Critical role of rhythmic poly(A) tail regulation in circadian gene expression

Xiangyu Yao[1], Shihoko Kojima[1,2,3], Jing Chen[1,2*]

1. Genetics, Bioinformatics, and Computational Biology program, Virginia Tech
2. Department of Biological Sciences, Virginia Tech
3. Biocomplexity Institute, Virginia Tech

* Corresponding author. Email: chenjing@vt.edu (J. C.)

Abstract

The circadian rhythmicity is deeply rooted in rhythmic regulation of gene expression. The core clock pathway that generates circadian oscillation has been well studied, but it remains unclear how this core rhythm controls rhythmic gene expression. Besides that several core clock components are transcriptional factors and mediate genome-wide rhythmic transcriptional control, mounting evidence has demonstrated the importance of rhythmic posttranscriptional controls. A recent study particularly highlighted rhythmic control of poly(A) tail lengths in hundreds of genes in mouse liver and a strong correlation of poly(A) rhythm with protein expression rhythm. In this work we constructed a simplistic model to study the effect of rhythmic poly(A) tail regulation on circadian mRNA expression. The model depicted rhythmic control imposed upon basic mRNA expression processes, including transcription, polyadenylation, deadenylation and degradation. The model results revealed rhythmicity in deadenylation as the strongest contributor to the rhythmicity in poly(A) tail length, and the rhythmicity in abundance of the mRNA subpopulation with long poly(A) tails, which serves a rough proxy for mRNA translatability. In line with this finding, our model further demonstrated that the rhythmic patterns found in the expression of deadenylases could funnel the peak phases of poly(A) tail length and long-tailed mRNA abundance, respectively, into three distinct groups, which could allow genes within each group to coordinate their functions around the clock. Last, our model suggested factors that contribute to the experimentally observed rhythmicity in poly(A) tail length and total mRNA abundance.

Introduction

Rhythmic control of gene expression is a hallmark of the circadian system. The daily rhythms in biochemistry, physiology and behavior ultimately stem from rhythmic gene expression in each cell (Takahashi, Hong et al. 2008, Mohawk, Green et al. 2012). In mammals, approximately 3-15% of mRNAs are rhythmically expressed with a ~24 hr period in any given tissue (Duffield 2003, Zhang, Lahens et al. 2014, Kojima and Green 2015). The rhythmicity originates from a cell-autonomous circadian clock machinery, which consists of a set of core clock genes interlocked by transcription-translation feedback loops (Zhang and Kay 2010, Partch, Green et al. 2014, Takahashi 2015). Many core clock genes encode transcription factors, and interact with their respective target enhancers to exert rhythmic transcriptional controls over circadian mRNA expression (Ueda, Hayashi et al. 2005, Takahashi 2015).

While rhythmic transcriptional control has been extensively studied, rhythmic control of gene expression also occurs beyond the transcription process (Kojima, Shingle et al. 2011, Hirano, Fu et al. 2016, Gobet and Naef 2017). Recent genome-wide analyses and mathematical modeling particularly highlighted the role of post-transcriptional regulations in driving rhythmic mRNA expressions (Koike, Yoo et al. 2012, Menet, Rodriguez et al. 2012, Luck, Thurley et al. 2014, Trott and Menet 2018, Wang, Symul et al. 2018). Post-transcriptional regulation targets various processes, such as RNA splicing, nuclear export, cellular translocation, dormancy and degradation of mRNA (Keene 2010). Many post-transcriptional processes are under circadian control, which, in turn, affect the phase and amplitude of mRNA abundance (Kim, Kim et al. 2005, Kojima, Gatfield et al. 2010, Kojima, Shingle et al. 2011, Du, Arpat et al. 2014). Ultimately, rhythmicity in the transcription and post-transcriptional processes couple with each other and jointly determine the rhythmicity in gene expression. For example, circadian

rhythmicity in RNA transcription and turnover (Luck, Thurley et al. 2014) jointly determine the rhythmicity in mRNA level. As yet, it remains poorly understood how rhythmicity in other post-transcriptional processes affect the rhythmicity in gene expression.

One of the post-transcriptional mechanisms that impact rhythmic gene expression is the regulation of poly(A) tail length. The tracts of adenosines at the 3' end of nearly all eukaryotic mRNAs are critical for controlling stability and translatability of the mRNA (Beilharz and Preiss 2007, Eckmann, Rammelt et al. 2011, Jalkanen, Coleman et al. 2014). Hundreds of mRNAs are recently discovered to exhibit robust circadian rhythms in their poly(A) tail lengths in mouse liver (Kojima, Sher-Chen et al. 2012). Interestingly, the rhythmicity in poly(A) tail length is closely correlated with rhythmicity in the level of the corresponding proteins, indicating that the rhythmic poly(A) regulation plays an important role in driving rhythmic protein expression (Kojima, Sher-Chen et al. 2012). Similar daily fluctuations in poly(A) tail lengths were also reported in the mouse brain (Robinson, Frim et al. 1988, Gerstner, Vanderheyden et al. 2012). In addition, the amplitude of mRNA rhythmicity increased in the absence of *Nocturnin*, a deadenylase (enzyme that removes poly(A) tails from mRNAs) which is rhythmically expressed in various mouse tissues (Wang, Osterbur et al. 2001, Stubblefield, Gao et al. 2018). These data collectively support the importance of poly(A) tail rhythmicity in regulating circadian gene expression.

In this work, we built a mathematical model that describes mRNA dynamics under the regulation of rhythmic transcription, polyadenylation, deadenylation and degradation. We used the model to systematically examine how coupling of rhythmic transcription and poly(A) tail regulation generates rhythmicity in poly(A) tail length and mRNA abundance. Our model results highlighted rhythmicity in deadenylation as the strongest determinant for rhythmicity in poly(A)

tail length and in abundance of long-tailed mRNA. The latter can be regarded as a rough proxy for mRNA translatability because sufficiently long poly(A) tail is necessary for mRNA translation (Jacobson and Favreau 1983, Gallie 1991, Gallie 1998, Wells, Hillner et al. 1998, Weill, Belloc et al. 2012). As a corollary of this general finding, we demonstrated that the experimentally observed distinct phases in expression of the deadenylases (Kojima, Sher-Chen et al. 2012) can robustly group the rhythmic mRNAs by their peak phases in mRNA translatability (~long-tailed mRNA level), regardless of other rhythmic controls. This result implies that rhythmic deadenylation could provide a potential mechanism to orchestrate rhythmicity of gene expression across the genome. Finally, we used the model to predict the factors contributing to different classes of combinatorial patterns in the rhythmicity of transcription rate, poly(A) tail length and mRNA abundance (Kojima, Sher-Chen et al. 2012), which provides information for future experimental tests.

Results

Model

In a typical RNA expression process, RNAs are first transcribed in the nucleus and acquire long poly(A) tails as a result of nuclear polyadenylation. After being exported into the cytoplasm, mRNAs undergo deadenylation and are ultimately degraded. Cytoplasmic polyadenylation, as another important post-transcriptional regulation, elongates the poly(A) tail to promote mRNA stability and translatability (Mendez and Richter 2001). Although cytoplasmic polyadenylation is typically associated with translational control in oocyte maturation, early embryo development and synaptic plasticity (Mendez and Richter 2001, Kwak, Drier et al. 2008, Charlesworth, Meijer et al. 2013, Cui, Sartain et al. 2013), it is recently suggested to play a role in circadian rhythmicity of mRNA (Kojima, Sher-Chen et al. 2012). The

expression level of Gld2, a poly(A) polymerase responsible for cytoplasmic polyadenylation, was found to exhibit circadian rhythmicity (Kojima, Sher-Chen et al. 2012). Hence, in the model we incorporate polyadenylation, together with transcription, deadenylation and degradation, to capture the major processes that dynamically regulate poly(A) tail length and mRNA abundance (**Fig. 1**). Instead of explicitly tracking the exact length of poly(A) tails, the model divides the mRNA population into a long-tailed fraction and a short-tailed fraction, which compares with the same division in the experiment (long-tailed >~100nt, short-tailed <~100nt, (Kojima, Sher-Chen et al. 2012)).

