

METHODOLOGY ARTICLE

Open Access



# The role of alternative Polyadenylation in regulation of rhythmic gene expression

Natalia Ptitsyna<sup>1</sup>, Sabri Boughorbel<sup>2</sup>, Mohammed El Anbari<sup>2</sup> and Andrey Ptitsyn<sup>2,3\*</sup>

## Abstract

**Background:** Alternative transcription is common in eukaryotic cells and plays important role in regulation of cellular processes. Alternative polyadenylation results from ambiguous PolyA signals in 3' untranslated region (UTR) of a gene. Such alternative transcripts share the same coding part, but differ by a stretch of UTR that may contain important functional sites.

**Methods:** The methodology of this study is based on mathematical modeling, analytical solution, and subsequent validation by datamining in multiple independent experimental data from previously published studies.

**Results:** In this study we propose a mathematical model that describes the population dynamics of alternatively polyadenylated transcripts in conjunction with rhythmic expression such as transcription oscillation driven by circadian or metabolic oscillators. Analysis of the model shows that alternative transcripts with different turnover rates acquire a phase shift if the transcript decay rate is different. Difference in decay rate is one of the consequences of alternative polyadenylation. Phase shift can reach values equal to half the period of oscillation, which makes alternative transcripts oscillate in abundance in counter-phase to each other. Since counter-phased transcripts share the coding part, the rate of translation becomes constant. We have analyzed a few data sets collected in circadian timeline for the occurrence of transcript behavior that fits the mathematical model.

**Conclusion:** Alternative transcripts with different turnover rate create the effect of rectifier. This "molecular diode" moderates or completely eliminates oscillation of individual transcripts and stabilizes overall protein production rate. In our observation this phenomenon is very common in different tissues in plants, mice, and humans. The occurrence of counter-phased alternative transcripts is also tissue-specific and affects functions of multiple biological pathways. Accounting for this mechanism is important for understanding the natural and engineering the synthetic cellular circuits.

**Keywords:** Alternative transcription, Oscillatory gene expression, Cellular circuits, Molecular diode, mathematical modeling, datamining

## Background

Circadian oscillation plays important role in regulation of gene expression. The number of reported cycling genes differs from study to study. Some publications report hundreds [1–3] others thousands [4] of oscillating transcripts, depending on experiment design and analysis of data. Some reports insist on majority or the entire transcriptome experiencing diurnal oscillations [5, 6]. In any case, a considerable fraction of rhythmically expressed genes is bound to modulate the activity in multiple biological pathways. Multiple other factors

are known to regulate gene expression in the context of biological pathways. Alternative polyadenylation is one of such factors, sometimes considered a form of alternative splicing, but rarely mentioned in connections with circadian oscillation. Recent reviews, such as [7, 8] provide a panoramic overview of the prominent role of alternative polyadenylation in various healthy and disease states, but make no connection to the rhythmic alternations in alternative transcript population. However, some studies point specifically to the importance of such connection [9] in *Arabidopsis thaliana* and *Drosophila melanogaster* [10]. Others point at the role of polyadenylation in regulation of rhythmic protein expres-

\* Correspondence: ptitsyn@aol.com

<sup>2</sup>Sidra Medical and Research Center, P.O. box 26999 Doha, Qatar

<sup>3</sup>Present affiliation: Gloucester Marine Genomics Institute, Gloucester, MA 01930, USA

Full list of author information is available at the end of the article



sion [11, 12] while observing the length of PolyA tail in various transcripts in mice. In some estimations up to 35% of all alternative 3'UTR transcripts may have different turnover rate [8]. Generalizing these observations we come to a conclusion that transcription factors are not the only mechanism regulating circadian expression. Post-transcriptional mechanisms such as alternative splicing, polyadenylation, nonsense-mediated decay, etc. are also important in formation of dynamic pattern of transcripts. In this connection we would like to remember one old study reporting a perplexing pattern of alternative transcripts of suppressor of cytokine signaling (SOCS3) in mice oscillating in opposite phase to each other [13]. The paper described the occurrence of alternative microarray probes traced to alternative transcripts sharing the coding part, but resulting from alternative polyadenylation sites. This effect was first discovered in JAK-STAT (Janus Kinase - Signal Transducer and Activator of Transcription) signal transduction pathway. Counter-phased alternative transcripts were observed in brown adipose tissue, but not in white adipose or liver samples. Other elements of the same pathway such as JAK were also showing counter-phase transcripts in one tissue, but not the other. The study proposed that such pattern of alternative transcripts may represent an adaptive mechanism regulating the pathway in a tissue-specific manner by creating a constant abundance of a particular protein. For example, constant production of SOCS3 from alternating short and long transcripts can block signal transduction in a particular tissue regardless of the diurnal change of the baseline. In the current study we attempt to generalize this observation and propose the model of molecular mechanism responsible for the observed pattern of counter-phase oscillation of alternative transcripts.

