

June 2015

Predicting and exploring network components involved in pathogenesis in the malaria parasite via novel subnetwork alignments

Hong Cai

University of Texas at San Antonio

Timothy G. Lilburn

Novozymes

Changjin Hong

University of Minnesota Twin Cities

Jiaying Gu

CUNY College of Staten Island

Rui Kuang

University of Minnesota Twin Cities

See next page for additional authors

[How does access to this work benefit you? Let us know!](#)

Follow this and additional works at: http://academicworks.cuny.edu/si_pubs

Recommended Citation

Cai, H., Lilburn, T. G., Hong, C., Gu, J., Kuang, R. & Wang, Y. (2015). Predicting and exploring network components involved in pathogenesis in the malaria parasite via novel subnetwork alignments. *BMC Systems Biology*, 9(Suppl 4), S1-S1. doi:10.1186/1752-0509-9-S4-S1.

This Article is brought to you for free and open access by the College of Staten Island at CUNY Academic Works. It has been accepted for inclusion in Publications and Research by an authorized administrator of CUNY Academic Works. For more information, please contact AcademicWorks@cuny.edu.

Authors

Hong Cai, Timothy G. Lilburn, Changjin Hong, Jianying Gu, Rui Kuang, and Yufeng Wang

RESEARCH

Open Access

Predicting and exploring network components involved in pathogenesis in the malaria parasite via novel subnetwork alignments

Hong Cai^{1†}, Timothy G Lilburn^{2†}, Changjin Hong^{3†}, Jianying Gu⁴, Rui Kuang^{3*}, Yufeng Wang^{1,5*}

From The International Conference on Intelligent Biology and Medicine (ICIBM) 2014
San Antonio, TX, USA. 04-06 December 2014

Abstract

Background: Malaria is a major health threat, affecting over 40% of the world's population. The latest report released by the World Health Organization estimated about 207 million cases of malaria infection, and about 627,000 deaths in 2012 alone. During the past decade, new therapeutic targets have been identified and are at various stages of characterization, thanks to the emerging omics-based technologies. However, the mechanism of malaria pathogenesis remains largely unknown. In this paper, we employ a novel neighborhood subnetwork alignment approach to identify network components that are potentially involved in pathogenesis.

Results: Our module-based subnetwork alignment approach identified 24 functional homologs of pathogenesis-related proteins in the malaria parasite *P. falciparum*, using the protein-protein interaction networks in *Escherichia coli* as references. Eighteen out of these 24 proteins are associated with 418 other proteins that are related to DNA replication, transcriptional regulation, translation, signaling, metabolism, cell cycle regulation, as well as cytoadherence and entry to the host.

Conclusions: The subnetwork alignments and subsequent protein-protein association network mining predicted a group of malarial proteins that may be involved in parasite development and parasite-host interaction, opening a new systems-level view of parasite pathogenesis and virulence.

Background

Malaria is a severe global infectious disease. The World Health Organization estimated 207 million clinical cases, and 627,000 deaths due to malaria in 2012. Malaria infection was reported in 97 countries in 2013, representing a major public health concern. The vast majority of the deaths occurred in children under age 5 in sub-Saharan areas; it is estimated that a child dies from malaria every minute.

The causative agents of malaria are a group of parasites from the Genus *Plasmodium*. Among five human malaria

parasites, *Plasmodium falciparum* causes the most severe form of malaria. It is notoriously difficult to study malaria biology as the parasite has a complex life cycle involving a mosquito vector and a human host. Within a human host, the parasite has a relatively dormant liver stage and an active red blood cell (RBC) stage. Clinical symptoms such as high fever, chills, headache, profuse sweats, fatigue, nausea, and vomiting are manifested at the blood stage, the stage that most drugs target against.

The effectiveness of antimalarial drugs, however, has been constantly challenged during the past decades due to the fast evolution of parasites that are resistant to multiple lines of drugs. This underscores an urgent need to search for novel drug targets. New lines of antimalarial targets have been identified [1-5] and are at various stages of functional and pharmacological characterization, thanks to the development of customized omic-based high

* Correspondence: kuang@cs.umn.edu; yufeng.wang@utsa.edu

† Contributed equally

¹Department of Biology, University of Texas at San Antonio, San Antonio, TX 78249, USA

³Department of Computer Science and Engineering, University of Minnesota Twin Cities, Minneapolis, MN 55455, USA

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

throughput technologies, including genome sequencing and annotation [6-11], microarray [12-23], RNA-Seq [21,22,24], mass spectrometry [4,5,25-32], yeast two-hybrid protein-protein interaction assays [33-36], and large-scale compound screening [37]. Moreover, these amassing genomic, transcriptomic, proteomic, metabolomics [38-42], and interactomic data are enabling a new systems biology paradigm in malaria research; the complex and dynamic cellular processes such as pathogenesis and virulence have recently begun to be unveiled.

By nature, pathogenicity is an emergent property, that is, it is the result of the concerted effects of many genes. The suggestion by Mu et al. that differences in drug resistance in *P. falciparum* parasites from different continents were due to “overlapping, but not identical, sets of genes, including many encoding unknown proteins” [43] and the observation that the costs of maintaining antibiotic resistance in bacteria can be offset or lost thanks to changes at other loci not directly involved in antibiotic resistance [44] are both strongly suggestive of the reality of systems-level adaptation and of the existence of systems producing the virulence phenotype [45]. However, unlike bacterial systems, whose pathogenesis mechanism is much better characterized, to date, malaria pathogenesis, i.e., the development of the disease condition, remains largely unknown. A critical barrier is our limited knowledge of the cellular components, cellular events and reactions and other pathologic mechanisms associated with the disease development.

To unveil previously unknown proteins that are potentially involved in pathogenesis in *P. falciparum*, we developed a module-based subnetwork alignment approach. Traditional homology transfer using a protein-protein network alignment approach [46-54] does not appear feasible to predict functional orthologs in the malaria parasite, because of the remote homology between *P. falciparum* and other known model organisms: BLAST-based sequence similarity search was unable to make functional annotation for over 60% of the genes in *P. falciparum* [11]. Our neighborhood subnetwork alignment algorithm [55] could potentially circumvent this limitation by searching for the similarities between functional modules. This algorithm, which was custom-developed for malaria research, was used to predict novel transcriptional regulators and cell cycle regulatory proteins in *P. falciparum* [55,56]. In this paper, we extended the subnetwork alignment approach to identify the proteins related to pathogenesis and explore their potential functional roles.

Results and discussion

Twenty-four (24) novel proteins were predicted to be associated with pathogenesis in malaria parasite

Malaria pathogenesis is an emergent property that involves dynamic regulation of parasite metabolism,

parasite signaling, and host-parasite interactions [57]. Gene Ontology using Term GO:0009405 (pathogenesis) predicted 95 *P. falciparum* genes (Additional File 1), not surprisingly, including 78 members of the *Plasmodium falciparum* erythrocyte membrane protein (PfEMP1) family. PfEMP1 is encoded by the highly polymorphic *Var* gene, and is one of the largest and most widely-studied protein families in *P. falciparum*. PfEMP1 is believed to play an important role in parasite pathogenesis and immune evasion [58]; it contributes to antigenic variation and cytoadherence, and the expression of specific PfEMP1 variants is associated with severe malaria in children [59].