For the sake of simplicity, we made the following assumptions in the model based on experimental evidence. First, degradation only occurs to the short-tailed mRNAs, because mRNA degradation typically requires the poly(A) tail length to be shortened to 10~15 nt (Schoenberg and Maquat 2012, Wu and Brewer 2012). Second, we lump together the nuclear events, i.e., transcription and nuclear polyadenylation, because nuclear polyadenylation is rather ubiquitous (Proudfoot 2011) and the poly(A) polymerases responsible for nuclear polyadenylation have not been found to be rhythmically expressed (Kojima, Sher-Chen et al. 2012). In our model, therefore, the transcription step directly leads to a long-tailed mRNA, and the downstream cytoplasmic deadenylation and polyadenylation are explicitly described. The ordinary differential equations that govern the temporal dynamics of long-tailed (L) and short-tailed (S) mRNAs read as Eqs. (1) and (2).

$$\text{Long-tailed mRNA:}\ \frac{dL}{dt} = \underbrace{\kappa_{\text{trsc}}(t)}_{\text{transcription}} - \underbrace{\kappa_{\text{deA}}(t)L}_{\text{deadenylation}} + \underbrace{\kappa_{\text{polyA}}(t)S}_{\text{polyadenylation}} \tag{1}$$

$$\text{Short-tailed mRNA:}\ \frac{dS}{dt} = \underbrace{\kappa_{\text{deA}}(t)L}_{\text{deadenylation}} - \underbrace{\kappa_{\text{polyA}}(t)S}_{\text{polyadenylation}} - \underbrace{\kappa_{\text{dgrd}}(t)S}_{\text{degradation}} \tag{2}$$

To capture the circadian regulation of the four processes in Eqs. (1) and (2), we represented each reaction rate term $\kappa(t)$ by a sinusoid function (Eq.(3)).

$$\kappa(t) = k\left(1 + A\cos\left(\omega t - \varphi\right)\right) \tag{3}$$

where k denotes the mean rate, A the relative amplitude, and φ the peak phase, of the process labeled by the subscript. The angular frequency, ω, equals $2\pi/(24\text{hr})$. ω is a fixed parameter, and the other parameters take different values.

In this paper we focus on how rhythmicity in the four modeled processes affects the rhythmicity in the abundance of mRNAs and the ratio between long-tailed and short-tailed mRNAs, because these are quantities measured in the previous circadian transcriptome studies (Koike, Yoo et al. 2012, Kojima, Sher-Chen et al. 2012). Furthermore, we are interested in the rhythmicity of abundance of long-tailed mRNAs, because elongation of the poly(A) tail activates translation (Jacobson and Favreau 1983, Gallie 1991, Gallie 1998, Wells, Hillner et al. 1998, Weill, Belloc et al. 2012). The abundance of the long-tailed subpopulation thus provides a rough proxy for mRNA translatability. From each simulation result based on Eqs. (1) and (2), i.e., time trajectories $L(t)$ and $S(t)$, we extracted the peak phases, relative amplitudes and mean levels of total mRNA abundance ($L(t) + S(t)$), long/short ratio ($LSR = L(t)/S(t)$), and long-tailed mRNA abundance, ($L(t)$) (see Methods). These quantities were subject to further analysis, as elaborated in the following Results sections. For the rest of the paper, we will refer to these quantities, e.g., peak phase of LSR, generally as the "output quantities", unless any specific quantity is referred to.

Rhythmic deadenylation is the dominant contributor to rhythmicity in poly(A) tail length and long-tailed mRNA abundance

To investigate the effects of coupled rhythmicity in transcription, degradation, polyadenylation and deadenylation on the output quantities, we ran 100,000 numeric simulations of the model (Eqs. (1) and (2)) with random parameter values for the mean rates, relative amplitudes, and phases of each process: transcription, deadenylation, degradation, and polyadenylation (**Table 1 and Fig. S1**), except for the mean rate of transcription, which only affected the overall abundance of mRNAs without any effect on the output quantities (**see Supplementary Materials**).

Our model results revealed rhythmicity in deadenylation as the strongest contributor to the rhythmicity of poly(A) tail length, followed by rhythmicity in polyadenylation. Over the sampled parameter sets, the scatter plots of the simulation results demonstrated that the peak phases of LSR were strongly concentrated within a ~3 hr-wide stripe centered around 10 hrs before the corresponding input peak phases of deadenylation (**Fig. 2A**), indicating a strong dependence of the LSR phase on the deadenylation phase. With a much more diffuse distribution of data points, the phase of LSR depended on the phase of polyadenylation much more weakly (**Fig. 2A**). The phases of transcription and degradation had very little correlation with the phase of LSR (**Fig. 2A**). We further performed variance-based sensitivity analysis (Sobol indices, (Sobol 2001, Iooss and Lemaitre 2015)) to systematically quantify the impacts of each model parameter on each output quantity. Based on simulation results from a large number of random parameter sets, each Sobol index quantifies the impact of a particular model parameter on a particular output quantity, reporting the percentage of variance in the output quantity contributed by the variance in the model parameter. The estimated Sobol indices (**Fig. 2B**) corroborated with the dependencies indicated by the scatter plots. For example, the peak phase of deadenylation was suggested by the Sobol indices to be the strongest contributor to the peak phase of LSR.

Variance in the peak phase of deadenylation alone explained ~50% of variance in the peak phase of LSR (longest bar in **Fig. 2B**). Consistently, the relative amplitude of deadenylation was also found to be the strongest contributor to the relative amplitude of LSR (**Fig. S2**), even though the mean level of LSR depended nearly equally on the mean rates of deadenylation and polyadenylation (**Fig. S2**).

Our model results also showed a moderate impact of rhythmic deadenylation and polyadenylation on the rhythmicity of total mRNA abundance. Although rhythmicity of total mRNA abundance, as expected, was most correlated with rhythmicity of transcription and degradation, variances in the peak phases of deadenylation and polyadenylation still made a moderate contribution to the variance in the peak phase of total mRNA (**Figs. 2C, D**). This contribution stemmed from the regulation of mRNA stability by poly(A) tail length. In the model, this regulation is depicted by the simplistic assumption that degradation is restricted to the short-tailed mRNAs (Eqs. (1) and (2)). For the same reason, relative amplitudes of deadenylation and polyadenylation contributed to the relative amplitudes of mRNA to an extent right next to the mean rate of degradation (**Fig. S4**).

We further used the model to examine the effects of the four processes on the rhythmicity of mRNA translatability, using the abundance of long-tailed mRNA as a proxy. Although it reflects the multiplication of total mRNA level and LSR, the rhythmicity of long-tailed mRNA abundance relied most heavily on rhythmicity of deadenylation, followed by that of polyadenylation (**Fig. 2F**). The peak phases of long-tailed mRNAs were strongly concentrated in the ~3 hr-wide strip centered around 10 hrs before the peak phases of deadenylation (**Fig. 2E**), highly similar to that observed in the peak phases of LSR (**Fig. 2A**). Consistently, the relative amplitudes of deadenylation and polyadenylation were also among the strongest contributors to

the relative amplitude of long-tailed mRNA (**Fig. S3**). Hence, deadenylation and polyadenylation were found to be the strongest contributors to the rhythmicity of mRNA translatability.

Taken together, these results underscore the importance of deadenylation and polyadenylation in regulating the rhythmicity of poly(A) tail length and mRNA abundance (especially abundance of the long-tailed subpopulation). We particularly found deadenylation the dominant contributor to the rhythmicity of poly(A) tail length and long-tailed mRNA abundance.

