Gene 518 (2013) 7–16

Contents lists available at SciVerse ScienceDirect

Gene

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

Co-expression network with protein–protein interaction and transcription regulation in malaria parasite *Plasmodium falciparum*

Fu-Dong Yu [a,b,1], Shao-You Yang [b,2], Yuan-Yuan Li [b,*], Wei Hu [a,**]

[a] School of Life Sciences, Fudan University, 220 Handan Road, Shanghai 200433, PR China
[b] Shanghai Center for Bioinformation Technology, 1278 Keyuan Road, Shanghai 201203, PR China

ARTICLE INFO

Available online 26 December 2012

Keywords:
Gene co-expression network
Malaria parasite
Module analysis

ABSTRACT

Background: Malaria continues to be one of the most severe global infectious diseases, as a major threat to human health and economic development. Network-based biological analysis is a promising approach to uncover key genes and biological processes from a network viewpoint, which could not be recognized from individual gene-based signatures.
Results: We integrated gene co-expression profile with protein–protein interaction and transcriptional regulation information to construct a comprehensive gene co-expression network of *Plasmodium falciparum*. Based on this network, we identified 10 core modules by using ICE (Iterative Clique Enumeration) algorithm, which were essential for malaria parasite development in intraerythrocytic developmental cycle (IDC) stages. In each module, all genes were highly correlated probably due to co-regulation or formation of a protein complex. Some of these genes were recognized to be differentially coexpressed among three close-by IDC stages. The gene of *prpf8* (PFD0265w) encoding pre-mRNA processing splicing factor 8 product was identified as DCGs (differentially co-expressed genes) among IDC stages, although this gene function was seldom reported in previous researches. Integrating the species-specific gene prediction and differential co-expression gene detection, we found some modules could perform species-specific functions according to some of genes in these modules were species-specific genes, like the module 10.
Furthermore, in order to reveal the underlying mechanisms of the erythrocyte invasion by *P. falciparum*, Steiner Tree algorithm was employed to identify the invasion subnetwork from our gene co-expression network. The subnetwork-based analysis indicated that some important *Plasmodium* parasite specific genes could corporate with each other and be co-regulated during the parasite invasion process, which including a head-to-head gene pair of PfRH2a (PF13_0198) and PfRH2b (MAL13P1.176).
Conclusions: This study based on gene co-expression network could shed new insights on the mechanisms of pathogenesis, even virulence and *P. falciparum* development.

1. Introduction

Malaria remains to be one of the most severe global infectious diseases, infecting 300–500 million people and causing millions of deaths every year. *Plasmodium falciparum*, the most devastating human malaria parasite, is responsible for over 90% of human deaths from malaria. However, no effective anti-malaria vaccines are available for clinical use in humans (Crompton et al., 2010). During the past several decades, the control of malaria still relies heavily on chemotherapy, which causing drug resistance in malaria parasites increased the morbidity and mortality rates in malaria endemic regions. It's urgently needed to develop new drug and vaccine. On the other hand, some mechanism researches about parasitism, pathogenesis and even virulence could improve and accelerate the development of drug and vaccine.

After the genome sequence of *P. falciparum* became available (Gardner et al., 2002), the gene expression profiles of its intraerythrocytic developmental cycle (IDC) (Bozdech et al., 2003; Llinas et al., 2006) and other asexual and bloodborne stages (Le Roch et al., 2003) have been illustrated by microarray. Several researches have discovered the possible regulation relationships between transcriptional factors of Apicomplexan-specific AP2 (ApiAP2) family and their binding motifs (Balaji et al., 2005; De Silva et al., 2008). ApiAP2 family putative transcriptional regulators were proved to play key roles in development and environmental stress response pathways in *Plasmodium* species (De Silva et al., 2008; Jurgelenaite et al., 2009; Sims et al., 2009).

Abbreviations: IDC, intraerythrocytic developmental cycle; DCGs, differentially co-expressed genes; MCL, Markov cluster algorithm; ICE, Iterative Clique Enumeration; PCC, Pearson Correlation Coefficient; PPI, protein–protein interaction; GO, Gene Ontology; ISSPG, Identification of Species-Specific eukaryotic Parasite Genes; TF, transcription factor.

* Corresponding author. Tel.: +86 21 20283720; fax: +86 21 20283780.
** Corresponding author. Tel.: +86 21 65643031; fax: +86 21 54653512.
E-mail addresses: fudongyu@gmail.com (F.-D. Yu), shyyang@scbit.org (S.-Y. Yang), yyli@scbit.org (Y.-Y. Li), huw@fudan.edu.cn (W. Hu).
[1] Tel.: +86 21 20283722; fax: +86 21 20283780.
[2] Tel.: +86 21 20283759; fax: +86 21 20283780.

0378-1119/$ – see front matter.
http://dx.doi.org/10.1016/j.gene.2012.11.092

Recently, experimental map of physical interactions between proteins of *P. falciparum* were released by yeast two-hybrid screen (LaCount et al., 2005). Some studies based on protein–protein interaction (PPI) network showed the insights of core interactome (Wuchty et al., 2009) and protease-associated cellular subnetworks (Lilburn et al., 2011) in *P. falciparum*. Although significant progresses have been made, there are still much more works needed to learn and decipher the functions of these genes or proteins, including how they interact with other proteins or regulate other genes in differential developmental stages to coordinate important biological processes or pathways related to growth, invasion, pathogenesis, response to drug treatment and resistance in malaria parasites.