Given the observed complexity of malaria pathogenesis, one might envisage that more network components have yet to be identified. Using a neighborhood subnetwork alignment algorithm, we predicted that 24 novel proteins in *P. falciparum* were functional orthologs of proteins known to be involved in pathogenesis in *E. coli* (Additional File 2). *E. coli* was chosen because its pathogenesis and virulence mechanisms are among the best-characterized of all known infectious agents. Seven of these predicted functional orthologs were annotated as “conserved *Plasmodium* proteins with unknown function”, six of which do not have any predicted interactors in malarial protein-protein association networks (Table 1). Seventeen of these 24 predicted functional orthologs may be involved in various biological processes from post-transcriptional regulation, protein translation, protein ubiquitination and modification, oxidation-reduction process, signaling, metabolism, to histone modification. Notably, one functional ortholog (PF3D7_0300100) is a variant of PfEMP1, and it is also consistent with the GO prediction that it may be associated with pathogenesis (Additional File 1).

The set of functional orthologs is involved in important biological processes that are related to pathogenesis and virulence

We performed network mining on the 24 predicted pathogenesis-related functional orthologs. Eighteen proteins are associated with 418 other proteins with high confidence score (>0.7). The network centralization parameter, a measure of the distribution of network density [60], is 0.0548, indicating that the network is decentralized. The degree of connectivity of these 18 proteins ranges from 180 to one and the degree distribution offers a good fit to a power law distribution (Table 1). This type of distribution indicates that most nodes are connected to a few others and a few nodes are connected to many others (Figure 1). The network heterogeneity coefficient assesses the relative number of hubs in a network; in this network it is relatively high (1.455), which is in accord with the observed power law distribution.

Table 2 shows representative functional categories of the proteins that are associated with predicted

Table 1 Novel *Plasmodium falciparum* proteins that were predicted to be functional orthologs of pathogenesis-related proteins in *E.coli*

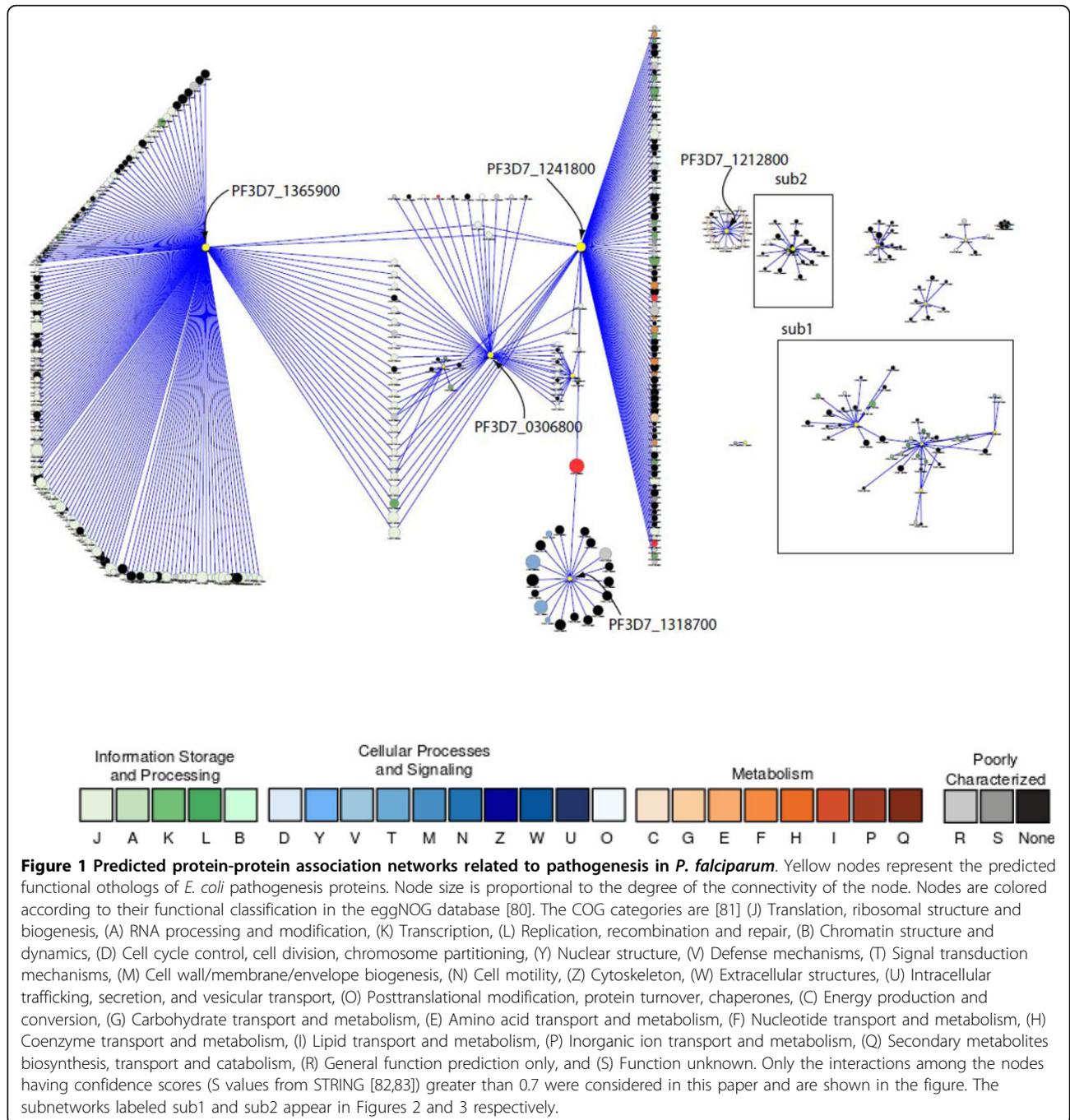
PlasmoDB Accession Number	Annotation	Degree of Connectivity
PF3D7_1365900	60S ribosomal protein L40/UBI, putative	180
PF3D7_1241800	DEAD/DEAH box ATP-dependent RNA helicase, putative	98
PF3D7_0306800	T-complex protein beta subunit, putative	43
PF3D7_0823300	histone acetyltransferase GCN5 (GCN5)	20
PF3D7_1318700	conserved Plasmodium protein, unknown function	19
PF3D7_1212800	iron-sulfur subunit of succinate dehydrogenase	18
PF3D7_0208600	ribosome-recycling factor, putative (RRF)	16
PF3D7_1416900	prefoldin subunit 2, putative	15
PF3D7_0731100	Plasmodium exported protein (PHISTc), unknown function (GEXP11)	14
PF3D7_1013500	phosphoinositide-specific phospholipase C (PI-PLC)	11
PF3D7_0810600	RNA helicase, putative	9
PF3D7_1337200	1-deoxy-D-xylulose 5-phosphate synthase	7
PF3D7_1148800	Plasmodium exported protein (hyp11), unknown function	6
PF3D7_1008000	inositol polyphosphate kinase, putative (IPK1)	6
PF3D7_0919000	nucleosome assembly protein (NAPS)	6
PF3D7_1418000	ubiquitin fusion degradation protein UFD1, putative (UFD1)	5
PF3D7_0300100	erythrocyte membrane protein 1, PfEMP1 (VAR)	5
PF3D7_1342400	casein kinase II beta chain (CK2beta2)	1
PF3D7_1318700	conserved Plasmodium protein, unknown function	0
PF3D7_1364200	conserved Plasmodium protein, unknown function	0
PF3D7_1229300	conserved Plasmodium protein, unknown function	0
PF3D7_1451200	conserved Plasmodium protein, unknown function	0
PF3D7_0310900	conserved Plasmodium protein, unknown function	0
PF3D7_1008100	conserved Plasmodium protein, unknown function	0

pathogenesis-related *P. falciparum* proteins. They are active players in crucial cellular processes in parasite development, growth, and response to host and environmental stresses. In the following sections, we discuss critical network components and their potential associations with pathogenesis and virulence.