Rhythmic deadenylation could robustly group genes by their poly(A) tail rhythms

Rhythmicity in transcription, deadenylation, polyadenylation and degradation of mRNA are ultimately controlled by rhythmicity in the abundance or activity of molecules mediating these processes, e.g., transcription factors, deadenylases and poly(A) polymerases. Interestingly, a cytoplasmic poly(A) polymerase, *Gld-2*, is rhythmically expressed with peak phase around ZT3.5 (Kojima, Sher-Chen et al. 2012), whereas five deadenylases are rhythmically expressed, with *Ccr4e/Angel1* peaking around ZT2, *Ccr4a/Cnot6*, *Ccr4b/Cnot6l* and *Caf1a/Cnot7/pop2* peaking around ZT5, and *Ccr4c/Nocturnin* peaking around ZT13 (Kojima, Sher-Chen et al. 2012). Furthermore, although the core clock pathway includes several transcription factors with different peak phases, the peak phases of nascent RNA synthesis are strongly concentrated around ZT15 (Koike, Yoo et al. 2012). Taken together, the rhythmic patterns of deadenylases appear much more diverse than the rhythmic patterns of poly(A) polymerases and nascent RNA transcription.

Intrigued by the above observation, in this section we use our model to explore the potential functional role of having diverse peak phases in deadenylases. First, we plugged in the model the observed distribution of transcription peak phases centered around ZT15 (Koike, Yoo

et al. 2012). In three separate *in silico* experiments, we chose degradation, deadenylation or polyadenylation, and let the process peak at three narrow windows centered around ZT0, 8, 16, and let the peak phases of the other two processes evenly distribute around the clock (2nd, 3rd and 4th columns in **Fig. 3**). The case with three deadenylation peak windows significantly grouped the peak phases of LSR and long-tailed mRNA abundance into three windows (3rd column in **Fig. 3**), whereas the cases with degradation or polyadenylation peak windows generated results were comparable to the case where all three processes have evenly distributed peak phases (1st column in **Fig. 3**), with wide distributions in all peak phases. To test the effect of the actual rhythmic patterns observed in deadenylases and polyadenylases, we further plugged in narrow peak phase windows centered around ZT3.5 for polyadenylation, and around ZT2, ZT5 and ZT13 for deadenylation. The simulations resulted in three highly distinct groups of mRNAs in their peak phases of poly(A) tail length and long-tailed mRNA abundance (5th column in **Fig. 3**).

These results corroborate with the strong sensitivity of rhythmicity in poly(A) tail length and long-tailed mRNA abundance to the rhythmicity in deadenylation (**Fig. 2B, F**). Note that the mean rates and relative amplitudes of all the processes assumed random variables as in the previous section (**Table 1 and Fig. S1**). Therefore, our results showed that multiple peak phases in deadenylation, but not other processes, could robustly group the rhythmicity of poly(A) tail length and mRNA translatability (represented by long-tailed mRNA abundance) into distinct peak phase windows, regardless of variations in the mean activity or rhythmicity of other processes.

mRNA half-life and other factors explain difference between Class I and Class II PAR

mRNAs

In the previous transcriptome-wide study of rhythmicity in poly(A) tail lengths (Kojima, Sher-Chen et al. 2012), the mRNAs with rhythmic changes in the poly(A) tail length were grouped into three classes according to their rhythmic characteristics, i.e., whether 24-hour rhythmicity was detected in the levels of pre-RNA and total mRNA. The former is an indicator for transcription rate. The Class I mRNAs are rhythmic not only in poly(A) tail length, but also in pre-RNA and total mRNA (**Fig. 4A**). The Class II mRNAs are rhythmic in poly(A) tail length and pre-RNA, but not in total mRNA (**Fig. 4A**). The Class III mRNA are only rhythmic in poly(A) tail length, but not the other two (**Fig. 5A**). Factors like differences in mRNA half-lives were suggested in the original study to explain why the rhythmic patterns were different in these classes (Kojima, Sher-Chen et al. 2012). Here we leveraged our model to systematically identify factors that can lead to the combinatorial rhythmic patterns in these classes.

We first attempted to identify the model parameters that contributed most to the distinction between Class I and Class II. Because the only difference between Class I and II are the rhythmicity of total mRNAs, we focused on identifying model parameters that contribute most to the relative amplitude of total mRNA abundance. The Sobol indices revealed the mean rate constant of mRNA degradation as the strongest contributor to the amplitude of total mRNA (**Fig. 4B**). Indeed, the mRNA half-lives found in the cases with Class II behaviors exhibited a much wider distribution than those with Class I behaviors, extending towards larger values (**Fig. 4C**). This finding corroborates with the experimental observation that the average mRNA half-life is longer in Class II than in Class I (Kojima, Sher-Chen et al. 2012). After the mean rate constant of mRNA degradation, the Sobol indices also revealed the relative amplitudes of transcription and degradation, and the mean rate constants of deadenylation and polyadenylation,

as contributors to the amplitude of total mRNA (**Fig. 4B**). Higher mean rate constant of deadenylation also contributed to Class I (**Fig. 4D**). This is related to the above finding about mRNA half-lives because deadenylation promotes degradation and hence increasing mean rate constant of deadenylation has a similar effect on mRNA turnover as increasing the mean rate constant of degradation. The Class I behaviors are also associated with stronger amplitudes in transcription and degradation (**Fig. 4F, G**). These factors were indeed shown to significantly affect amplitude of mRNA abundance in a previous theoretical study on coupling of rhythmic transcription and degradation (Luck, Thurley et al. 2014). Furthermore, the Class I behaviors were found to associate with antiphasic rhythms in transcription and degradation (**Fig. 4H**). These additional factors differentiating Class I and Class II awaits future experimental testing.

Class III PAR mRNAs likely have longer mRNA half-life and non-rhythmic degradation

Compared with Class I and Class II, Class III does not have rhythmic transcription. Because rhythmicity of transcription serves as input to our model, we cannot use the model to investigate the origin of lack of transcriptional rhythmicity. However, we were interested in understanding why all PAR mRNAs without transcriptional rhythmicity also lack rhythmicity in total mRNA level, and only exhibit rhythmicity in LSR (Kojima, Sher-Chen et al. 2012). For the sake of discussion, we use “Class IV” to refer to the non-existing group of PAR mRNAs that would exhibit rhythmicity in total mRNAs and poly(A) tails, but not in pre-RNA (**Fig. 5A**). We tried to identify model parameters that could make the difference between Class III and the non-existing Class IV. Because both Class III and IV do not have rhythmic transcription, we ran 100,000 model simulations with non-rhythmic transcription (i.e., relative amplitude of transcription = 0, while keeping the other parameters sampled from the same distributions as

before (**Table 1 and Fig. S1**). Without rhythmic transcription, the resulting Sobol indices again revealed the mean rate of degradation as the most sensitive contributor to the relative amplitude of mRNA abundance, which explained the difference between Class III and IV (**Fig. 5B**). The Class IV behaviors required much shorter half-life of mRNA (**Fig. 5C**). Therefore, the absence of Class IV indicates that the group of mRNAs without rhythmic control at the transcriptional level likely have long mRNA half-life. Indeed, Class III has the longest average mRNA half-life measured among all mRNAs that are rhythmically expressed (Fig. 2D in (Kojima, Sher-Chen et al. 2012)). We also found the relative amplitude of degradation as the next strongest factor that differentiates Class III from IV (**Fig. 5B)**. The rhythmic characteristics of Class IV would require markedly higher amplitude of mRNA degradation (**Fig. 5D**). According to this result, the complete missing of Class IV PAR mRNAs implies that the group of mRNAs without transcriptional rhythmicity might also lack rhythmic degradation control, and hence all the mRNAs in this group exhibit the Class III, but not the Class IV behavior. Such RNAs likely depend on rhythmic poly(A) tail regulation to mediate downstream rhythmicity in protein translation. This conjecture awaits future experimental testing.

