Gene 587 (2016) 163–168

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

# Post-transcriptional and translational regulation modulates gene co-expression behavior in more synchronized pace to carry out molecular function in the cell

Jinzeng Wang, Qi Lv, Xujuan Li, Ya Liu, Kang Liu, Haiyun Wang *

*School of Life Science and Technology, Tongji University, Shanghai 200092, China*

ARTICLE INFO

*Article history:*
Received 3 February 2016
Received in revised form 16 April 2016
Accepted 28 April 2016
Available online 2 May 2016

*Keywords:*
Gene co-expression
Transcriptome
Proteome
Post-transcriptional regulation

ABSTRACT

*Motivation:* Biological processes involve much complex interplay between cellular molecules at different molecular levels, and this interplay may exhibit various co-expression patterns explicitly representing the cellular inner regulation mechanism. Whereas, coexpression patterns cannot be necessarily conserved across the different molecular levels for complex regulation processes involved even after transcripts being produced. Investigation of co-expression propagation from transcript level to protein level will reflect inner regulation effects in function states of cells.

*Results:* In this study, we perform a comparative analysis of gene coexpression patterns in Plasmodium falciparum. We investigate coexpression patterns propagation from transcript level to protein level to reveal the underlying biological meaning of post-transcriptional and translational mechanism. Our systems-level approach shows after posttranscriptional and translational regulation gene co-expression pace at protein level is mechanistically adjusted to higher synchronicity. Moreover, co-expression patterns at protein level are more linked to function categories, such as co-expression at the same time point is more related with binding categories, and co-expression delayed by several time points is more related with activity categories. Therefore, posttranscriptional and translational regulation modulates co-expression relationships between molecules for meeting the function demands.



## 1. Introduction

Regulation of gene expression is a complex multilayered process from the transcription, post-transcription, translation to the post-translational modification of a protein. Any step of gene expression may be modulated. Gene expression values can be considered as important phenotypic consequences, and the direct or indirect regulation mechanisms across the different steps can be inferred by relationships of gene expression that may display various co-expression patterns. For instance, some gene encoding regulators within the same category are bound by the same transcriptional regulators, and these genes may act in simultaneous expression (Song et al., 2005). One gene may control or activate a downstream gene in a pathway and therefore their expression relationship may be time-shifted (Qian et al., 2001; Song et al., 2005). Different coordination patterns should be the explicit representation of cellular inner regulation mechanism (Wang et al., 2008).

Correlation between mRNA levels will not necessarily be conserved in the corresponding protein levels because of post-transcriptional and translational regulations. After transcripts being produced, the stability and distribution of the different transcripts as well as correlation between genes will be regulated by means of RNA binding protein that control the various steps and rates of the transcripts: events such as alternative splicing, nuclear degradation, processing, nuclear export, sequestration in *DCP2*-bodies for storage or degradation, and ultimately translation. Transcriptional control has received much attention (Kadonaga, 2004; Bertone et al., 2005; Birney et al., 2007; Kouzarides, 2007; Lieberman-Aiden et al., 2009; Schoenfelder et al., 2010). Presently post-transcriptional control has been mostly focused on the field of non-coding RNAs after they have been implicated in biological, developmental and pathological processes and act through mechanisms such as chromatin reprogramming, *cis* regulation at enhancers and post-transcriptional regulation of mRNA processing (Humphreys et al., 2005; Petersen et al., 2006; Mathonnet et al., 2007; Covarrubias and Reyes, 2010; Prensner and Chinnaiyan, 2011; Ulitsky and Bartel, 2013; Iyer et al., 2015; Wang et al., 2015). Translational control has been

*Abbreviation:* GO, Gene ontology; PAP, pace adjustment pattern; Pfa, plasmodium falciparum; ST-SP, simultaneous co-expression at transcript level and simultaneous co-expression at protein level; ST-DP, simultaneous co-expression at transcript level and time-delayed co-expression at protein level; DT-SP, time-delayed co-expression at transcript level and simultaneous co-expression at protein level; DT-DP, time-delayed co-expression at transcript level and time-delayed co-expression at protein level.

* Corresponding author.
*E-mail addresses:* wangjinzeng@gmail.com (J. Wang), iris0172@126.com (Q. Lv), libbyjuan69@gmail.com (X. Li), liuyaguai@163.com (Y. Liu), 360134299@qq.com (K. Liu), wanghaiyun@tongji.edu.cn (H. Wang).

http://dx.doi.org/10.1016/j.gene.2016.04.055
0378-1119/

speculated for years that the amount of translating mRNAs (mRNAs bound to ribosome-nascent chain complex, RNC-mRNA) may better reflect protein abundance (Greenbaum et al., 2003; Wang et al., 2013). The limits of non-coding RNAs and translating mRNAs information, respectively, only provide post-transcriptional or translational regulation in phase.