1. Tight associations between translation, protein turnover, proteolysis, signaling, and stress response

The predicted functional ortholog with the highest connectivity (180) is a putative 60S ribosomal protein L40/UBI (PF3D7_1365900) (Figure 1). This protein is unique as it is a fusion protein of two domains: a C-terminal ribosomal L40e family domain and an N-terminal ubiquitin domain. It therefore plays dual functions: On one hand, it is a central member of the ribosomal protein complex in *P. falciparum*, which includes small and large subunit ribosomal proteins of varying sizes. This complex is associated with several translation initiation and elongation factors. On the other hand, L40/UBI is implicated in the ubiquitin-proteasome system (UPS) [61]. UPS is an essential mechanism in *P. falciparum* for degrading misfolded or unneeded proteins [62]. UPS is becoming a promising target for antimalarial development due to its critical roles

in cell cycle regulation and stress response and the relative low toxicity of its inhibitors [63]. Notably, malarial L40/UBI may be involved in other processes that are currently not fully appreciated, such as (1) signaling. L40/UBI and PfRACK (*P. falciparum* receptor for activated C kinase homolog, PF3D7_0826700) are associated with a high confidence score (0.908), suggested by the co-expression of their homologs in *Caenorhabditis elegans*, *Drosophila melanogaster*, and humans, and protein-protein interaction found by affinity capture-MS assay in humans. The evidence that PfRACK could directly inhibit inositol 1,4,5-trisphosphate receptor-mediated calcium-signaling in mammalian host cells is indicative of a pathogenesis mechanism acting through the disruption of host activities [64]. (2) Genome stability. L40/UBI is associated with two putative proliferating cell nuclear antigens (PCNAs), auxiliary factors of DNA polymerase. Ubiquitin, along with SUMO (small ubiquitin-related modifier), were shown to effectively regulate DNA damage recognition and repair in yeast [65]. A similar mechanism, if it exists in the malaria parasite, would promote the accuracy of replication and timely repair, both of which are crucial for parasite survival within the host.



2. RNA metabolism and transcriptional regulation

The predicted pathogenesis-related functional orthologs with the second largest connectivity (98) is a putative DEAD/DEAH box ATP-dependent RNA helicase (PF3D7_1241800) (Figure 1). The main function of this family of RNA helicases is to unwind RNA [66]. Helicase is considered to be a potential antimalarial target due to its essentiality for parasite life cycle [67]. Fifty-one DEAD/DEAH helicase homologs were identified in *P. falciparum* by genomic analysis [68]. Our network mining indicated

that PF3D7_1241800 is broadly involved in RNA metabolism; it is associated with (1) several RNA polymerase subunits and a number of other helicases required for transcription, (2) nucleolar GTP-binding proteins, putative GTPases, and BRIX-domain containing proteins for ribosomal biogenesis and maturation, (3) at least three putative U3 snoRNP associated proteins for pre-rRNA processing, (4) a ribosomal subunit export protein for RNA transport, and (5) an eukaryotic initiation factor required for translation. In addition to RNA metabolism,

Table 2 Representative *P.falciparum* proteins that were predicted to be associated with pathogenesis.

Functional category	PlasmoDB Accession Number	Annotation
DNA replication	PF3D7_0111300	replication factor c protein, putative
Transcription and transcriptional regulation	PF3D7_0923000	DNA-directed RNA polymerase II, putative
	PF3D7_1206600	DNA-directed RNA polymerase III subunit, putative
	PF3D7_0206600	transcription factor, putative
	PF3D7_1451400	transcriptional regulatory protein sir2b (Sir2B)
	PF3D7_1342700	transcription activator, putative
Translation	PF3D7_1460700	60S ribosomal protein L27, putative
	PF3D7_1421200	40S ribosomal protein S25, putative
	PF3D7_1451100	elongation factor 2
	PF3D7_0422700	eukaryotic translation initiation factor 4F complex
Protein phosphorylation and signaling	PF3D7_0826700	receptor for activated C kinase homolog, PfrACK
	PF3D7_0424500	serine/threonine protein kinase, FIKK family (FIKK4.1)
	PF3D7_0623800	protein kinase, putative (TKL4)
Regulation of cell cycle	PF3D7_1428300	proliferation-associated protein 2g4, putative
	PF3D7_0619400	cell division cycle protein 48 homologue, putative
Proteolysis	PF3D7_0933600	mitochondrial-processing peptidase subunit beta, putative (MAS1)
	PF3D7_0904400	signal peptidase complex subunit 3, putative (SPC3)
	PF3D7_1331300	signal peptidase 21 kDa subunit (SP21)
	PF3D7_1130400	26S protease subunit regulatory subunit 6a, putative
Heat shock response	PF3D7_0708800	heat shock protein 110 (HSP110c)
	PF3D7_1357800	TCP-1/cpn60 chaperonin family, putative
	PF3D7_1215300	10 kd chaperonin, putative
	PF3D7_0308200	TCP-1/cpn60 chaperonin family, putative
	PF3D7_0214000	TCP-1/cpn60 chaperonin family, putative

PF3D7_1241800 is associated with a putative transcription factor (PF3D7_0206600). This transcription factor is a member of the general transcription factor TFIIS family with a characteristic zinc ribbon conformation. The accurate information flow of basal transcriptional regulation and RNA metabolism is essential for parasite development.