Discussion

Many posttranscriptional steps are shown to participate in circadian control of gene expression (Kojima, Shingle et al. 2011), among which rhythmic regulation of poly(A) tail length was recently discovered and characterized at the transcriptome level (Kojima, Sher-Chen et al. 2012). Notably, the rhythmicity in poly(A) tail is both the cause and result of rhythmic gene expression. On one hand, rhythmicity in poly(A) tail length controls rhythmicity in mRNA degradation (Luck, Thurley et al. 2014), because mRNA stability relies on poly(A) tail length

(Jacobson and Favreau 1983, Wells, Hillner et al. 1998, Eckmann, Rammelt et al. 2011, Weill, Belloc et al. 2012). On the other hand, rhythmicity in poly(A) tail length itself stems from rhythmicity in different gene expression processes, such as polyadenylation and deadenylation. Down to the mechanistic basis, it is the dynamic coupling between multiple rhythmic processes, i.e., transcription, polyadenylation, deadenylation and degradation, that generates the concomitant rhythmic patterns in poly(A) tail lengths and mRNA abundance. Such dynamic coupling strongly influences the rhythmic patterns in the gene product, as shown previously for coupling between rhythmic transcription and rhythmic degradation (Luck, Thurley et al. 2014).

In this work, we developed a simplistic mathematical model to study how rhythmic inputs from the transcription, degradation, polyadenylation and deadenylation processes control the output rhythmicity in mRNA abundance, poly(A) tail length and long-tailed mRNA abundance (as a proxy for mRNA translatability). Our model results highlighted rhythmic deadenylation as the strongest factor in controlling the peak phases and amplitudes of rhythmic poly(A) tail length and abundance of long-tailed mRNA (**Fig. 2**). This finding revealed a crucial role of rhythmic poly(A) regulation, especially rhythmic deadenylation, in rhythmic control of mRNA.

Following the identification of rhythmic deadenylation as the most important factor in controlling poly(A) tail length and long-tailed mRNA abundance, we used the model to test the potential effects of having three distinct peak phases in deadenylases as those suggested by previous experiments (Kojima, Sher-Chen et al. 2012). According to the model results, these three groups of deadenylases, coupled with the experimentally observed single peak phase of poly(A) polymerase (Kojima, Sher-Chen et al. 2012) and concentration of transcription peak phases around ZT15 (Koike, Yoo et al. 2012), can robustly divide peak phases in poly(A) tail

length and abundance of long-tailed mRNAs into three distinct groups, regardless of further rhythmic control (e.g., by rhythmic degradation) (**Fig. 3**). This finding indicates that rhythmic deadenylation may not only diversify the rhythmic patterns in mRNA functionality, but also organize the rhythmic genes into groups with close phases. Such grouping could potentially provide an effective mechanism to coordinate expression of genes that synergistically mediate physiological functions at distinct times of the day. Genes related to the same cellular or physiological functions could potentially be regulated by the same deadenylase, leading to similar phases in their expression and subsequent synergistic functioning. Our model further confirmed that such grouping effect is not significant if the same three peak phases are associated with the polyadenylation or degradation processes, instead of the deadenylation process. In other words, the diversity of rhythmicity in deadenylases is more effective than diversity in other rhythmic processes in imposing rhythmic grouping on mRNAs.

Our model suggested factors that likely contribute to the rhythmic characteristics of three observed classes of PAR mRNAs. Generally, multiple factors could affect the rhythmic characteristics of an mRNA, reflecting complexity and flexibility of the mechanisms controlling circadian gene expression. In principle, four classes of PAR mRNAs should exist, depending on whether total mRNA and preRNA are rhythmic or not (**Fig. 5A**). Only three classes, however, were reported previously (Kojima, Sher-Chen et al. 2012). For PAR mRNAs without rhythmicity in preRNA, they ubiquitously lack rhythmicity in total abundance of mRNA. In addition to the previous suggestion of longer half-life (Kojima, Sher-Chen et al. 2012), our model analysis suggested non-rhythmic mRNA degradation as another possible contributing factor to this phenomenon (**Fig. 5B, D**). It is up to future experiments to test the effects of factors predicted by the model that lead to different classes.

The finding that rhythmic deadenylation plays a dominant role in regulating rhythmicity of poly(A) tail and mRNA translatability poses two interesting questions. First, could deadenylation help bridge the circadian rhythm with other rhythmicity? For example, genes related to poly(A) tail regulation, such as *Parn*, one of the rhythmically expressed deadenylases, are suggested to gain rhythmicity from clock-independent signals (Wang, Symul et al. 2018). Equipped with a strong impact on mRNA rhythmicity, such deadenylation process could potentially help couple the circadian rhythm to other rhythmic signals, e.g., feeding, to optimize physiological functions around the clock. Second, how do different deadenylases interact with each other in the regulation poly(A) tail rhythmicity? In other words, could deadenylases that peak at different phases regulate the same mRNA and impose a more complex rhythmic pattern? These questions need future studies to resolve.

Methods

Model simulation and extraction of phase, amplitude and mean from simulation results

For any given parameter set, Eqs. (1) and (2) were simulated using the ordinary differential equation (ODE) solver, ode45, in MATLAB. For a simulated time trajectory $\{L(t), S(t)\}$, the peak phases, relative amplitudes and mean levels of $L(t) + S(t), L(t)/S(t)$ and $L(t)$ were analyzed. First, the time trajectory for the output quantity of interest, e.g., $L(t)/S(t)$, was calculated from $\{L(t), S(t)\}$. Then a 48-hr window after 700 hrs (sufficiently long to pass the initial transient) were extracted from the simulated trajectory for data analysis. The trajectories typically have irregular time spacing (due to automatic time stepping in the ode45 solver) and often have insufficient time resolution for accurate determination of the peak phase. To improve the accuracy of the estimated peak phase, the 48-hr trajectory were interpolated upon 500 equally spaced time points spanning the 48 hrs. The peak

phase was evaluated from the time for the maximum value, t_{max}, i.e., $\text{peak phase} = mod(t_{max} + 700, 24)$ (hr). The mean value was estimated by taking the average of the 500 interpolated data values within the 48-hr window. The relative amplitude was estimated by taking the maximum and minimum values within the 48-hr window and calculating $(\max - \min)/(2 \times \text{mean})$.

Parameter sampling

We performed global parameter sensitivity analysis (Saltelli 2008) on the model to analyze the general contribution of each parameter to any specific output quantity (i.e., peak phase, relative amplitude and mean of $L(t) + S(t), L(t)/S(t)$ and $L(t)$). To perform such global sensitivity analysis, one needs to simulate the model with randomly chosen parameter values that maximally represent the parameter space. In this study we drew random parameter values from the distributions listed in **Table 1** and plotted in **Fig. S1**. The peak phases and relative amplitudes were sampled form uniform distributions of their possible ranges by definition (**Table**). The mean reaction rates were sampled from log-normal distributions suggested by previous genomic scale measurements (see sources indicated in **Table 1**). It is noteworthy that we set the mean transcription rate as constant, as it only causes proportional changes in $L(t)$ and $S(t)$, and does not affect the rhythmic patterns of any quantity (see Supplementary Materials). To improve the accuracy of the global sensitivity analysis for models with many parameters, one needs sampled parameter values that well represent the parameter space. To this end, we used the sampling method of Sobol sequence, which was known to ensure good representation of the high-dimensional parameter space (Iooss and Lemaitre 2015).

Table 1. Parameter distribution for sampling.