## Results

### Mathematical model

Let  $n_1(t)$  denote the change in abundance (for instance, relative to invariant sum of intensity of microarray control spots) of the long isoform in time and  $n_2(t)$  stands for the change abundance of the short isoform of the transcript  $n$  in time.

Let  $r_p$  describe the expression rate of the gene from which both isoforms are transcribed. Since they share the same promoter and all other functional sites except 3' UTR polyadenylation signal, the rate is the same for both short and long transcripts. Let  $p$  denote the probability of transcription resulting in production of the long isoform. Then  $1-p$  is the probability of transcription resulting in the short isoform. The UTRs of these transcripts are different, thus we introduce separate variables for the degradation rates:

$r_{d1}$  describes the degradation rate for the long isoform of transcript  $n_1$

$r_{d2}$  describes the degradation rate for the long isoform of transcript  $n_2$

$$\begin{cases} \frac{dn_1}{dt} = pr_p - r_{d1}; \\ \frac{dn_2}{dt} = (1-p)r_p - r_{d2}; \end{cases} \quad (1)$$

Let's consider the case when the baseline of expression is modulated by a simple harmonic process, such as circadian rhythm. Since the entire cell (or even the organism consisting of magnitude of cells) is modulated by the same factors, we consider the period of oscillation equal in all equations. The baseline oscillation is described by the travelling wave equation

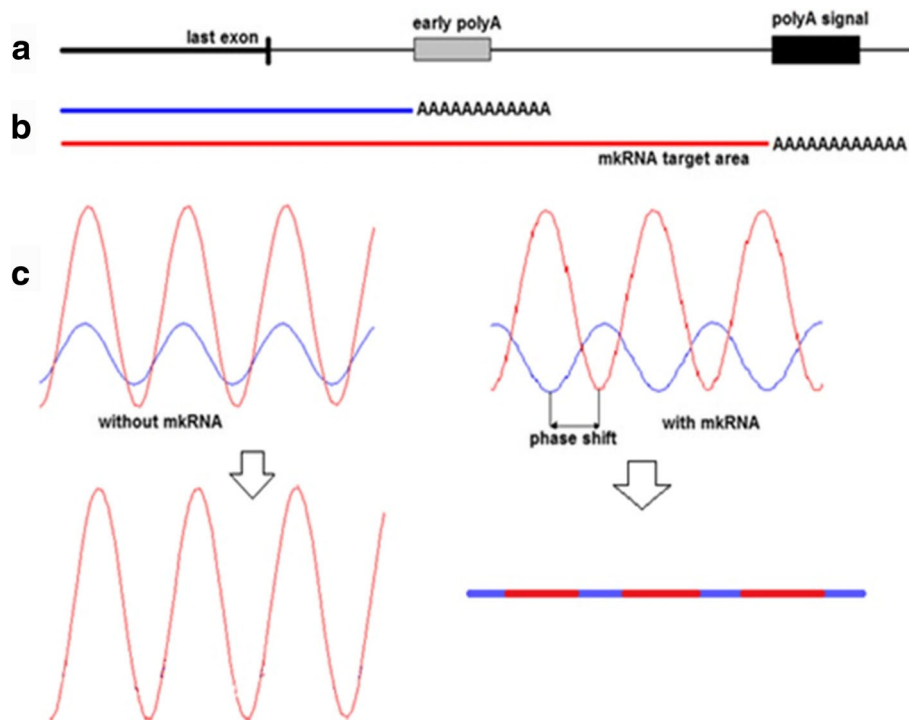
$$\begin{aligned} r_p &= a \sin(\omega t + \alpha_1); r_{d1} = b \sin(\omega t + \alpha_2); r_{d2} \\ &= c \sin(\omega t + \alpha_3); \end{aligned} \quad (2)$$

Here we assume that  $b > c$ , which means that longer transcripts have a shorter life span. This assumption models the action of miRNA that can bind the longer transcript and facilitate the decay. The shorter isoform lacks the miRNA binding site and thus can last longer, transcribing more copies of encoded protein. The overall model takes the following form:

$$\begin{cases} \frac{dn_1}{dt} = pa \sin(\omega t + \alpha_1) - b \sin(\omega t + \alpha_2); \\ \frac{dn_2}{dt} = (1-p)a \sin(\omega t + \alpha_1) - c \sin(\omega t + \alpha_3); \end{cases} \quad (3)$$

The formula (14) (see Methods for complete analytic solution) allows direct calculation of the phase shift from the estimated degradation rates of short and long isoforms. These values can be estimated experimentally.