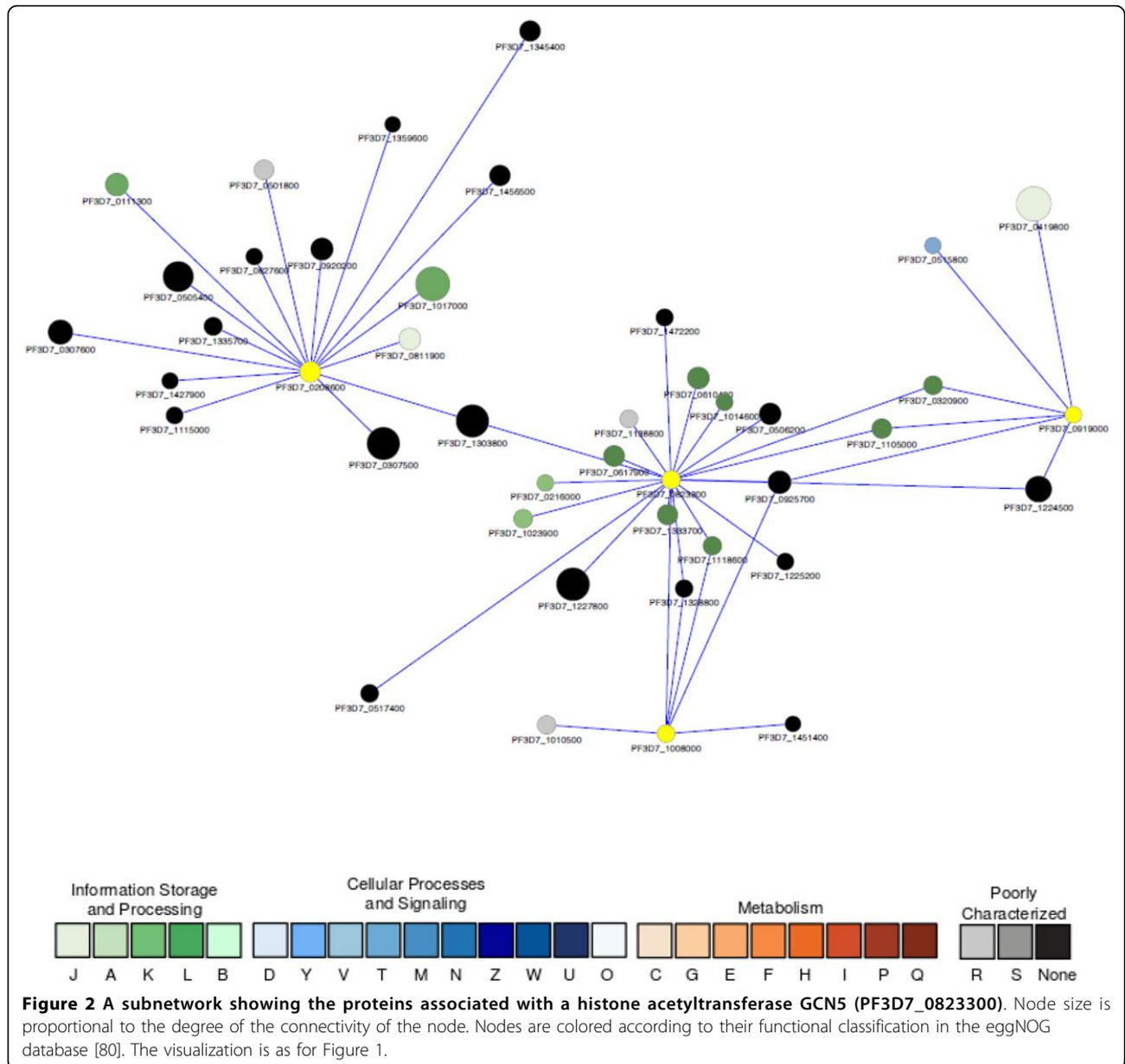
3. Chromatin remodeling, epigenetic regulation, and antigenic variation

It is becoming clear that chromatin remodeling/histone modification controls genome-wide gene expression in *P. falciparum* and epigenetic regulation may be a neglected major contributing factor to antigenic variation and evasion of the human immune system during parasite pathogenesis [69-71]. Our network analysis identified a subnetwork module centered on PfGCN5 (PF3D7_0823300), a bona fide histone acetyltransferase. PfGCN5 is associated with 20 other proteins (Figure 2), including histones H2A, H3, and H4, DNA helicase, three histone deacetylases, a putative chromodomain-helicase-DNA-binding protein, and a putative chromatin assembly protein ASF1, suggesting its role in chromatin remodeling. In addition, PfGCN5 may be involved in transcriptional regulation: it is associated with

the transcriptional activator ADA2 with a confidence score of 0.955 [72]; in yeast, GCN5, ADA2, and ADA3 interact and form a catalytic core for histone acetyltransferase (HAT) activity. PfGCN5 may also have protein-protein interactions with a general transcription initiation factor TFIID and a putative transcriptional regulator. Interestingly, PfGCN5 is predicted to be associated with a transcriptional regulatory protein Sir2a (PF3D7_1328800) (Figure 2). Sir2a and its paralog Sir2b are implicated in epigenetic regulation of PfEMP1 antigenic variation, thus controlling the pathogenicity of malaria [73,74].

PfGCN5, is also indirectly associated with three predicted pathogenesis-related functional orthologs (labelled in yellow in Figure 2): (1) an NAPS protein (PF3D7_0919000) involved in nucleosome assembly, (2) a putative ribosome-recycling factor (RRE, PF3D7_0208600) that is essential for recycling of organellar ribosomes during protein translation, and (3) a putative inositol polyphosphate kinase (IPK1, PF3D7_1008000) that may be involved in signaling, chromatin remodeling, and mRNA export.

Unlike the PfGCN5-centered chromatin-remodeling module, very little is known about another subnetwork centered on two *Plasmodium* exported proteins hyp11



(PF3D7_1148800) and GEXP11 (PF3D7_0731100) (Figure 3). The function of these two proteins is unclear, and the majority of their interactors are hypothetical proteins as well; the evidence of association is mainly established based on co-expression profiles and does not necessarily suggest direct functional association.

Conclusions

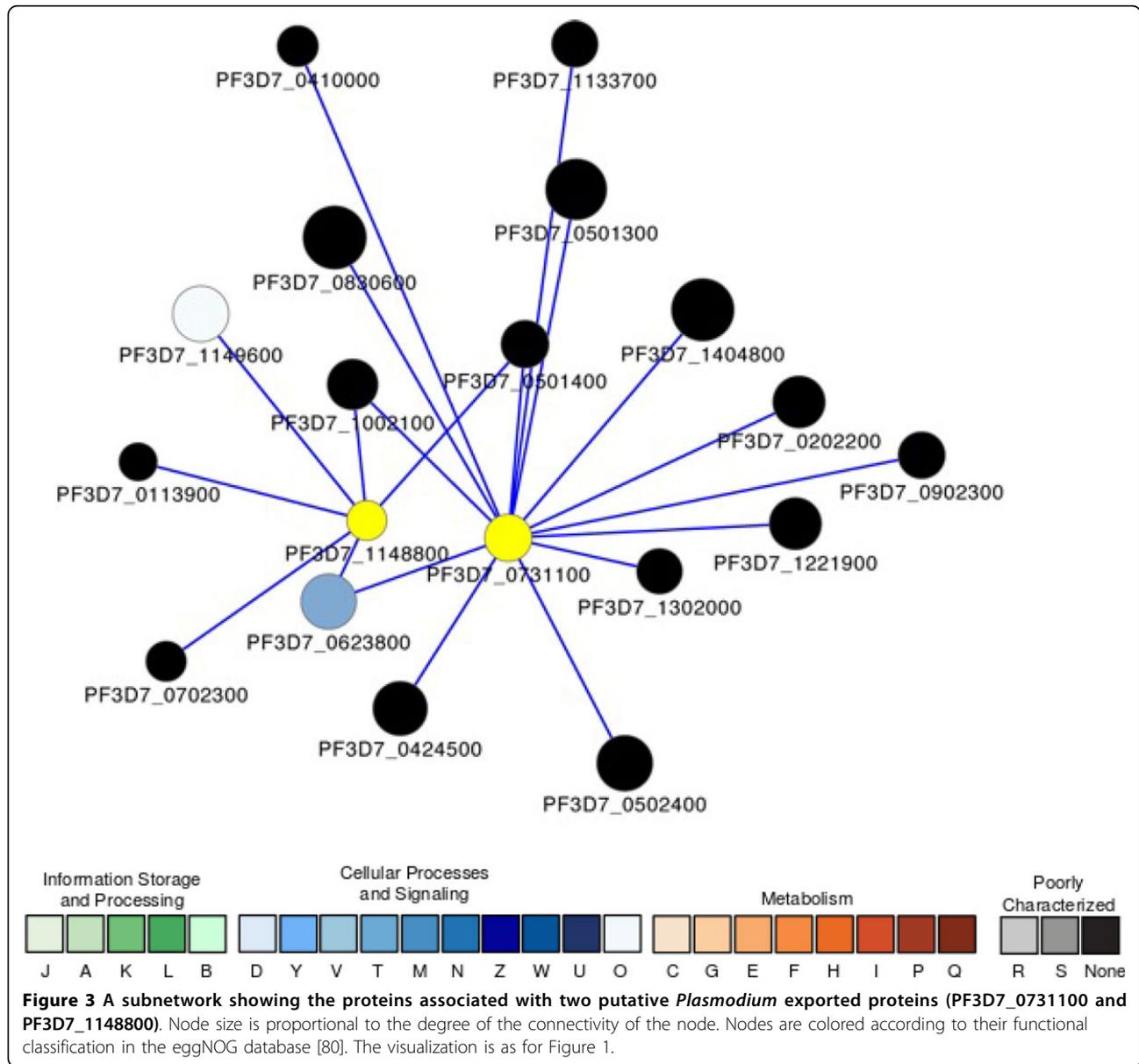
Using a neighborhood subnetwork alignment algorithm, we identified network components associated with 24 potential pathogenesis-related proteins that may be involved in malaria pathogenesis. The proteins play roles in parasite information processing, signaling, epigenetic regulation, and entry to the host, all processes that