Parameter	Symbol	Distribution	Source
Mean rate of transcription	k_{trsc}	1 (constant)	Has no effect on rhythmic patterns (see **Supplementary Materials**)
Mean rate of degradation	k_{dgrd}	$\ln(k/\mathrm{hr}^{-1}) \sim \mathcal{N}(0.033, 0.23^2)$ (Log-normal)	Fitting with half-life distribution measured in (Schwanhausser, Busse et al. 2011)
Mean rate of deadenylation	k_{deA}	$\ln(k/\mathrm{hr}^{-1}) \sim \mathcal{N}(0.15, 0.23^2)$ (Log-normal)	Mean value of deadenylation rate taken from (Yamashita, Chang et al. 2005); deviation same as mRNA degradation
Mean rate of polyadenylation	k_{polyA}	$\ln(k/\mathrm{hr}^{-1}) \sim \mathcal{N}(0.15, 0.23^2)$ (Log-normal)	Same as deadenylation
Relative amplitudes	$A_{\mathrm{trsc}}, A_{\mathrm{dgrd}}, A_{\mathrm{deA}}, A_{\mathrm{polyA}}$	$A \sim U(0,1)$ (Uniform)	
Peak phases	$\varphi_{\mathrm{trsc}}, \varphi_{\mathrm{dgrd}}, \varphi_{\mathrm{deA}}, \varphi_{\mathrm{polyA}}$	$\varphi \sim U(0,24)$ (Uniform)	

Sensitivity analysis

To evaluate the impact of each model parameter (e.g., phase of deadenylation) on each model output (e.g., relative amplitude of LSR), we used a variance-based global parameter sensitivity analysis method, the Sobol indices (Iooss and Lemaitre 2015). For each pair of

parameter X_i and output Y_j in the model, the first-order Sobol index, S_{ij}, characterizes the contribution of variance in X_i alone to the total variance in Y_j (Eq.(4)). The total-effect index, S_{Tij}, characterizes the contribution of variance in X_i, as well as the variance caused by its coupling with other parameters, to the total variance in Y_j (Eq.(5)). The larger these indices are, the more sensitive Y_j is to X_i, or the more impact X_i has on Y_j.

$$S_{ij} = \frac{\mathrm{Var}_{X_i}\left(\mathrm{E}_{X_{\sim i}}\left(Y_j \mid X_i\right)\right)}{\mathrm{Var}\left(Y_j\right)} \tag{4}$$

$$S_{Tij} = \frac{\mathrm{E}_{X_{\sim i}}\left(\mathrm{Var}_{X_i}\left(Y_j \mid X_{\sim i}\right)\right)}{\mathrm{Var}\left(Y_j\right)} \tag{5}$$

where the subscript $\sim i$ indicates all indices except for i.

We followed the specific algorithms given in (Sobol 2001) and (Saltelli, Annoni et al. 2010) for evaluating the first-order indices (Eq.(4)) and total-effect indices (Eq.(5)), respectively. The calculation of variance was slightly modified for the peak phases, the values of which live on a circular domain. This results in a sudden jump of the numeric values of the phase at the ends of the circular domain, e.g., 24hr to 0hr, even though it reflects a continuous change. This problem can artificially increase the estimated variance. To correct the problem, the difference between phases, $\varphi_1 - \varphi_2$, was replaced by its sine value, $\sin(\varphi_1 - \varphi_2)$, in the evaluation of the Sobol indices.

References

Beilharz, T. H. and T. Preiss (2007). "Widespread use of poly(A) tail length control to accentuate expression of the yeast transcriptome." Rna-a Publication of the Rna Society **13**(7): 982-997.

Charlesworth, A., H. A. Meijer and C. H. de Moor (2013). "Specificity factors in cytoplasmic polyadenylation." Wiley Interdiscip Rev RNA **4**(4): 437-461.

Cui, J., C. V. Sartain, J. A. Pleiss and M. F. Wolfner (2013). "Cytoplasmic polyadenylation is a major mRNA regulator during oogenesis and egg activation in Drosophila." Developmental

Biology **383**(1): 121-131.

Du, N. H., A. B. Arpat, M. De Matos and D. Gatfield (2014). "MicroRNAs shape circadian hepatic gene expression on a transcriptome-wide scale." Elife **3**: e02510.

Duffield, G. E. (2003). "DNA microarray analyses of circadian timing: the genomic basis of biological time." J Neuroendocrinol **15**(10): 991-1002.

Eckmann, C. R., C. Rammelt and E. Wahle (2011). "Control of poly(A) tail length." Wiley Interdiscip Rev RNA **2**(3): 348-361.

Gallie, D. R. (1991). "The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency." Genes Dev **5**(11): 2108-2116.

Gallie, D. R. (1998). "A tale of two termini: a functional interaction between the termini of an mRNA is a prerequisite for efficient translation initiation." Gene **216**(1): 1-11.

Gerstner, J. R., W. M. Vanderheyden, T. LaVaute, C. J. Westmark, L. Rouhana, A. I. Pack, M. Wickens and C. F. Landry (2012). "Time of day regulates subcellular trafficking, tripartite synaptic localization, and polyadenylation of the astrocytic Fabp7 mRNA." J Neurosci **32**(4): 1383-1394.

Gobet, C. and F. Naef (2017). "Ribosome profiling and dynamic regulation of translation in mammals." Curr Opin Genet Dev **43**: 120-127.

Hirano, A., Y. H. Fu and L. J. Ptacek (2016). "The intricate dance of post-translational modifications in the rhythm of life." Nat Struct Mol Biol **23**(12): 1053-1060.

Iooss, B. and P. Lemaitre (2015). A Review on Global Sensitivity Analysis Methods. Uncertainty Management in Simulation-Optimization of Complex Systems. G. Dellino and C. Meloni. New York, Springer Science+Business Media.

Jacobson, A. and M. Favreau (1983). "Possible involvement of poly(A) in protein synthesis." Nucleic Acids Res **11**(18): 6353-6368.

Jalkanen, A. L., S. J. Coleman and J. Wilusz (2014). "Determinants and implications of mRNA poly(A) tail size--does this protein make my tail look big?" Semin Cell Dev Biol **34**: 24-32.

Keene, J. D. (2010). "Minireview: global regulation and dynamics of ribonucleic Acid." Endocrinology **151**(4): 1391-1397.

Kim, T. D., J. S. Kim, J. H. Kim, J. Myung, H. D. Chae, K. C. Woo, S. K. Jang, D. S. Koh and K. T. Kim (2005). "Rhythmic serotonin N-acetyltransferase mRNA degradation is essential for the maintenance of its circadian oscillation." Mol Cell Biol **25**(8): 3232-3246.

Koike, N., S. H. Yoo, H. C. Huang, V. Kumar, C. Lee, T. K. Kim and J. S. Takahashi (2012). "Transcriptional Architecture and Chromatin Landscape of the Core Circadian Clock in Mammals." Science **338**: 349.

Koike, N., S. H. Yoo, H. C. Huang, V. Kumar, C. Lee, T. K. Kim and J. S. Takahashi (2012). "Transcriptional architecture and chromatin landscape of the core circadian clock in mammals." Science **338**(6105): 349-354.

Kojima, S., D. Gatfield, C. C. Esau and C. B. Green (2010). "MicroRNA-122 modulates the rhythmic expression profile of the circadian deadenylase Nocturnin in mouse liver." PLoS One **5**(6): e11264.

Kojima, S. and C. B. Green (2015). "Circadian genomics reveal a role for post-transcriptional regulation in mammals." Biochemistry **54**(2): 124-133.

Kojima, S., E. L. Sher-Chen and C. B. Green (2012). "Circadian control of mRNA polyadenylation dynamics regulates rhythmic protein expression." Genes Dev **26**(24): 2724-2736.

Kojima, S., D. L. Shingle and C. B. Green (2011). "Post-transcriptional control of circadian rhythms." J Cell Sci **124**(Pt 3): 311-320.

Kwak, J. E., E. Drier, S. A. Barbee, M. Ramaswami, J. C. P. Yin and M. Wickens (2008). "GLD2 poly(A) polymerase is required for long-term memory." Proceedings of the National Academy of Sciences of the United States of America **105**(38): 14644-14649.

Luck, S., K. Thurley, P. F. Thaben and P. O. Westermark (2014). "Rhythmic degradation explains and unifies circadian transcriptome and proteome data." Cell Rep **9**(2): 741-751.

Mendez, R. and J. D. Richter (2001). "Translational control by CPEB: A means to the end." Nature Reviews Molecular Cell Biology **2**(7): 521-529.

Menet, J. S., J. Rodriguez, K. C. Abruzzi and M. Rosbash (2012). "Nascent-Seq reveals novel features of mouse circadian transcriptional regulation." Elife **1**: e00011.