Summing up isoforms  $n_1$  and  $n_2$  we can estimate the overall level of expression and amplitude of oscillation for the entire population of alternative transcripts of gene  $n$ . While  $n_1 + n_2 = n$  at all times, the amplitude of the resulting curve for  $n$  depends on the phase shift between isoforms  $n_1$  and  $n_2$ . The phase lag between isoforms may have values varying between 0 and  $2\pi$ . In the middle of this range, when  $\beta_2 - \beta_1 = \pi$  the amplitude of  $n$  is reduced to 0. In terms of biology, this means that gene expression oscillatory in nature at the origin can produce a constant steady production of peptides using the mechanism of differentially polyadenylated transcriptional isoforms. This mechanism provides the "power rectifier" element for the cellular circuitry. Figure 1 illustrates the action of molecular circuit rectifier. The degradation rate of mRNA which determines the turnover rate of mRNA copies and eventually the amount of synthesized protein can be affected by many factors, such as post-transcriptional modification, mRNA transport, tertiary structure, etc. However, the most well-known factor is the action of miRNA.



**Fig. 1** Molecular mechanism of a cellular circuit rectifier. **a** Two subpopulations of transcripts are created by occurrence of canonical distant PolyA signal and proximal non-canonical signal. **b** Transcription produces two types of mRNA that differ by a stretch of RNA that may contain functional sites such as microRNA target areas. Both transcripts share the same coding part. **c** When both transcripts have the same turnover rate, the transcript abundance has oscillating baseline. If more abundant transcript decays faster the peak abundance also shifts in time and can reach complete counter-phase (see Analytic Solution). In such case the sum of two transcripts approaches non-oscillatory steady line

### Pattern datamining

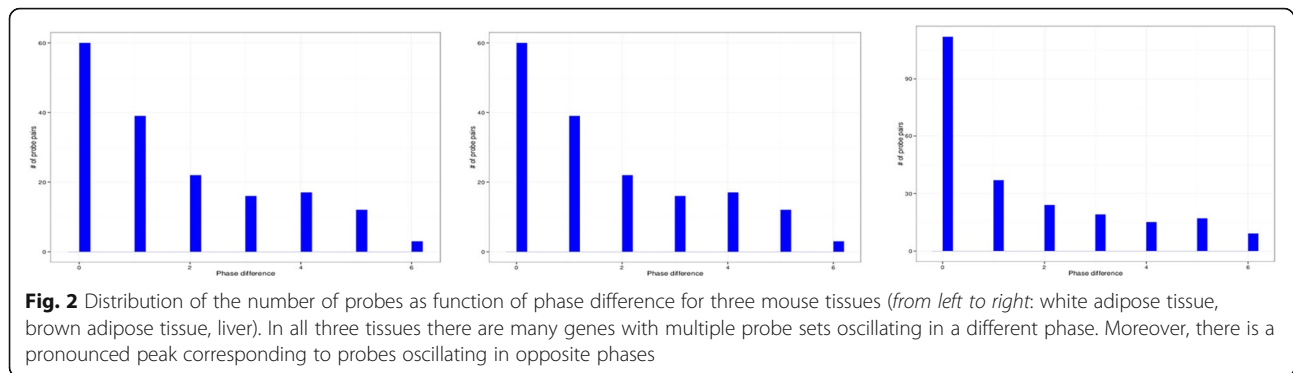
Frequently referencing the same or similar data sets from independent circadian studies we could not help but notice that the pattern of alternative probe sets for the same gene showing oscillations in a different phase is quite common in different plants and animals. Here we present the results of systematic search for expression patterns indicative of counter-phase transcripts.

We present a conservative estimation of the counter-phase transcript occurrence. The standard expression microarrays are poorly suited for observation of alternative transcripts. The higher representation of 3' UTR is usually viewed as unwanted bias that designers strive to avoid. Full length mRNAs from Refseq database are given priority over alternative shorter ESTs. Engineers also try to avoid excess probe sets interrogating the same gene in order to make quantitative estimation of gene expression more consistent. As a result we are only able to observe alternative polyadenylation through unintended imperfection on microarray design.

The phase estimation procedure described previously provides for each probe an estimation of the phase among one of six phase classes discretized by cyclic shift  $\pi i / 6$ , ( $i = 1..6$ ) and a corresponding  $p$ -value for each estimate. The  $p$ -value calculation is obtained from the bootstrap

analysis described in Algorithm 1. The latter can be used to filter probes with low statistical confidence on their phase estimate. We used the mouse annotation data (available from Affymetrix and on the shared github source code) to identify multiple probe sets interrogating expression of the same gene. All probes that correspond to the same gene symbol are gathered in a same probe set. The next step of the analysis is to generate phase differences within each probe set. All probe pairs in each probe set are used to compute the absolute value of phase difference.