are related to pathogenesis and virulence. A better understanding of the network components containing these proteins creates a new list of potential rational targets for antimalarial intervention.

Methods

Querying subnetworks using neighborhood alignments

We have framed the prediction of functional orthologs in *P. falciparum* as a subnetwork querying problem. Network alignment algorithms search the network of the organism of interest for subnetworks that are similar to subnetworks in other, better-understood organisms [75,76]. The network we used to search the “target” network is a well-studied functional module from a model



organism. Such queries can find similar modules in the less studied target organism, permitting us to link information about biological functions across organisms [77]. We have used this method in previous studies to predict hitherto unrecognized transcriptional regulators and cell cycle regulators playing important roles in the parasite life cycle [55,56]. We adopted the same algorithm to identify proteins involved in pathogenesis.

We first mapped a set of proteins related to pathogenesis (GO: 0009405) from *E. coli* onto its own PPI network. A set of “neighbors” was selected for each pathogenesis protein in *E. coli*, creating a set of subnetworks, which, by inference, form a network of subnetworks in the query network. Similarly, each *P. falciparum* protein was

mapped into its own PPI network, and a subnetwork of neighbors was constructed. Subnetwork size was controlled by including only proteins that are k hops from the central protein in each case and k was chosen such that the neighbor size was under 500, unless the central protein had more than 500 directly connected neighbors.

After deducing the neighborhood subnetworks for both the query *E. coli* pathogenesis-related proteins and all *P. falciparum* proteins, the *E. coli* subnetworks were combinatorically aligned against the *P. falciparum* subnetworks. When good alignments were found (see below), the central protein of the best-aligned *P. falciparum* subnetwork was labeled a functional

ortholog of the cognate protein at the center of the relevant pathogenesis subnetwork in *E. coli*.

In order to analyze and score the degree of similarity between the *P. falciparum* and *E. coli* neighborhood subnetworks a shortest-path graph kernel was used to measure the similarity between two labeled networks, and a numerical score for each alignment was assigned [78]. To optimize the graph kernel, only paths between the central protein and other subnetwork proteins are counted. Each shortest path through the central protein can be considered a chain of cellular activities, and the path defines the dynamic function of this protein. Given a subnetwork S_p with central protein p and a query subnetwork S_q with central protein q , the shortest path similarity function is defined as follows,

$$K(S_q, S_p) = \frac{1}{|S_q| + |S_p|} \prod_{\forall (i1, i2) \in S_q} B((i1, i2), S_p)$$

where

$$B((i1, i2), S_p) = \max_{\forall (j1, j2) \in S_p} \frac{2E(i1, j1)E(i2, j2)}{dist(i1, i2) + dist(j1, j2)},$$

$$E(x, y) = \exp\left(-\frac{Eval(x, y)}{\sigma}\right) \text{ with the normalization}$$

parameter $\sigma = 10$ measures the sequence similarity between proteins x and y based on the E-value of the sequence alignment, and $dist(x, y)$ is the length of the shortest path connecting proteins x and y in the PPI subnetwork. The computation was done on a $-\log_{10}$ scale. The method outlined here takes each pair of proteins $(i1, i2)$ from one subnetwork and seeks the maximum ratio of sequence similarity with respect to the closeness (shortest path through the central protein) of the networks, in order to identify proteins $(j1, j2)$ in the target subnetwork. Using this algorithm, a subnetwork alignment score is obtained by collecting the shortest paths between two neighborhood subnetworks, getting an alignment score for each pair of proteins, and totaling all of the alignment values. Hence, by quantifying the sequence similarity and network similarity and evaluating the role of the central protein in the query network we can summarize the functional coherence, and distance between two central proteins, as a numerical score.

Network data and network analysis

We downloaded protein-protein interaction data for *E. coli* from the IntAct database [79], and protein-protein association data for *P. falciparum* from the STRING database [45]. STRING uses numerous data types, including sequence similarity estimates, pathway analysis, chromosome synteny, genome organization, and phylogenetic reconstruction, as well as literature text mining to estimate

association confidence scores (S), ranging from 0.15 to 0.999. The data are integrated using a Bayesian approach and the scores approximate the likelihood of finding the pairs of proteins in the same pathway. Cytoscape 3.1 was used for network visualization [59]. EggNOG database was used for functional classification of the network nodes [48]. NetworkAnalyzer was used to compute topological parameters and centrality measures of the cellular networks [49].

List of abbreviations used:

GO: Gene Ontology

HAT: histone acetyltransferase

IPK: inositol polyphosphate kinase

PCNA: proliferating cell nuclear antigen

PfEMP1: *Plasmodium falciparum* erythrocyte membrane protein

PfRACK: *P. falciparum* receptor for activated C kinase

PPI: protein-protein interaction

RBC: red blood cell

RRE: ribosome-recycling factor

SUMO: small ubiquitin-related modifier

UPS: ubiquitin-proteasome system

Additional material

Additional file 1: *P. falciparum* proteins that are annotated to be associated with Gene Ontology (GO) term GO0009405 (pathogenesis).