Mohawk, J. A., C. B. Green and J. S. Takahashi (2012). "Central and peripheral circadian clocks in mammals." Annu Rev Neurosci **35**: 445-462.

Partch, C. L., C. B. Green and J. S. Takahashi (2014). "Molecular architecture of the mammalian circadian clock." Trends Cell Biol **24**(2): 90-99.

Proudfoot, N. J. (2011). "Ending the message: poly(A) signals then and now." Genes Dev **25**(17): 1770-1782.

Robinson, B. G., D. M. Frim, W. J. Schwartz and J. A. Majzoub (1988). "Vasopressin mRNA in the suprachiasmatic nuclei: daily regulation of polyadenylate tail length." Science **241**(4863): 342-344.

Saltelli, A. (2008). Global sensitivity analysis : the primer. Chichester, England ;, John Wiley.

Saltelli, A., P. Annoni, I. Azzini, F. Campolongo, M. Ratto and S. Tarantola (2010). "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index." Computer Physics Communications **181**(2): 259-270.

Schoenberg, D. R. and L. E. Maquat (2012). "Regulation of cytoplasmic mRNA decay." Nat Rev Genet **13**(4): 246-259.

Schwanhausser, B., D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen and M. Selbach (2011). "Global quantification of mammalian gene expression control." Nature **473**(7347): 337-342.

Sobol, I. M. (2001). "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates." Mathematics and Computers in Simulation **55**(1-3): 271-280.

Stubblefield, J. J., P. Gao, G. Kilaru, B. Mukadam, J. Terrien and C. B. Green (2018). "Temporal Control of Metabolic Amplitude by Nocturnin." Cell Rep **22**(5): 1225-1235.

Takahashi, J. S. (2015). "Molecular components of the circadian clock in mammals." Diabetes

Obes Metab **17 Suppl 1**: 6-11.

Takahashi, J. S., H. K. Hong, C. H. Ko and E. L. McDearmon (2008). "The genetics of mammalian circadian order and disorder: implications for physiology and disease." Nat Rev Genet **9**(10): 764-775.

Trott, A. J. and J. S. Menet (2018). "Regulation of circadian clock transcriptional output by CLOCK:BMAL1." PLoS Genet **14**(1): e1007156.

Ueda, H. R., S. Hayashi, W. Chen, M. Sano, M. Machida, Y. Shigeyoshi, M. Iino and S. Hashimoto (2005). "System-level identification of transcriptional circuits underlying mammalian circadian clocks." Nat Genet **37**(2): 187-192.

Wang, J., L. Symul, J. Yeung, C. Gobet, J. Sobel, S. Luck, P. O. Westermark, N. Molina and F. Naef (2018). "Circadian clock-dependent and -independent posttranscriptional regulation underlies temporal mRNA accumulation in mouse liver." Proc Natl Acad Sci U S A.

Wang, Y., D. L. Osterbur, P. L. Megaw, G. Tosini, C. Fukuhara, C. B. Green and J. C. Besharse (2001). "Rhythmic expression of Nocturnin mRNA in multiple tissues of the mouse." BMC Dev Biol **1**: 9.

Weill, L., E. Belloc, F. A. Bava and R. Mendez (2012). "Translational control by changes in poly(A) tail length: recycling mRNAs." Nature Structural & Molecular Biology **19**(6): 577-585.

Wells, S. E., P. E. Hillner, R. D. Vale and A. B. Sachs (1998). "Circularization of mRNA by eukaryotic translation initiation factors." Mol Cell **2**(1): 135-140.

Wu, X. and G. Brewer (2012). "The regulation of mRNA stability in mammalian cells: 2.0." Gene **500**(1): 10-21.

Yamashita, A., T. C. Chang, Y. Yamashita, W. M. Zhu, Z. P. Zhong, C. Y. A. Chen and A. B. Shyu (2005). "Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover." Nature Structural & Molecular Biology **12**(12): 1054-1063.

Zhang, E. E. and S. A. Kay (2010). "Clocks not winding down: unravelling circadian networks." Nat Rev Mol Cell Biol **11**(11): 764-776.

Zhang, R., N. F. Lahens, H. I. Ballance, M. E. Hughes and J. B. Hogenesch (2014). "A circadian gene expression atlas in mammals: implications for biology and medicine." Proc Natl Acad Sci U S A **111**(45): 16219-16224.

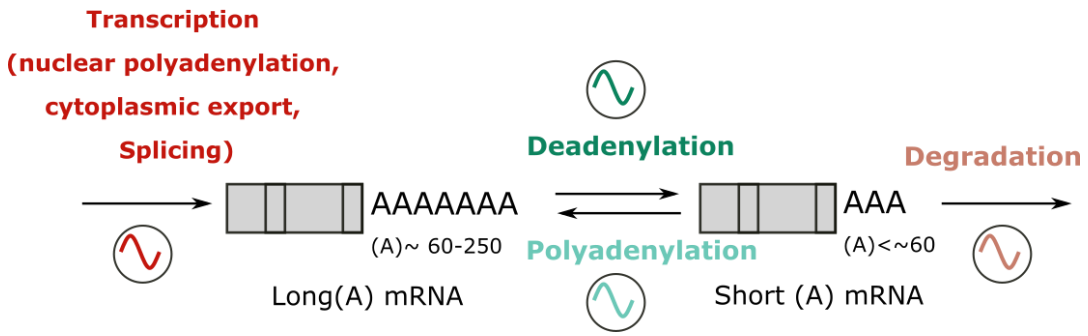

Figure 1. Schematic diagram of the model. The model describes four processes which control the poly(A) tail length and mRNA abundance: transcription, degradation, cytoplasmic deadenylation and polyadenylation. All four processes are allowed variable rhythmicity.