Figure 2 shows the distribution of phase differences for three mouse tissues. We used a threshold  $p$ -value of 0.1 to filter probes with very low confidence on phase estimation. There is a peak around the zero phases for the different tissues. This result is expected since the probes are designed to provide a robust estimation of expression levels. Probe sets and separate probes within each set reporting results inconsistent with other probes and probe sets for the same gene tend to be eliminated from the chip on early design stages. As a result, the majority of alternative probe sets report abundance of the same transcript and shows no phase difference. There is a degree of uncertainty in identification of phase, considering the low sampling rate and high level of technical variation in microarray data. Thus, we expect high number of



alternative probesets with phase difference by one time point (second bar on the diagrams in Fig. 2). Likewise, there should be progressively fewer alternative probesets with larger phase difference. However, the diagrams show pronounced peaks corresponding to approximately opposite phases of oscillation.

We tested the significance of these visible bumps on diagrams in Fig. 2. Let's assume that in ideal microarray design all probe sets report consistent abundance of transcripts and there is no phase difference between alternative probe sets. In this case the first bar on Fig. 2 would be dominating, but we would still observe other bars due to technical variation and uncertainty in peak time estimations. However, if perceived phase difference was caused by stochastic variation only, we would observe progressively fewer cases with larger phase difference. It is expected that as the phase difference increases the count would decrease in an exponential manner. Therefore we can test the hypothesis that the observed distribution of phase difference has some stochastic decay. In order to test this hypothesis we fitted the phase difference distribution by a Poisson distribution. The advantage of using a Poisson distribution is that it can capture random variables that have stochastic decays. In addition since we have a discrete number of bins for phase difference, Poisson distribution is a good choice for discrete support. After fitting the distribution to the phase difference data, we applied a Chi-Square test to verify if this fitting is statistically valid. Table 1 summarizes the Poisson distribution fitting and

**Table 1** Phase difference distribution and stochastic decay hypothesis testing

Dataset	Lambda	X-squared	p-value
Brown adipose tissue	1.64	74.21	5.5e-14
White adipose tissue	1.5	46.45	2.4e-08
Liver	1.46	267.06	2.2e-16
Arabidopsis (UC Davis)	2.12	403.19	2.2e-16
Arabidopsis (Warwick)	2.07	272.88	2.2e-16

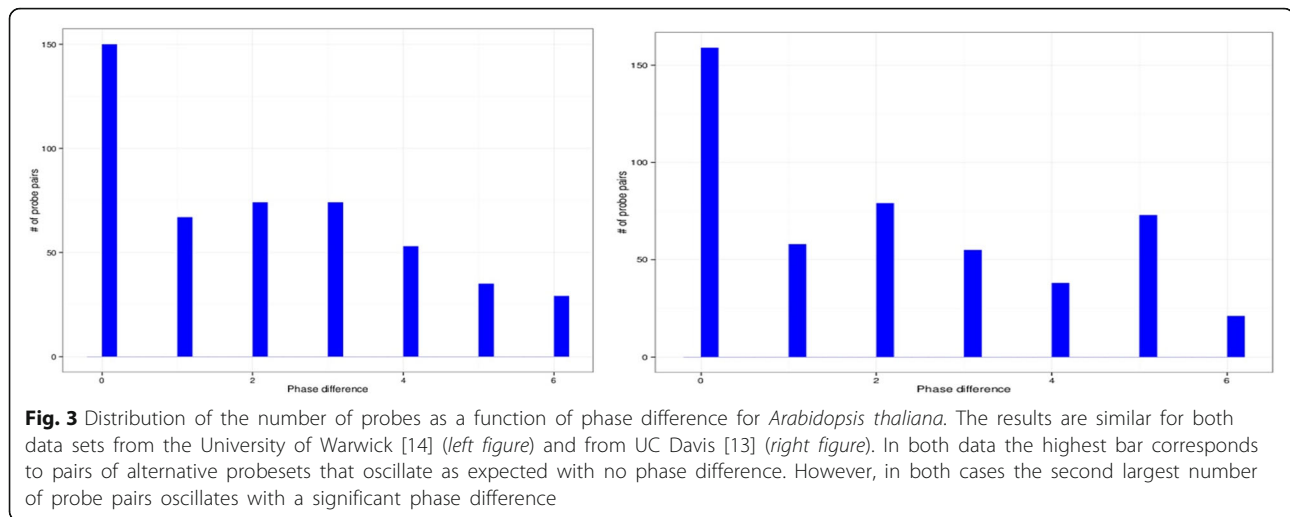
Lambda denotes the parameter of the Poisson distribution, X-square and p-value are the  $\chi^2$  hypothesis testing result. The Null hypothesis is that the phase difference distribution follows a Poisson distribution

the hypothesis testing results. We observe that, in the five datasets, the null hypothesis can be rejected and that the stochastic decay does not completely explain the phase difference distribution. We are aware that Poisson distribution does not capture all possible distributions with stochastic decay. We also performed non-parametric estimation of the phase difference distribution. The results are consistent with the tests in Table 1 (data not shown, see Additional file 1: Supplemental method). The distributions in Figs. 2 and 3 cannot be explained by stochastic variation and reflects a fraction of probes that oscillate consistently in opposite or near-opposite phase to each other. This observation is also true for all analyzed datasets (see complete list of probe sets with phase differences in Additional files 2 and 3).