Additional File 2: Predicted functional orthologs involved in pathogenesis in *P. falciparum*. The query genome is *P. falciparum*, and the target genome is *E. coli*. GO: Gene Ontology. BP: Biological Process. MF: Molecular Function. CC: Cellular Component.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

YW and RK conceived and designed the study. All authors performed bioinformatics data analysis and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank PlasmoDB for providing access to malaria omic data. This work is supported by NIH grants GM100806, GM081068 and AI080579 to YW. KR and CH are supported by University of Minnesota Grant-in-Aid of Research, Artistry and Scholarship. We thank the Computational Biology Initiative at UTSA for providing computational support. This work received computational support from Computational Systems Biology Core, funded by the National Institute on Minority Health and Health Disparities (G12MD007591) from the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences, National Institute of Allergy and Infectious Diseases, National Institute on Minority Health and Health Disparities, or the National Institutes of Health.

Declarations

The publication costs for this article were funded by the corresponding author YW.

This article has been published as part of BMC Systems Biology Volume 9 Supplement 4, 2015: Selected articles from The International Conference on

Intelligent Biology and Medicine (ICIBM) 2014: Systems Biology. The full contents of the supplement are available online at <http://www.biomedcentral.com/bmcsystbiol/supplements/9/S4>.

Authors' details

¹Department of Biology, University of Texas at San Antonio, San Antonio, TX 78249, USA. ²Novozymes NA, Durham, NC 27709, USA. ³Department of Computer Science and Engineering, University of Minnesota Twin Cities, Minneapolis, MN 55455, USA. ⁴Department of Biology, College of Staten Island, City University of New York, Staten Island, NY 10314, USA. ⁵South Texas Center for Emerging Infectious Diseases, University of Texas at San Antonio, San Antonio, TX 78249, USA.

Published: 11 June 2015

References

- Wu Y, Wang X, Liu X, Wang Y: **Data-mining approaches reveal hidden families of proteases in the genome of malaria parasite.** *Genome research* 2003, **13**(4):601-616.
- Wiesner J, Jomaa H: **Isoprenoid biosynthesis of the apicoplast as drug target.** *Curr Drug Targets* 2007, **8**(1):3-13.
- Dvorin JD, Martyn DC, Patel SD, Grimley JS, Collins CR, Hopp CS, Bright AT, Westenberger S, Winzler E, Blackman MJ, *et al*: **A plant-like kinase in *Plasmodium falciparum* regulates parasite egress from erythrocytes.** *Science* 2010, **328**(5980):910-912.
- Solyakov L, Halbert J, Alam MM, Semblat JP, Dorin-Semblat D, Reininger L, Bottrill AR, Mistry S, Abdi A, Fennell C, *et al*: **Global kinomic and phospho-proteomic analyses of the human malaria parasite *Plasmodium falciparum*.** *Nature communications* 2011, **2**:565.
- Acharya P, Kumar R, Tatu U: **Chaperoning a cellular upheaval in malaria: Heat shock proteins in *Plasmodium falciparum*.** *Molecular and biochemical parasitology* 2007, **153**(2):85-94.
- Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, Crabtree J, Angiuoli SV, Merino EF, Amedeo P, *et al*: **Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*.** *Nature* 2008, **455**(7214):757-763.
- Carlton JM, Angiuoli SV, Suh BB, Kooij TW, Perlea M, Silva JC, Ermolaeva MD, Allen JE, Selengut JD, Koo HL, *et al*: **Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*.** *Nature* 2002, **419**(6906):512-519.
- Hall N, Karras M, Raine JD, Carlton JM, Kooij TW, Berriman M, Florens L, Janssen CS, Pain A, Christophides GK, *et al*: **A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses.** *Science* 2005, **307**(5706):82-86.
- Pain A, Bohme U, Berry AE, Mungall K, Finn RD, Jackson AP, Mourier T, Mistry J, Pasini EM, Aslett MA, *et al*: **The genome of the simian and human malaria parasite *Plasmodium knowlesi*.** *Nature* 2008, **455**(7214):799-803.
- Tachibana S, Sullivan SA, Kawai S, Nakamura S, Kim HR, Goto N, Arisue N, Palacpac NM, Honma H, Yagi M, *et al*: ***Plasmodium cynomolgi* genome sequences provide insight into *Plasmodium vivax* and the monkey malaria clade.** *Nature genetics* 2012, **44**(9):1051-1055.
- Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, *et al*: **Genome sequence of the human malaria parasite *Plasmodium falciparum*.** *Nature* 2002, **419**(6906):498-511.
- Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, DeRisi JL: **The Transcriptome of the Intraerythrocytic Developmental Cycle of *Plasmodium falciparum*.** *PLoS biology* 2003, **1**(1):E5.
- Bozdech Z, Mok S, Hu G, Imwong M, Jaidee A, Russell B, Ginsburg H, Nosten F, Day NP, White NJ, *et al*: **The transcriptome of *Plasmodium vivax* reveals divergence and diversity of transcriptional regulation in malaria parasites.** *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**(42):16290-16295.
- Dharia NV, Bright AT, Westenberger SJ, Barnes SW, Batalov S, Kuhlen K, Borboa R, Federe GC, McClean CM, Vinetz JM, *et al*: **Whole-genome sequencing and microarray analysis of ex vivo *Plasmodium vivax* reveal selective pressure on putative drug resistance genes.** *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**(46):20045-20050.
- Dharia NV, Sidhu AB, Cassera MB, Westenberger SJ, Bopp SE, Eastman RT, Plouffe D, Batalov S, Park DJ, Volkman SK, *et al*: **Use of high-density tiling microarrays to identify mutations globally and elucidate mechanisms of drug resistance in *Plasmodium falciparum*.** *Genome biology* 2009, **10**(2):R21.
- Florent I, Porcel BM, Guillaume E, Da Silva C, Artiguenave F, Marechal E, Brehelin L, Gascuel O, Chameau S, Wincker P, *et al*: **A *Plasmodium falciparum* FcB1-schizont-EST collection providing clues to schizont specific gene structure and polymorphism.** *BMC genomics* 2009, **10**:235.
- Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, De La Vega P, Holder AA, Batalov S, Carucci DJ, *et al*: **Discovery of gene function by expression profiling of the malaria parasite life cycle.** *Science* 2003, **301**(5639):1503-1508.
- Llinas M, Bozdech Z, Wong ED, Adai AT, DeRisi JL: **Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains.** *Nucleic acids research* 2006, **34**(4):1166-1173.
- Lopez-Barragan MJ, Lemieux J, Quinones M, Williamson KC, Molina-Cruz A, Cui K, Barillas-Mury C, Zhao K, Su XZ: **Directional gene expression and antisense transcripts in sexual and asexual stages of *Plasmodium falciparum*.** *BMC genomics* 2011, **12**:587.
- Mair GR, Braks JA, Garver LS, Wiegant JC, Hall N, Dirks RW, Khan SM, Dimopoulos G, Janse CJ, Waters AP: **Regulation of sexual development of *Plasmodium* by translational repression.** *Science* 2006, **313**(5787):667-669.
- Otto TD, Wilinski D, Assefa S, Keane TM, Sarry LR, Bohme U, Lemieux J, Barrell B, Pain A, Berriman M, *et al*: **New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq.** *Molecular microbiology* 2010, **76**(1):12-24.
- Vignali M, Armour CD, Chen J, Morrison R, Castle JC, Biery MC, Bouzek H, Moon W, Babak T, Fried M, *et al*: **NSR-seq transcriptional profiling enables identification of a gene signature of *Plasmodium falciparum* parasites infecting children.** *J Clin Invest* 2011, **121**(3):1119-1129.
- Watanabe J, Wakaguri H, Sasaki M, Suzuki Y, Sugano S: **Comparative study of transcriptomes of parasites defined by full-length cDNAs.** *Nucleic acids research* 2007, **35**(Database):D431-438.
- Otto TD, Bohme U, Jackson AP, Hunt M, Franke-Fayard B, Hoeijmakers WA, Religa AA, Robertson L, Sanders M, Ogun SA, *et al*: **A comprehensive evaluation of rodent malaria parasite genomes and gene expression.** *BMC biology* 2014, **12**(1):86.
- Bowyer PW, Simon GM, Cravatt BF, Bogoy M: **Global profiling of proteolysis during rupture of *Plasmodium falciparum* from the host erythrocyte.** *Mol Cell Proteomics* 2010, **10**(5):M110 001636.
- Florens L, Liu X, Wang YF, Yang SG, Schwartz O, Peglar M, Carucci DJ, Yates JR, Wu YM: **Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes.** *Molecular and biochemical parasitology* 2004, **135**(1):1-11.
- Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, Haynes JD, Moch JK, Muster N, Sacchi JB, Tabb DL, *et al*: **A proteomic view of the *Plasmodium falciparum* life cycle.** *Nature* 2002, **419**(6906):520-526.
- Lasonder E, Ishihama Y, Andersen JS, Vermunt AMW, Pain A, Sauerwein RW, Eling WMC, Hall N, Waters AP, Stunnenberg HG, *et al*: **Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry.** *Nature* 2002, **419**(6906):537-542.
- Lasonder E, Janse CJ, van Gemert GJ, Mair GR, Vermunt AM, Douradinha BG, van Noort V, Huynen MA, Luty AJ, Kroeze H, *et al*: **Proteomic profiling of *Plasmodium* sporozoite maturation identifies new proteins essential for parasite development and infectivity.** *PLoS pathogens* 2008, **4**(10):e1000195.
- Roobsoong W, Roytrakul S, Sattabongkot J, Li J, Udomsangpetch R, Cui L: **Determination of the *Plasmodium vivax* schizont stage proteome.** *Journal of proteomics* 2011, **74**(9):1701-1710.
- Silvestrini F, Lasonder E, Olivieri A, Camarda G, van Schaijk B, Sanchez M, Younis S, Sauerwein R, Alano P: **Protein export marks the early phase of gametocytogenesis of the human malaria parasite *Plasmodium falciparum*.** *Mol Cell Proteomics* 2010, **9**(7):1437-1448.
- Tarun AS, Peng X, Dumpit RF, Ogata Y, Silva-Rivera H, Camargo N, Daly TM, Bergman LW, Kappe SH: **A combined transcriptome and proteome survey of malaria parasite liver stages.** *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**(1):305-310.
- Date SV, Stoeckert CJ Jr: **Computational modeling of the *Plasmodium falciparum* interactome reveals protein function on a genome-wide scale.** *Genome research* 2006, **16**(4):542-549.
- LaCount DJ, Schoenfeld LW, Fields S: **Selection of yeast strains with enhanced expression of *Plasmodium falciparum* proteins.** *Molecular and biochemical parasitology* 2009, **163**(2):119-122.