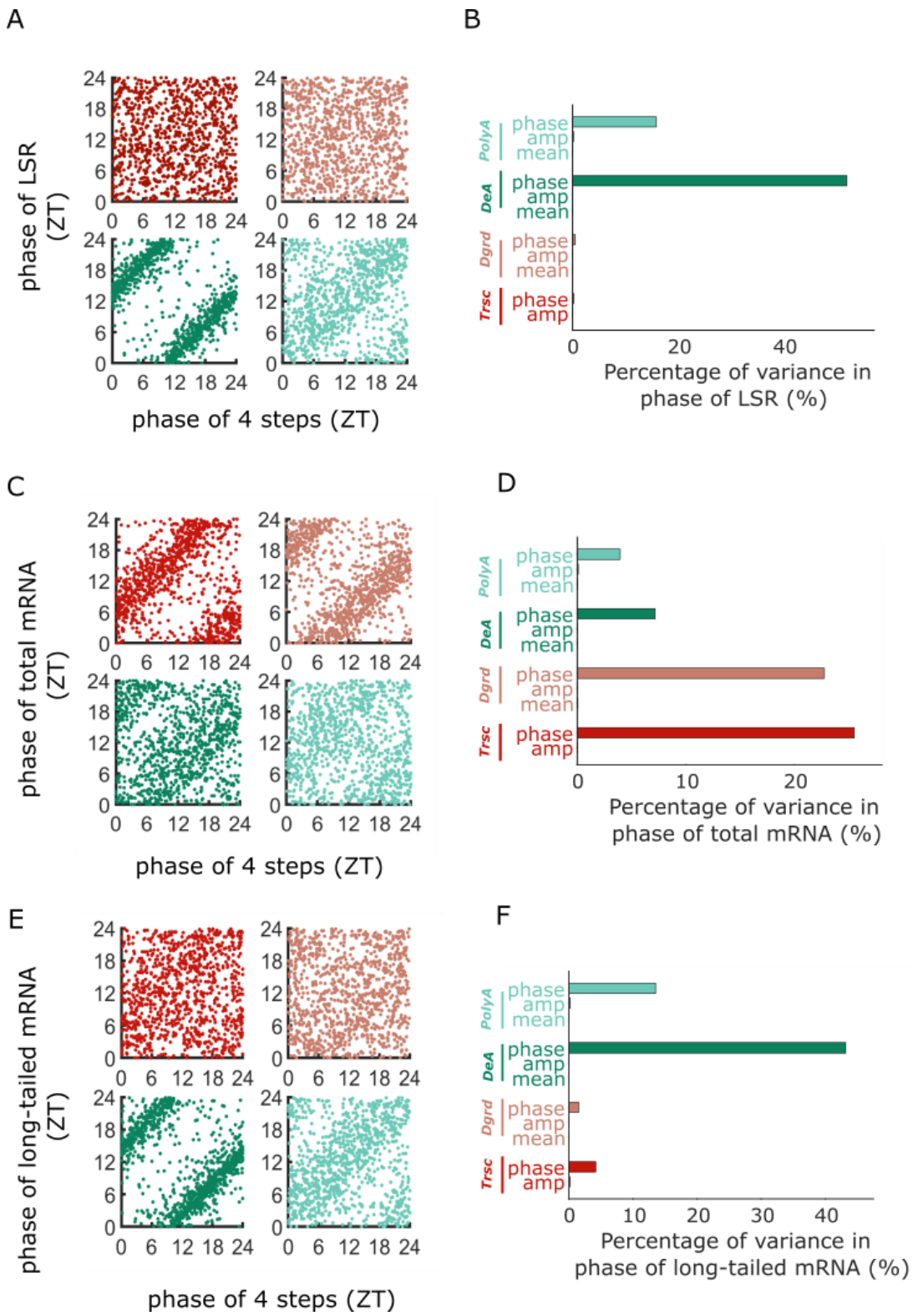

Figure 2. **Rhythmicity of poly(A) tail length and long-tailed mRNA abundance is strongly controlled by rhythmic deadenylation.** (A) Scatter plot of the peak phases of input processes

versus the peak phases of LSR. (B) Sobol indices for the peak phase of LSR. (C) Scatter plot of the peak phases of input processes versus the peak phases of total mRNA abundance. (D) Sobol indices of the peak phase of total mRNA abundance. (E) Scatter plot of the peak phases of input processes versus the peak phases of long-tailed mRNA abundance. (F) Sobol indices of the peak phase of long-tailed mRNA abundance. Each scatter plot shows 10,000 data points randomly chosen from the original 100,000 simulations for the sake of clarity. Parameters of simulations are further described in the Supplemental Materials.

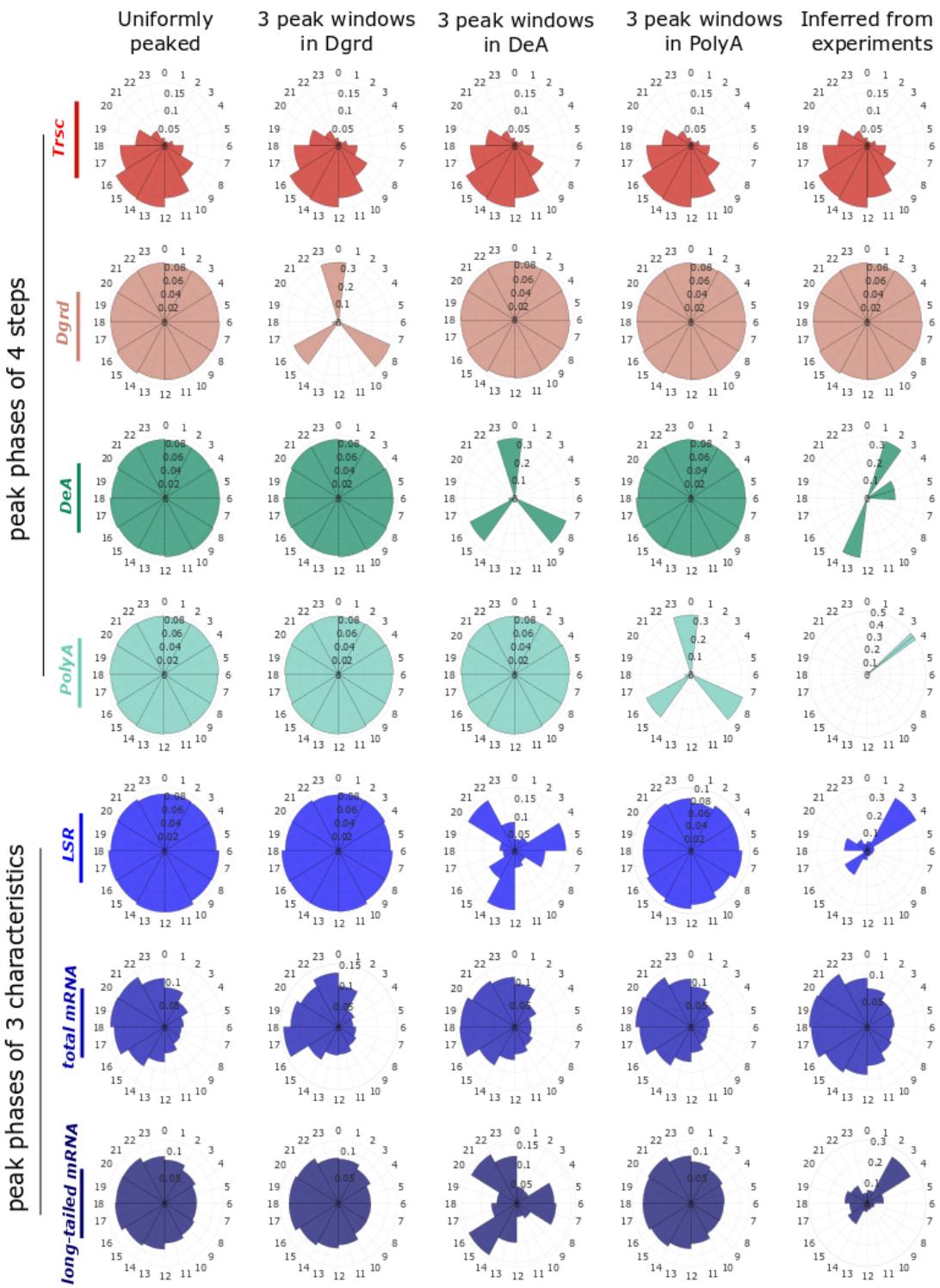

Figure 3. Peak phase windows in deadenylases group transcripts according to peak phases

of poly(A) tail length and long-tailed mRNA abundance. Upper four rows: distributions of the input peak phases of transcription (Trsc), degradation (Dgrd), deadenylation (DeA) and polyadenylation (PolyA). Bottom three rows: distributions of the output peak phases of LSR, total mRNA and long-tailed mRNA. Each column corresponds to a specific scenario for distribution of input peak phases as labeled on the top.