We applied the same analysis to *Arabidopsis thaliana* circadian gene expression data. The data comes from the published sources (GEO GSE8365, GSE5612) and the primary analysis has been previously published [14, 15] and later re-analyzed and published again [16]. Additional challenges in identifying alternative probesets in *Arabidopsis thaliana* microarray come from the less extensive annotation (also available from Affymetrix). For this analysis we considered probesets representing the same gene if any of the following fields were identical: RefSeq Transcript ID, AGI and Entrez Gene. Figure 3 shows that overall the phase difference distribution is similar to the analysis based on mouse data with some differences in the distribution shape. The occurrence of alternative transcripts oscillating with pronounced phase difference in such distant organism leads to conclusion that the mechanism creating phenomena is likely to be common for all eukaryotic organisms.

## Discussion

Regardless of the source of oscillation, the rhythmic nature of expression demands a significant revision of the way we understand and model the function and regulation of genes. One of the previously published models predicted that for a rhythmically expressed gene addition of miRNA may have two different effects: either expected decrease or surprising increase of transcript abundance, depending on timing of miRNA action [17]. In case of



alternative polyadenylation, we have first noted, investigated and reported a strange abnormality in expression of alternative probe sets reporting activity of the same genes [13]. Disagreement in intensity among alternative probe sets is usually attributed to cross-hybridization, flaws in microarray design or manufacturing or other factors reflecting technical rather than biological variation. Indeed, experiments comparing only single points in time are insufficient to explain such discordance. Observation of complete circadian (or other periodic) time leaves no doubt that at least some of the alternative probe sets report biologically relevant rather than technical variation.

Our model shows that such strange behavior of alternative probes is not only natural; they are performing an important function. This function eliminates the effect of oscillation in transcription mechanism. Other studies have already reported pervasive oscillation of the entire transcriptome (see [18, 19, 20] for review). Current study presents the mechanism for rectification of constant baseline oscillation. If there is such mechanism than the default state of gene expression must be rhythmic. Our observations show that this mechanism is common in among plant, mouse and human genes. However, our study tends to underestimate the occurrence of alternative transcripts. On the chips used to produce this data thousands of genes are interrogated by a single probe set only. In cases when original CEL files are available for Affymetrix microarrays it is possible to analyze more genes with alternative transcripts by low level single probe analysis. However, individual probes may not be uniformly distributed to represent all transcripts and are less reliable in quantitative estimation of transcript abundance. The true occurrence of such mechanism is yet to be determined in a specially designed experimental study using a different detection mechanism.

Most of approaches in bioengineering and synthetic biology make no account for oscillation [21]. We believe a

significant advancement in Pathway Engineering will require better understanding of the principles on which complex biological systems are organized. One of the principles is oscillation in production of cellular components, signal transduction and energy metabolism. Ignoring the fact of oscillation is only possible on the early steps or when implementing most primitive constructions. This study offers one of the components for building artificial cellular systems or re-engineering the existing cells based on the knowledge of the rhythmic nature of gene expression. This component is a functional analog of a diode in electronic circuits and can rectify oscillations occurring in biological circuits due to the oscillatory nature of gene expression.

## Conclusion

The findings described in this paper may have practical applications in pathway engineering and synthetic biology. Our model provides the mechanism for re-engineering of existing biological pathways in a living cell or de novo design of cellular circuits. The model predicts that it is possible to find the parameters (such as miRNA site with certain affinity) regulating the ratio of alternatively adenylated transcripts. Manipulating such ration allows changing the amplitude of a particular gene expression or even complete elimination of oscillation. Constant abundance of a gene product can be used for production purpose to maximize the output of a peptide or an enzyme producing the product of interest. Alternatively this mechanism can be engineered to block unwanted pathways such as apoptosis or cell motivity, etc., or to keep certain pathways active at any time. Likewise, the same model can be used to create a blueprint for constructing artificial genes with certain properties. For example, the formula given in the description can be used to select parameters of the artificial gene (affinity of early and late PolyA sites, affinity of



microRNA binding site) in order to create the desirable amplitude of oscillating product abundance.

**Methods**

**Data sources**

The mouse gene expression profiles was obtained in the original study of circadian gene expression in adipose tissues [22]. The AKR/J mice acclimated to a 12 h light: 12 h dark cycle, were harvested in sets of 3–5 mice at 4 h intervals in duplicates over a 24 h period. Total RNA samples from inguinal (iWAT) white adipose tissue, brown adipose tissue (BAT), and liver have been assayed by Affymetrix U74 GeneChip microarrays.