One of the most popular methods to uncover these potential functions in biological systems depends on comprehensive network biology study. As usually knowledge-based information is used to guide to build a network of proteins or genes. In this study we incorporated the protein–protein interaction data from STRING database (Szklarczyk et al.) and the yeast 2-hybrid (Y2H) screened physical protein interaction data (Aurrecoechea et al., 2009; LaCount et al., 2005), ApiAP2 transcriptional regulators driven regulation data (De Silva et al., 2008), IDC stages gene expression data (Llinas et al., 2006), and literature text mining information into building a comprehensive gene co-expression network with protein–protein interaction or regulation of *P. falciparum* (see the workflow of this study, Fig. 1). In this network, each node represented a gene or protein and each edge represented co-expression relation between each two gene-pair if co-expression correlation value was above the threshold. One of clique-percolation algorithms named ICE (Iterative Clique Enumeration) algorithm (Shi et al., 2010) was employed to identify functional modules (or named subnetworks), which could represent major transcriptional programs in co-expression network or protein complex, based on the topology of our constructed comprehensive gene co-expression network. To address each module function, a hypergeometric distribution enrichment analysis was used to analyze the function of modules identified by ICE algorithms using Gene Ontology (GO) annotation and KEGG pathway information. Thus we obtained the mostly significant enriched function of each module using this functional enrichment analysis although each module covered diversity heterogeneity functions with significant statistics. Therefore, genes highly co-expressed in each module could work together to perform potential homogeneity function, no matter the genes in one module with poorly understood or putative annotation. Besides, this functional module analysis is also one of the popular approaches used to annotate those genes with poorly functional annotation, and even postulate previously unrecognized members of some subnetworks.

As we all know, more than half of the genes in *P. falciparum* are predicted with poor annotation, and almost all of them are species-specific genes or genus-specific genes. And these genes are attracted for drug development. Considering the significance of species-specific genes for drug development, we detected the potential species-specific genes in function modules in this study. For each module, all genes implemented functional biological process or regulation complex together with highly co-expression. Some of them co-expressed

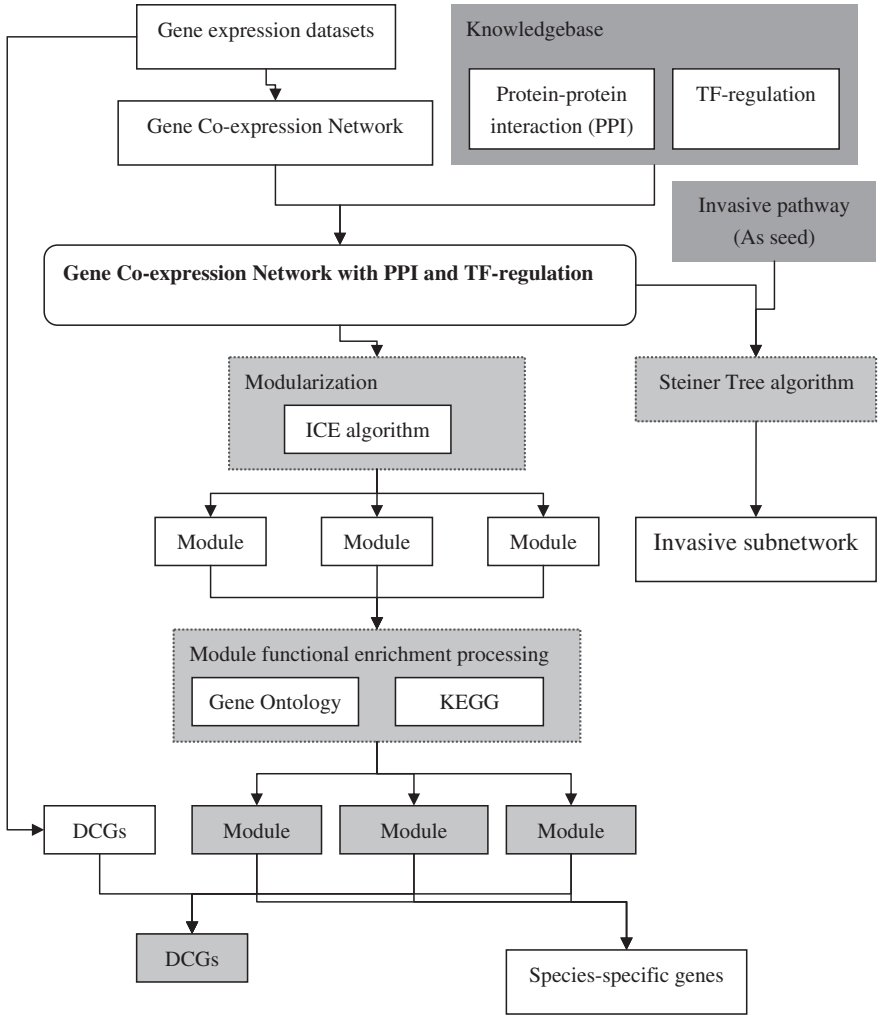

Fig. 1. The workflow of comprehensive gene co-expression network analysis in *Plasmodium falciparum*.