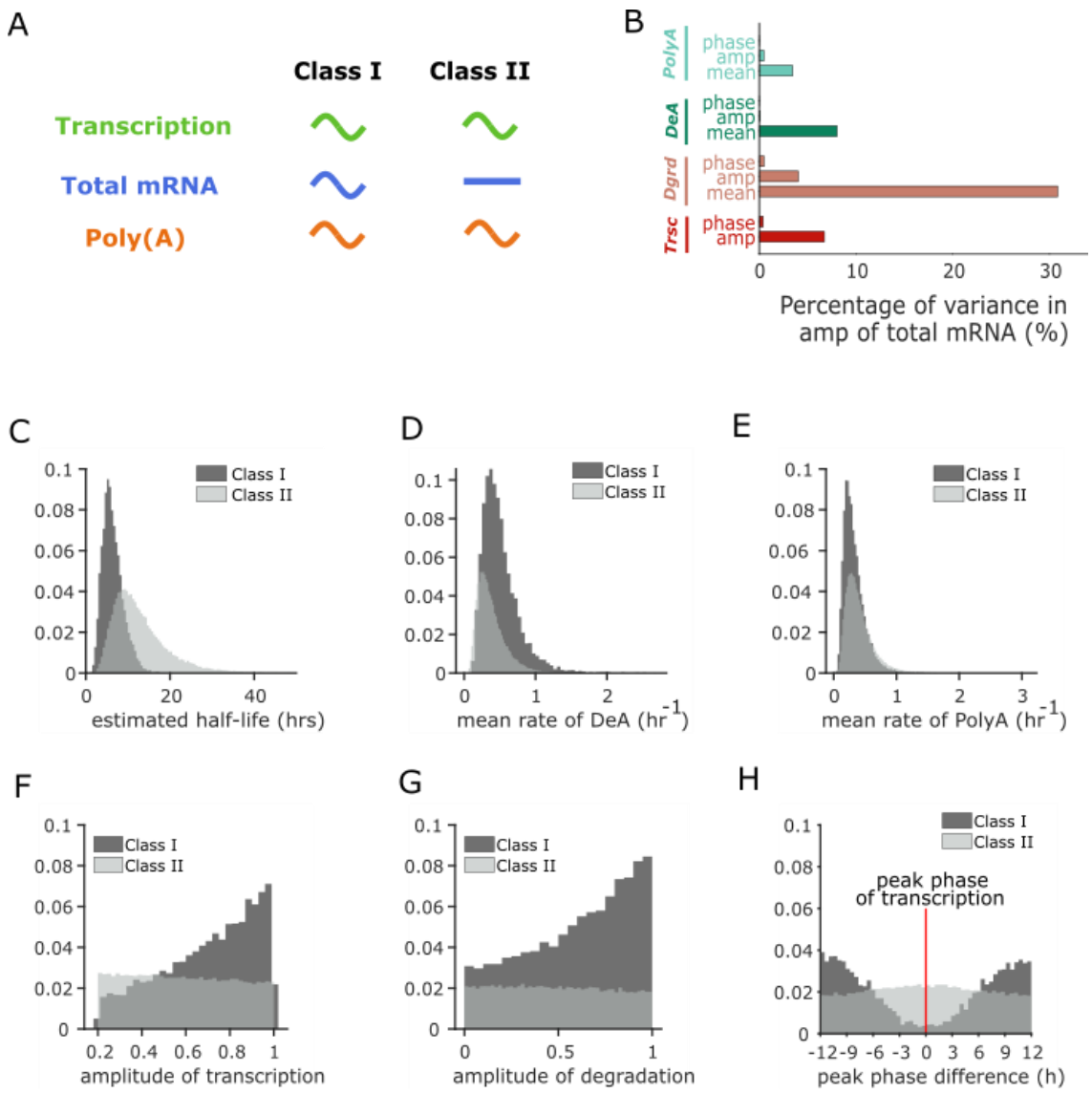

Figure 4. Factors distinguishing Class I and Class II PAR mRNAs. (A) Characteristic behaviors of Class I and Class II PAR mRNAs. (B) Sobol indices for the amplitude of total mRNA. (C) Distributions of mRNA half-lives for the two classes. (D) Distributions of mean deadenylation rate for the two classes. (E) Distributions of mean polyadenylation rate for the two classes. (F) Distribution of the relative amplitude of transcription for the two classes. (G) Distributions of the relative amplitude of degradation for the two classes. (H) Distribution of peak phase difference between transcription and degradation for the two classes. Results from

100,000 simulations with parameters randomly sampled according to Table 1. Class I is defined for the cases with >=0.2 relative amplitude in total mRNA abundance, while Class II is defined for the cases with <0.2 relative amplitude in total mRNA abundance.

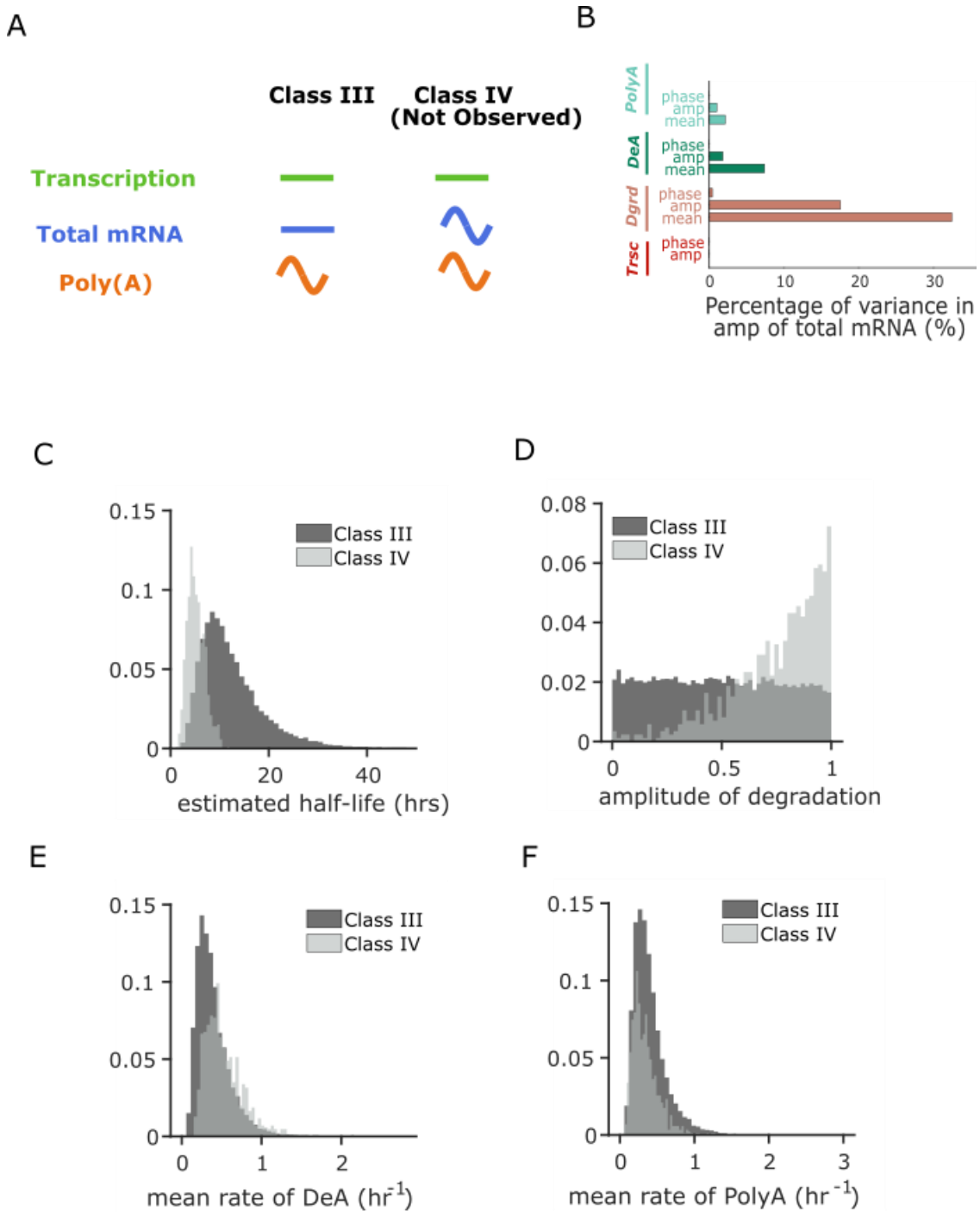

Figure 5. Factors distinguishing Class III and Class IV PAR mRNAs. (A) Characteristic behaviors of Class III and the hypothetical Class IV PAR mRNAs. (B) Sobol indices of the amplitude of total mRNA for a model without rhythmic transcription. (C) Distributions of

mRNA half-lives for the two classes. (D) Distributions of the relative amplitude of degradation for the two classes. (E) Distributions of the mean deadenylation rate for the two classes. (F) Distributions of the mean polyadenylation rate for the two classes. Results from 100,000 simulations with parameters randomly sampled according to Table 1, but without rhythmic transcription ($A_{\mathrm{trsc}} = 0$). Class III is defined for the cases with >=0.2 relative amplitude in total mRNA abundance, while Class IV is defined for the cases with <0.2 relative amplitude in total mRNA abundance.