35. LaCount DJ, Vignali M, Chettier R, Phansalkar A, Bell R, Hesselberth JR, Schoenfeld LW, Ota I, Sahasrabudhe S, Kurschner C, et al: **A protein interaction network of the malaria parasite *Plasmodium falciparum*.** *Nature* 2005, **438**(7064):103-107.
36. Suthram S, Sittler T, Ideker T: **The *Plasmodium* protein network diverges from those of other eukaryotes.** *Nature* 2005, **438**(7064):108-112.
37. Yuan J, Cheng KC, Johnson RL, Huang R, Pattaradilokrat S, Liu A, Guha R, Fidock DA, Inglese J, Wellems TE, et al: **Chemical genomic profiling for antimalarial therapies, response signatures, and molecular targets.** *Science* 2011, **333**(6043):724-729.
38. Ginsburg H: **Progress in silico functional genomics: the malaria Metabolic Pathways database.** *Trends in parasitology* 2006, **22**(6):238-240.
39. Ginsburg H, Tilley L: ***Plasmodium falciparum* metabolic pathways (MPMP) project upgraded with a database of subcellular locations of gene products.** *Trends in parasitology* 2011.
40. Lakshmanan V, Rhee KY, Daily JP: **Metabolomics and malaria biology.** *Molecular and biochemical parasitology* 2011, **175**(2):104-111.
41. Plata G, Hsiao TL, Olszewski KL, Llinas M, Vitkup D: **Reconstruction and flux-balance analysis of the *Plasmodium falciparum* metabolic network.** *Mol Syst Biol* 2010, **6**:408.
42. Yeh I, Hanekamp T, Tsoka S, Karp PD, Altman RB: **Computational analysis of *Plasmodium falciparum* metabolism: organizing genomic information to facilitate drug discovery.** *Genome research* 2004, **14**(5):917-924.
43. Mu JB, Ferdig MT, Feng XR, Joy DA, Duan JH, Furuya T, Subramanian G, Aravind L, Cooper RA, Wootton JC, et al: **Multiple transporters associated with malaria parasite responses to chloroquine and quinine.** *Molecular microbiology* 2003, **49**(4):977-989.
44. Schrag SJ, Perrot V, Levin BR: **Adaptation to the fitness costs of antibiotic resistance in *Escherichia coli*.** *Proceedings Biological sciences / The Royal Society* 1997, **264**(1386):1287-1291.
45. Frank SA, Schmid-Hempel P: **Mechanisms of pathogenesis and the evolution of parasite virulence.** *Journal of evolutionary biology* 2008, **21**(2):396-404.
46. Bandyopadhyay S, Sharan R, Ideker T: **Systematic identification of functional orthologs based on protein network comparison.** *Genome research* 2006, **16**(3):428-435.
47. Bruckner S, Huffner F, Karp RM, Shamir R, Sharan R: **TORQUE: topology-free querying of protein interaction networks.** *Nucleic acids research* 2009, **37**(Web Server):W106-108.
48. Kelley BP, Yuan B, Lewitter F, Sharan R, Stockwell BR, Ideker T: **PathBLAST: a tool for alignment of protein interaction networks.** *Nucleic acids research* 2004, **32**(Web Server):W83-88.
49. Klau GW: **A new graph-based method for pairwise global network alignment.** *BMC bioinformatics* 2009, **10**(Suppl 1):S59.
50. Koyuturk M, Kim Y, Subramaniam S, Szpankowski W, Grama A: **Detecting conserved interaction patterns in biological networks.** *Journal of computational biology : a journal of computational molecular cell biology* 2006, **13**(7):1299-1322.
51. Kuang R, le E, Wang K, Wang K, Siddiqi M, Freund Y, Leslie C: **Profile-based string kernels for remote homology detection and motif extraction.** *J Bioinform Comput Biol* 2005, **3**(3):527-550.
52. Sharan R, Suthram S, Kelley RM, Kuhn T, McCuine S, Uetz P, Sittler T, Karp RM, Ideker T: **Conserved patterns of protein interaction in multiple species.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**(6):1974-1979.
53. Singh R, Xu J, Berger B: **Global alignment of multiple protein interaction networks.** *Pac Symp Biocomput* 2008, 303-314.
54. Zhenping L, Zhang S, Wang Y, Zhang XS, Chen L: **Alignment of molecular networks by integer quadratic programming.** *Bioinformatics* 2007, **23**(13):1631-1639.
55. Cai H, Hong CJ, Gu JY, Lilburn TG, Kuang R, Wang YF: **Module-based subnetwork alignments reveal novel transcriptional regulators in malaria parasite *Plasmodium falciparum*.** *BMC Systems Biology* 2012, **6**.
56. Cai H, Hong C, Lilburn TG, Rodriguez AL, Chen S, Gu J, Kuang R, Wang Y: **A novel subnetwork alignment approach predicts new components of the cell cycle regulatory apparatus in *Plasmodium falciparum*.** *BMC bioinformatics* 2013, **14**(Suppl 12):S2.
57. Miller LH, Ackerman HC, Su XZ, Wellems TE: **Malaria biology and disease pathogenesis: insights for new treatments.** *Nature medicine* 2013, **19**(2):156-167.
58. Pasternak ND, Dzikowski R: **PfEMP1: an antigen that plays a key role in the pathogenicity and immune evasion of the malaria parasite *Plasmodium falciparum*.** *The international journal of biochemistry & cell biology* 2009, **41**(7):1463-1466.
59. Lavstsen T, Turner L, Saguti F, Magistrado P, Rask TS, Jespersen JS, Wang CW, Berger SS, Baraka V, Marquard AM, et al: ***Plasmodium falciparum* erythrocyte membrane protein 1 domain cassettes 8 and 13 are associated with severe malaria in children.** *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**(26):E1791-1800.
