale measurement of mRNA turnover and the mechanisms of action of the anti-cancer drug flavopiridol. *Genome Biology*, 2(10), research0041.1-research0041.11.
- Lee, K.-H., Kim, S.-H., Kim, D.-Y., Kim, S., & Kim, K.-T. (2012). Internal Ribosomal Entry Site-Mediated Translation Is Important for Rhythmic PERIOD1 Expression. *PLOS ONE*, 7(5), e37936. <https://doi.org/10.1371/journal.pone.0037936>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550.  
<https://doi.org/10.1186/s13059-014-0550-8>
- Lowrey, P. L., & Takahashi, J. S. (2000). Genetics of the Mammalian Circadian System:

Photic Entrainment, Circadian Pacemaker Mechanisms, and Posttranslational Regulation. *Annual Review of Genetics*, 34(1), 533–562.

<https://doi.org/10.1146/annurev.genet.34.1.533>

Lück, S., Thurley, K., Thaben, P. F., & Westermark, P. O. (2014). Rhythmic Degradation Explains and Unifies Circadian Transcriptome and Proteome Data. *Cell Reports*, 9(2), 741–751. <https://doi.org/10.1016/j.celrep.2014.09.021>

Lugowski, A., Nicholson, B., & Rissland, O. S. (2018). DRUID: A pipeline for transcriptome-wide measurements of mRNA stability. *RNA*, 24(5), 623–632. <https://doi.org/10.1261/rna.062877.117>

Martelot, G. L., Canella, D., Symul, L., Migliavacca, E., Gilardi, F., Liechti, R., Martin, O., Harshman, K., Delorenzi, M., Desvergne, B., Herr, W., Deplancke, B., Schibler, U., Rougemont, J., Guex, N., Hernandez, N., Naef, F., & Consortium, the C. (2012). Genome-Wide RNA Polymerase II Profiles and RNA Accumulation Reveal Kinetics of Transcription and Associated Epigenetic Changes During Diurnal Cycles. *PLOS Biology*, 10(11), e1001442. <https://doi.org/10.1371/journal.pbio.1001442>

Menet, J. S., Rodriguez, J., Abruzzi, K. C., & Rosbash, M. (2012). Nascent-Seq reveals novel features of mouse circadian transcriptional regulation. *ELife*, 1, e00011. <https://doi.org/10.7554/eLife.00011>

Mermet, J., Yeung, J., Hurni, C., Mauvoisin, D., Gustafson, K., Jouffe, C., Nicolas, D., Emmenegger, Y., Gobet, C., Franken, P., Gachon, F., & Naef, F. (2018). Clock-dependent chromatin topology modulates circadian transcription and behavior. *Genes & Development*, 32(5–6), 347–358. <https://doi.org/10.1101/gad.312397.118>

Mermet, J., Yeung, J., & Naef, F. (2017). Systems Chronobiology: Global Analysis of Gene Regulation in a 24-Hour Periodic World. *Cold Spring Harbor Perspectives in Biology*, 9(3), a028720. <https://doi.org/10.1101/cshperspect.a028720>

Mor, A., Suliman, S., Ben-Yishay, R., Yunger, S., Brody, Y., & Shav-Tal, Y. (2010). Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells. *Nature Cell Biology*, 12(6), 543–552. <https://doi.org/10.1038/ncb2056>



- Morf, J., Rey, G., Schneider, K., Stratmann, M., Fujita, J., Naef, F., & Schibler, U. (2012). Cold-Inducible RNA-Binding Protein Modulates Circadian Gene Expression Posttranscriptionally. *Science*, *338*(6105), 379–383.  
<https://doi.org/10.1126/science.1217726>
- Müller-McNicoll, M., & Neugebauer, K. M. (2013). How cells get the message: Dynamic assembly and function of mRNA–protein complexes. *Nature Reviews Genetics*, *14*(4), 275–287. <https://doi.org/10.1038/nrg3434>
- Neugebauer, K. M. (2019). Nascent RNA and the Coordination of Splicing with Transcription. *Cold Spring Harbor Perspectives in Biology*, *11*(8), a032227.  
<https://doi.org/10.1101/cshperspect.a032227>
- Palazzo, A. F., & Lee, E. S. (2018). Sequence Determinants for Nuclear Retention and Cytoplasmic Export of mRNAs and lncRNAs. *Frontiers in Genetics*, *0*.  
<https://doi.org/10.3389/fgene.2018.00440>
- Rabani, M., Levin, J. Z., Fan, L., Adiconis, X., Raychowdhury, R., Garber, M., Gnirke, A., Nusbaum, C., Hacohen, N., Friedman, N., Amit, I., & Regev, A. (2011). Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nature Biotechnology*, *29*(5), 436–442.  
<https://doi.org/10.1038/nbt.1861>
- Rabani, M., Pieper, L., Chew, G.-L., & Schier, A. F. (2017). A Massively Parallel Reporter Assay of 3' UTR Sequences Identifies In Vivo Rules for mRNA Degradation. *Molecular Cell*, *68*(6), 1083-1094.e5. <https://doi.org/10.1016/j.molcel.2017.11.014>
- Saldi, T., Cortazar, M. A., Sheridan, R. M., & Bentley, D. L. (2016). Coupling of RNA polymerase II transcription elongation with pre-mRNA splicing. *Journal of Molecular Biology*, *428*(12), 2623–2635. <https://doi.org/10.1016/j.jmb.2016.04.017>
- Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., & Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature*, *473*(7347), 337–342. <https://doi.org/10.1038/nature10098>
- Shearman, L. P. (2000). Interacting Molecular Loops in the Mammalian Circadian Clock.