**The plant (*Arabidopsis thaliana*) data sets**

We used two independent data sets similar in experiment design [14, 15]. Seedlings were entrained in 12-h white light (light source was cool white fluorescence tubes)/12-h dark cycles for 7 days before being released into free-running conditions of continuous white light at 22 °C. Starting at subjective dawn of day 8 [14] or day 9 [15], tissue was harvested every 4 h over the course of the next 44 h. Following standard protocols labelled cRNA targets were prepared from total RNA and hybridized to Affymetrix Arabidopsis expression GeneChips according to the manufacturer’s instructions.

**Analytic solution**

We integrate each equation separately with respect to time. The solution of the system is:

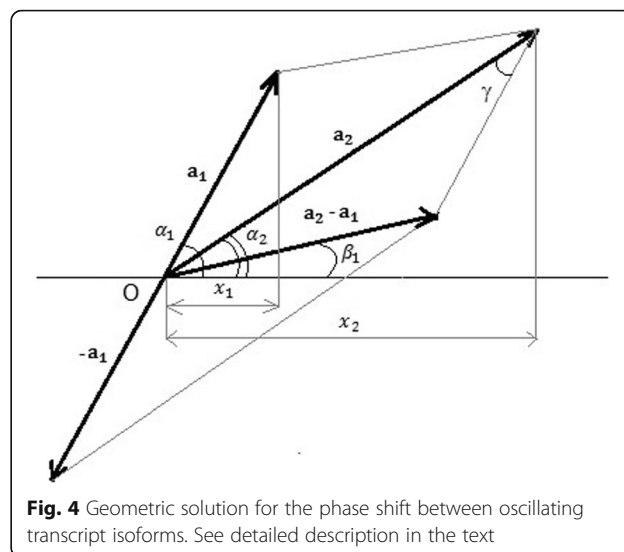
$$\begin{cases} n_1(t) = -\frac{pa}{\omega} \cos(\omega t + \alpha_1) + \frac{b}{\omega} \cos(\omega t + \alpha_2); \\ n_2(t) = -(1-p)\frac{a}{\omega} \cos(\omega t + \alpha_1) + \frac{c}{\omega} \cos(\omega t + \alpha_3); \end{cases} \tag{4}$$

The rotating-vector description of simple harmonic oscillation provides a neat way of rewriting  $n_1$  and  $n_2$  as single harmonic oscillations:

$$n_1(t) = A \cos(\omega t + \beta_1); \tag{5}$$

$$n_2(t) = B \cos(\omega t + \beta_2); \tag{6}$$

The following geometric solution is illustrated in Fig. 4. The harmonic oscillations  $x_1$  and  $x_2$  can be represented as two vectors  $a_1$  and  $a_2$  that rotate around their tails, which are pivoted at the origin  $O$ . The angular speed of the rotation is equal to  $\omega$ . As the vectors rotate around the origin their projections  $x_1$  and  $x_2$  on the horizontal axis vary sinusoidally. Hence we have



**Fig. 4** Geometric solution for the phase shift between oscillating transcript isoforms. See detailed description in the text

$$\gamma = \frac{2\pi - 2(\pi - \alpha_1 + \alpha_2)}{2} = \alpha_1 - \alpha_2; \tag{7}$$

$$A^2 = |a_1|^2 + |a_2|^2 - 2|a_1||a_2|\cos(\alpha_1 - \alpha_2); \tag{8}$$

$$A^2 = \left(\frac{pa}{\omega}\right)^2 + \left(\frac{b}{\omega}\right)^2 - \frac{2pab}{\omega^2} \cos(\alpha_1 - \alpha_2); \tag{9}$$

Therefore  $n_1(t) = A \cos(\omega t + \beta_1)$ , with

$$A = \sqrt{\left(\frac{pa}{\omega}\right)^2 + \left(\frac{b}{\omega}\right)^2 - \frac{2pab}{\omega^2} \cos(\alpha_1 - \alpha_2)}; \tag{10}$$

$$\begin{aligned} \tan\beta_1 &= \frac{|a_2| \sin\alpha_2 - |a_1| \sin\alpha_1}{|a_2| \cos\alpha_2 - |a_1| \cos\alpha_1} \\ &= \frac{b \sin\alpha_2 - pa \sin\alpha_1}{b \cos\alpha_2 - pa \cos\alpha_1}; \end{aligned} \tag{11}$$

Similarly, we can get the expression for  $n_2$ :

$$n_2(t) = B \cos(\omega t + \beta_2);$$

where.

$$B = \sqrt{(1-p)^2 \left(\frac{a}{\omega}\right)^2 + \left(\frac{c}{\omega}\right)^2 - \frac{2(1-p)ac}{\omega^2} \cos(\alpha_1 - \alpha_3)}; \tag{12}$$

$$\tan\beta_2 = \frac{c \sin\alpha_3 - (1-p)a \sin\alpha_1}{c \cos\alpha_3 - (1-p)a \cos\alpha_1}; \tag{13}$$

And the phase difference is.