60. Assenov Y, Ramirez F, Schelhorn SE, Lengauer T, Albrecht M: **Computing topological parameters of biological networks.** *Bioinformatics* 2008, **24**(2):282-284.
61. Hershko A, Ciechanover A: **The ubiquitin system.** *Annual review of biochemistry* 1998, **67**:425-479.
62. Komander D: **The emerging complexity of protein ubiquitination.** *Biochemical Society transactions* 2009, **37**(Pt 5):937-953.
63. Chung DW, Le Roch KG: **Targeting the *Plasmodium* ubiquitin/proteasome system with anti-malarial compounds: promises for the future.** *Infectious disorders drug targets* 2010, **10**(3):158-164.
64. Sartorello R, Amaya MJ, Nathanson MH, Garcia CR: **The plasmodium receptor for activated C kinase protein inhibits Ca(2+) signaling in mammalian cells.** *Biochemical and biophysical research communications* 2009, **389**(4):586-592.
65. Stelter P, Ulrich HD: **Control of spontaneous and damage-induced mutagenesis by SUMO and ubiquitin conjugation.** *Nature* 2003, **425**(6954):188-191.
66. de la Cruz J, Kressler D, Linder P: **Unwinding RNA in *Saccharomyces cerevisiae*: DEAD-box proteins and related families.** *Trends in biochemical sciences* 1999, **24**(5):192-198.
67. Pradhan A, Tuteja R: **Bipolar, Dual *Plasmodium falciparum* helicase 45 expressed in the intraerythrocytic developmental cycle is required for parasite growth.** *Journal of molecular biology* 2007, **373**(2):268-281.
68. Tuteja R: **Genome wide identification of *Plasmodium falciparum* helicases: a comparison with human host.** *Cell cycle* 2010, **9**(1):104-120.
69. Gupta AP, Chin WH, Zhu L, Mok S, Luah YH, Lim EH, Bozdech Z: **Dynamic epigenetic regulation of gene expression during the life cycle of malaria parasite *Plasmodium falciparum*.** *PLoS pathogens* 2013, **9**(2):e1003170.
70. Cui L, Miao J: **Chromatin-mediated epigenetic regulation in the malaria parasite *Plasmodium falciparum*.** *Eukaryotic cell* 2010, **9**(8):1138-1149.
71. Ponts N, Harris EY, Prudhomme J, Wick I, Eckhardt-Ludka C, Hicks GR, Hardiman G, Lonardi S, Le Roch KG: **Nucleosome landscape and control of transcription in the human malaria parasite.** *Genome research* 2010, **20**(2):228-238.
72. Fan Q, An L, Cui L: ***Plasmodium falciparum* histone acetyltransferase, a yeast GCN5 homologue involved in chromatin remodeling.** *Eukaryotic cell* 2004, **3**(2):264-276.
73. Tonkin CJ, Carret CK, Duraisingh MT, Voss TS, Ralph SA, Hommel M, Duffy MF, Silva LM, Scherf A, Ivens A, et al: **Sir2 paralogs cooperate to regulate virulence genes and antigenic variation in *Plasmodium falciparum*.** *PLoS biology* 2009, **7**(4):e84.
74. Duraisingh MT, Voss TS, Marty AJ, Duffy MF, Good RT, Thompson JK, Freitas-Junior LH, Scherf A, Crabb BS, Cowman AF: **Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*.** *Cell* 2005, **121**(1):13-24.
75. Sharan R, Ideker T: **Modeling cellular machinery through biological network comparison.** *Nature biotechnology* 2006, **24**(4):427-433.
76. Barabasi AL, Oltvai ZN: **Network biology: understanding the cell's functional organization.** *Nature reviews Genetics* 2004, **5**(2):101-113.
77. Sahraeian SME, Yoon BJ: **Fast Network Querying Algorithm for Searching Large-Scale Biological Networks.** *Int Conf Acoust Spee* 2011, 6008-6011.
78. Borgwardt KM, Kriegel HP, Vishwanathan SV, Schraudolph NN: **Graph kernels for disease outcome prediction from protein-protein interaction networks.** *Pac Symp Biocomput* 2007, 4-15.
79. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, Campbell NH, Chavali G, Chen C, del-Toro N, et al: **The MintAct project-IntAct as a common curation platform for 11 molecular interaction databases.** *Nucleic acids research* 2014, **42**(Database):D358-363.
80. Muller J, Szklarczyk D, Julien P, Letunic I, Roth A, Kuhn M, Powell S, von Mering C, Doerks T, Jensen LJ, et al: **eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised**

orthologous groups, species and functional annotations. *Nucleic acids research* 2010, **38**(Database):D190-195.

81. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, *et al*: **The COG database: an updated version includes eukaryotes.** *BMC bioinformatics* 2003, **4**:-.
82. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguéz P, Doerks T, Stark M, Muller J, Bork P, *et al*: **The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored.** *Nucleic acids research* 2010, **39**(Database):D561-568.
83. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguéz P, Bork P, von Mering C, *et al*: **STRING v9.1: protein-protein interaction networks, with increased coverage and integration.** *Nucleic acids research* 2013, **41**(Database):D808-815.

doi:10.1186/1752-0509-9-S4-S1

Cite this article as: Cai *et al*: Predicting and exploring network components involved in pathogenesis in the malaria parasite via novel subnetwork alignments. *BMC Systems Biology* 2015 **9**(Suppl 4):S1.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