- Science*, 288(5468), 1013–1019. <https://doi.org/10.1126/science.288.5468.1013>
- So, W. V., & Rosbash, M. (1997). Post-transcriptional regulation contributes to *Drosophila* clock gene mRNA cycling. *The EMBO Journal*, 16(23), 7146–7155.  
<https://doi.org/10.1093/emboj/16.23.7146>
- Sobel, J. A., Krier, I., Andersin, T., Raghav, S., Canella, D., Gilardi, F., Kalantzi, A. S., Rey, G., Wegers, B., Gachon, F., Peraro, M. D., Hernandez, N., Schibler, U., Deplancke, B., Naef, F., & Consortium, C. (2017). Transcriptional regulatory logic of the diurnal cycle in the mouse liver. *PLOS Biology*, 15(4), e2001069.  
<https://doi.org/10.1371/journal.pbio.2001069>
- Stoeger, T., Battich, N., & Pelkmans, L. (2016). Passive Noise Filtering by Cellular Compartmentalization. *Cell*, 164(6), 1151–1161.  
<https://doi.org/10.1016/j.cell.2016.02.005>
- Tani, H., Mizutani, R., Salam, K. A., Tano, K., Ijiri, K., Wakamatsu, A., Isogai, T., Suzuki, Y., & Akimitsu, N. (2012). Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Research*, 22(5), 947–956.  
<https://doi.org/10.1101/gr.130559.111>
- Tilgner, H., Knowles, D. G., Johnson, R., Davis, C. A., Chakraborty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T. R., & Guigó, R. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Research*, 22(9), 1616–1625.  
<https://doi.org/10.1101/gr.134445.111>
- Torres, M., Becquet, D., Blanchard, M.-P., Guillen, S., Boyer, B., Moreno, M., Franc, J.-L., & François-Bellan, A.-M. (2016). Circadian RNA expression elicited by 3'-UTR IRAlu-paraspeckle associated elements. *ELife*, 5, e14837.  
<https://doi.org/10.7554/eLife.14837>
- Vollmers, C., Gill, S., DiTacchio, L., Pulivarthy, S. R., Le, H. D., & Panda, S. (2009). Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression. *Proceedings of the National Academy of Sciences*, 106(50), 21453–21458.

<https://doi.org/10.1073/pnas.0909591106>

Wang, J., Mauvoisin, D., Martin, E., Atger, F., Galindo, A. N., Dayon, L., Sizzano, F., Palini, A., Kussmann, M., Waridel, P., Quadroni, M., Dulić, V., Naef, F., & Gachon, F. (2017). Nuclear Proteomics Uncovers Diurnal Regulatory Landscapes in Mouse Liver. *Cell Metabolism*, 25(1), 102–117. <https://doi.org/10.1016/j.cmet.2016.10.003>

Wang, J., Symul, L., Yeung, J., Gobet, C., Sobel, J., Lück, S., Westermarck, P. O., Molina, N., & Naef, F. (2018). Circadian clock-dependent and -independent posttranscriptional regulation underlies temporal mRNA accumulation in mouse liver. *Proceedings of the National Academy of Sciences*, 115(8), E1916–E1925.

<https://doi.org/10.1073/pnas.1715225115>

Weger, B. D., Gobet, C., David, F. P. A., Atger, F., Martin, E., Phillips, N. E., Charpagne, A., Weger, M., Naef, F., & Gachon, F. (2021). Systematic analysis of differential rhythmic liver gene expression mediated by the circadian clock and feeding rhythms. *Proceedings of the National Academy of Sciences*, 118(3).

<https://doi.org/10.1073/pnas.2015803118>

Xia, C., Babcock, H. P., Moffitt, J. R., & Zhuang, X. (2019). Multiplexed detection of RNA using MERFISH and branched DNA amplification. *Scientific Reports*, 9, 7721.

<https://doi.org/10.1038/s41598-019-43943-8>

Xiong, L., Ma, Y., & Tang, L. (2009). Attenuation of transcriptional bursting in mRNA transport. *Physical Biology*, 7(1), 016005.

<https://doi.org/10.1088/1478-3975/7/1/016005>

Zeisel, A., Köstler, W. J., Molotski, N., Tsai, J. M., Krauthgamer, R., Jacob-Hirsch, J., Rechavi, G., Soen, Y., Jung, S., Yarden, Y., & Domany, E. (2011). Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli. *Molecular Systems Biology*, 7(1), 529.

<https://doi.org/10.1038/msb.2011.62>

Zhang, R., Lahens, N. F., Ballance, H. I., Hughes, M. E., & Hogenesch, J. B. (2014). A circadian gene expression atlas in mammals: Implications for biology and medicine.

*Proceedings of the National Academy of Sciences*, 111(45), 16219–16224.

<https://doi.org/10.1073/pnas.1408886111>

Zuckerman, B., & Ulitsky, I. (2019). Predictive models of subcellular localization of long RNAs. *RNA*, 25(5), 557–572. <https://doi.org/10.1261/rna.068288.118>