## Algorithm 1.

The algorithm consists of the following steps:

1. Generate cosine waves of period of 24h, two cycles long (48h), with phase  $\varphi$ , intervals between adjacent phases is 4h. The equations are given by
 
$$C(t) = \cos\left(2\frac{\pi}{24}t - \varphi\right),$$
2. For  $b = 1, \dots, 999$ 
  - a- Using the maximum entropy algorithm, generate a replication of the gene expression time series;
  - b- Calculate the vector of correlations between the generated time series and each of the ideal cosine waves generated in step 1;
  - c- Estimate the first peak time of the generated gene;
3. Calculate a bootstrapped  $p$ -value and use it to test for the genes exhibiting a circadian rhythm;
4. Calculate a bootstrap percentile confidence interval and use it to assign a phase to each identified circadian gene.

$$\beta_2 - \beta_1 = \arctan\left(\frac{c \sin\alpha_3 - (1-p)a \sin\alpha_1}{c \cos\alpha_3 - (1-p)a \cos\alpha_1}\right) - \arctan\left(\frac{b \sin\alpha_2 - pa \sin\alpha_1}{b \cos\alpha_2 - pa \cos\alpha_1}\right). \quad (14)$$

**Phase assignment and phase confidence algorithm**

The algorithm used in the analysis of the data is based on resampling techniques. Indeed, we use the maximum entropy bootstrap algorithm to generate a large number of replications of a given gene expression time series. Then, we calculate a bootstrapped  $p$ -value to test for circadian genes, and finally we construct a bootstrap percentile confidence interval that will be used to assign a phase to each oscillating gene.

The complete description with source code and test results has been published in [23].

**Additional files**

**Additional file 1:** Supplemental method. In addition to Chi-Square test summarized in Table 1 we run a Monte Carlo simulation with 1000 repetitions with the option `simulate.p.value` as described in <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/chisq.test.html>. This supplemental file provides description and  $p$ -values obtained in simulations. (DOCX 10 kb)

**Additional file 2:** Supplemental Data Tables. This zip archive contains the results of analysis of phase difference among redundant probe sets in mouse liver, mouse brown fat, mouse white fat and two independent studies of *Arabidopsis thaliana* timeline gene expression. (ZIP 32 kb)

**Additional file 3:** Supplemental Initial Data Tables. This zip archive contains the initial timeline data for probe sets in mouse liver, mouse brown fat, mouse white fat originally published in Zvonicek et al. [21]. (ZIP 3710 kb)

**Abbreviations**

AGI: Arabidopsis Genome Initiative; BAT: Brown adipose tissue; cRNA: Antisense amplified RNA; GEO: Gene Expression Omnibus; iWAT: Inguinal white adipose tissue; JAK: Janus Kinase; miRNA: micro Ribonucleic Acid; mRNA: Messenger Ribonucleic Acid; SOCS3: Suppressor of cytokine signaling; STAT: Signal Transducer and Activator of Transcription; UTR: Untranslated region

**Acknowledgements**

We would like to thank Dr. Jeffrey Gimble for kindly providing the raw data from timeline microarray experiments.

**Funding**

Not applicable.

**Availability of data and materials**

All initial data is taken from previously published public sources. Intermediate data used in this study is included in supplemental materials. The murine circadian gene expression data is available by request from the corresponding author, GimbleJM@pbr.edu. The data from *A. thaliana* studies can be downloaded from Gene Expression Omnibus database (GEO) (<http://www.ncbi.nlm.nih.gov/geo>) using accession number GSE8365 and NASCArrays database (<http://bar.utoronto.ca/NASCArrays/index.php?ExplD=108>) under the accession NASCARRAYS-108 (circadian time course).

**Software availability.** All scripts and test data sets can be downloaded from the Github project directory <https://github.com/sidratools/>, following the link to /bsabri/gx\_phase\_shift. All software is provided free of charge on the open source basis.

**Authors' contributions**

NP designed the model and provided analytical solution; SB performed data mining and extraction of specific patterns of expression in multiple data; ME contributed the software and analysis of oscillation phase; AP designed the model and wrote the paper. All authors read and approved the final manuscript.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not Applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

<sup>1</sup>Embry-Riddle Aeronautical University, Daytona Beach, FL 32114, USA. <sup>2</sup>Sidra Medical and Research Center, P.O. box 26999 Doha, Qatar. <sup>3</sup>Present affiliation: Gloucester Marine Genomics Institute, Gloucester, MA 01930, USA.