With the high-throughput transcriptome and proteome data available, to the extent it has become possible to study gene post-transcriptional and translational regulation mechanism in a systematic way at the genome-scale. Initial efforts at the comparative study of gene co-expression change in small scale have yielded some interesting hints (Wang et al., 2010). Whereas it is not clear yet about whether co-expression change shows significant over-representation of some biological function characteristic, and how co-expression propagation from transcript level to protein level reflect inner regulation effects in function states of cells. In this article, we employ the high-throughput mRNA and protein expression profiles for *Plasmodium falciparum* and perform a comparative analysis of gene co-expression patterns and their corresponding function association between transcript level and protein level. Then we assess the extent of co-expression divergence across two molecular levels and the connections between gene co-expression patterns and molecular function, to generally explore the biological significance of post-transcript and translation mechanism from the view of gene co-expression propagation across two molecular levels.

## 2. Algorithm and methods

### 2.1. Methods for unravelling simultaneous and time-delayed co-expression

The expression vectors of two genes $g_x$ and $g_y$, are denoted by $X=(x_1,x_2\cdots x_i\cdots x_N)$ and $Y=(y_1,y_2\cdots y_i\cdots y_N)$. We calculate the co-expression score $\gamma$ of $g_x$ and $g_y$ at the same time point. The basic definition of $\gamma$ follows from the Spearman rank correlation coefficient

$$\gamma = 1-\frac{6\sum d^2}{L\left(L^2-1\right)}$$

with $L$ denoting the dimension of $X$ and $Y$. Here, we rank both $X$ and $Y$ from the highest to the lowest values. Then, we subtract the two sets of ranks to obtain the difference $d$.

We apply a hypothesis test $T$ for the co-expression score based on a permutation approach using Monte Carlo techniques. Based on this procedure, we can test whether a calculated score $\gamma$ for two genes is a random sample from the background distribution of scores. The p value was further corrected with the Bonferroni method for controlling false discovery rate. The test procedure is as follows:

(1) Create reference expression vectors of $g_x$ and $g_y$ under $H_0$ by permuting experimental conditions of $X$ and $Y$.
(2) Calculate co-expression score $\gamma_0$ of permuted $X$ and $Y$.
(3) Repeat step the two previous steps 500 times.
(4) Create cumulative distribution of $\gamma_0$ (null distribution).
(5) Calculate $p(\gamma|H_0)$ after the Bonferroni correstion, if $p<0.05$, reject $H_0$.

Only gene pairs with significant co-expression scores are selected.

If there is no statistically simultaneous co-expression correlation between them, time-delayed co-expression score $\gamma$ is further calculated. The detailed algorithm is described in our another work (Wang et al., 2010).

### 2.2. Enrichment analysis of gene pairs with different co-expression pattern

Given an a priori defined set of gene pairs (e.g. gene pairs with co-expression at same time point or delayed several time point), the goal of this method is to determine whether the members of set are randomly distributed across Gene Ontology (GO) (Ashburner et al., 2000) terms.

This method compares the number of gene pairs that are annotated in the same GO term with the number of gene pairs that are randomly annotated in the given GO term just by chance. If the observed number is statistically greater than the one expected by chance, the GO term is reported as significant. A statistical model, binomial distribution is used to calculate the probability of observing the actual number of gene pairs just by chance, i.e., $p$ value. The $p$ value was further corrected with the Bonferroni method for controlling false discovery rate. Each GO term is calculated with this method.

### 2.3. Determining pace adjustment pattern in each category

The goal of this method is to check how co-expression pace is adjusted across two molecular levels in each GO term. There are three steps of this method:

Step 1: For each GO term, we calculate the number of gene pairs with specific co-expression pattern, including (i) co-expression at same time point, (ii) delayed by one time point, (iii) two time point and (iv) three time point, and (v) no relationship, respectively, at the transcript level and at the protein level.

Step 2: We compare the number of gene pairs with specific co-expression pattern at the protein level with the ones with the same co-expression pattern at the transcript level, and estimate the statistical significance ($p$ value) by using one-side Fisher test. Also, the $p$ value was further adjusted by the Bonferroni correction method.

Step 3: Considering five co-expression patterns in step 1, we develop a pace adjustment pattern (PAP) to evaluate the pace difference across two molecular levels for each GO term. If the number of gene pairs with a specific co-expression pattern at the protein level is statistically greater than the one with the same co-expression pattern at the transcript level, pace adjustment score is 1. Oppositely, pace adjustment score is −1. And if there is no difference across two molecular levels, pace adjustment score is 0. PAP for each GO term is a vector which consists of the pace adjustment scores for five co-expression patterns.

## 3. Results

### 3.1. Gene co-expression pace at the protein level is mechanistically adjusted to higher synchronicity

The data comprise mRNA transcript and protein sets of *Plasmodium falciparum* (Pfa) at six common stages (Mero, Ring, Troph, Schiz, Gameto and Sporo) (Le Roch et al., 2004). The transcriptome and proteome data of Pfa dataset are taken from a pioneering study from Le Roch KG et al. (Le Roch et al., 2004). In this study, the abundance of mRNA transcripts is calculated by applying the MOID algorithm for high-density oligonucleotide array analysis. The MOID algorithm provides a $p$ value for each measurement and thus a metric to evaluate the confidence of each data point. Transcripts are considered to be present if their expression levels are greater than 10 and the logrithm of the $p$ value is less than −0.5. Applying this methodology, 4292 transcripts are detected in at least one of the six stages examined. On the protein level, point were measured using the Redi Micro BCA protein assay system (Pierce), 2904 proteins are detected in at least