Received: 21 March 2017 Accepted: 25 July 2017

Published online: 04 August 2017

**References**

- Panda S, Hogenesch JB, Kay SA. Circadian rhythms from flies to human. *Nature*. 2002;417(6886):329–35.
- Storch KF, Lipan O, Leykin I, Viswanathan N, Davis FC, Wong WH, et al. Extensive and divergent circadian gene expression in liver and heart. *Nature*. 2002;417(6884):78–83.
- Keegan KP, Pradhan S, Wang JP, Allada R. Meta-analysis of drosophila circadian microarray studies identifies a novel set of rhythmically expressed genes. *PLoS Comput Biol*. 2007;3(11):e208.
- Hogenesch JB, Panda S, Kay S, Takahashi JS. Circadian transcriptional output in the SCN and liver of the mouse. *Novartis Found Symp*. 2003;253:171–80. discussion 52–5, 02–9, 80–3 passim.
- Ptitsyn AA, Zvonic S, Gimble JM. Digital signal processing reveals circadian baseline oscillation in majority of mammalian genes. *PLoS Comput Biol*. 2007;3(6):e120.
- Zhang R, Lahens NF, Ballance HI, Hughes ME, Hogenesch JB. A circadian gene expression atlas in mammals: implications for biology and medicine. *Proc Natl Acad Sci U S A*. 2014;111(45):16219–24.
- Elkon R, Ugalde AP, Agami R. Alternative cleavage and polyadenylation: extent, regulation and function. *Nat Rev Genet*. 2013;14(7):496–506.
- Mayr C. Evolution and biological roles of alternative 3'UTRs. *Trends Cell Biol*. 2016;26(3):227–37. doi:10.1016/j.tcb.2015.10.012. Epub 2015 Nov 18. Review. PubMed PMID: 26597575; PubMed Central PMCID: PMC4955613.
- Perez-Santangelo S, Schlaen RG, Yanovsky MJ. Genomic analysis reveals novel connections between alternative splicing and circadian regulatory networks. *Brief Funct Genomics*. 2013;12(1):13–24.
- Petrillo E, Sanchez SE, Kornblihtt AR, Yanovsky MJ. Alternative splicing adds a new loop to the circadian clock. *Commun Integr Biol*. 2011;4(3):284–6.
- Kojima S, Sher-Chen EL, Green CB. Circadian control of mRNA polyadenylation dynamics regulates rhythmic protein expression. *Genes Dev*. 2012;26(24):2724–36.
- Kojima S, Green CB. Analysis of circadian regulation of poly(a)-tail length. *Methods Enzymol*. 2015;551:387–403.
- Ptitsyn AA, Gimble JM. Analysis of circadian pattern reveals tissue-specific alternative transcription in leptin signaling pathway. *BMC bioinformatics*. 2007;8(Suppl 7):S15.
- Covington MF, Harmer SL. The circadian clock regulates auxin signaling and responses in *Arabidopsis*. *PLoS Biol*. 2007;5(8):e222.
- Edwards KD, Anderson PE, Hall A, Salathia NS, Locke JC, Lynn JR, et al. FLOWERING LOCUS C mediates natural variation in the high-temperature response of the *Arabidopsis* circadian clock. *Plant Cell*. 2006;18(3):639–50.
- Ptitsyn A. Comprehensive analysis of circadian periodic pattern in plant transcriptome. *BMC bioinformatics*. 2008;9(Suppl 9):S18.
- Ptitsyn AP, Natalia; Elsebakhi, Emad; Marincola, Francesco; Al-Ali, Rashid; Temanni, M-Ramzi; AlSaad, Rawan; Wang, Ena Modulation of mRNA circadian transcription cycle by microRNAs. Middle East Conference on Biomedical Engineering (MECBME); 17–20 Feb. 2014; Doha, Qatar 2014.
- Ptitsyn AA, Gimble JM. True or false: all genes are rhythmic. *Ann Med*. 2011;43(1):1–12.
- Klevecz RR, Li CM, Marcus I, Frankel PH. Collective behavior in gene regulation: the cell is an oscillator, the cell cycle a developmental process. *FEBS J*. 2008;275(10):2372–84.
- Lloyd D, Eshantha L, Salgado J, Turner MP, Murray DB. Respiratory oscillations in yeast: clock-driven mitochondrial cycles of energization. *FEBS Lett*. 2002;519(1–3):41–4.
- Raman K, Chandra N. Flux balance analysis of biological systems: applications and challenges. *Brief Bioinform*. 2009;10(4):435–49.
- Zvonic S, Ptitsyn AA, Conrad SA, Scott LK, Floyd ZE, Kilroy G, et al. Characterization of peripheral circadian clocks in adipose tissues. *Diabetes*. 2006;55(4):962–70.
- El Anbari M, Fadda A, Ptitsyn A. Confidence in phase definition for periodicity in genes expression time series. *PLoS One*. 2015;10(7):e0131111.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
www.biomedcentral.com/submit





Reproduced with permission of copyright owner.  
Further reproduction prohibited without permission.