nature biotechnology

RESOURCE

Transcriptional profiling of growth perturbations of the human malaria parasite *Plasmodium falciparum*

Guangan Hu[1,3], Ana Cabrera[2,3], Maya Kono[2], Sachel Mok[1], Balbir K Chaal[1], Silvia Haase[2], Klemens Engelberg[2], Sabna Cheemadan[1], Tobias Spielmann[2], Peter R Preiser[1], Tim-W Gilberger[2] & Zbynek Bozdech[1]

Functions have yet to be defined for the majority of genes of *Plasmodium falciparum*, the agent responsible for the most serious form of human malaria. Here we report changes in *P. falciparum* gene expression induced by 20 compounds that inhibit growth of the schizont stage of the intraerythrocytic development cycle. In contrast with previous studies, which reported only minimal changes in response to chemically induced perturbations of *P. falciparum* growth, we find that ~59% of its coding genes display over three-fold changes in expression in response to at least one of the chemicals we tested. We use this compendium for guilt-by-association prediction of protein function using an interaction network constructed from gene co-expression, sequence homology, domain-domain and yeast two-hybrid data. The subcellular localizations of 31 of 42 proteins linked with merozoite invasion is consistent with their role in this process, a key target for malaria control. Our network may facilitate identification of novel antimalarial drugs and vaccines.

Together with AIDS/HIV and tuberculosis, human malaria represents one of the three most dangerous infectious diseases of humankind[1]. In 2007, 1.38 billion people were estimated to be at risk of infection with *P. falciparum*, the protozoan endoparasite responsible for up to 2 million annual human deaths from malaria[2,3]. The lack of an effective vaccine and the rapid spread of resistance to most antimalarial drugs are major concerns for the control of this unicellular eukaryote. In particular, the complexity of the *P. falciparum* life cycle, which is associated with many unique morphological and metabolic states, has challenged efforts to identify parasite-specific molecular mechanisms that can be targeted by new malaria intervention strategies[4].

The genome of *P. falciparum* encodes ~5,300 genes. This obligate endoparasite has lost many basic metabolic abilities, such as a majority of the enzymes of amino acid synthesis, but expanded its repertoire of proteins involved in many parasite-specific functions, such as interaction with its host, antigenic variation and host-cell invasion[5]. This is consistent with the difficulty in predicting functions for the majority of *P. falciparum* proteins. Genome-wide approaches offer an attractive method to accelerate functional annotation of the *P. falciparum* genome.

The haploid state of the genome throughout the majority of the *P. falciparum* life cycle and lack of inducible knockout or RNAi-mediated knockdown systems for this parasite limits the application of forward and reverse genetic approaches to assess gene function in this species[6,7]. Moreover, the low efficiency of the available transfection technologies makes genetic modification of *P. falciparum* too costly and time consuming for genome-wide analyses. Although the potential of systems biology approaches to derive functional gene predictions is widely appreciated[8], previous efforts to predict the functions of uncharacterized *P. falciparum* gene products were based on gene interaction networks derived mainly from probabilistic integration of transcriptome data collected at different stages of the *P. falciparum* life cycle[9–11]. Largely because many genes with unrelated functions exhibit similar transcriptional profiles across the *P. falciparum* life cycle[12,13], these approaches provided relatively low-confidence predictions of gene function.

Although studies with model organisms such as yeast and *Caenorhabditis elegans* suggest that microarray analyses of global transcriptional responses to growth perturbations can substantially improve the accuracy and coverage of probabilistic interaction networks[14,15], the utility of monitoring changes in gene expression in response to growth perturbations for predicting *P. falciparum* gene function has been controversial. Some perturbations, including those associated with several antimalarial drugs, such as chloroquine and several antifolates, induced only low-amplitude mRNA changes with no particular link to their presumed mode of action[16,17]. On the other hand, exposure of *P. falciparum* parasites to febrile temperatures[18], artesunate[19] and an inhibitor of sphingomyeline synthase[20] induced biologically relevant transcriptional changes that led to the identification of proteins associated with these processes.

Here we demonstrate that DNA microarray-based profiling of growth perturbations in *P. falciparum* can generate a high-resolution transcriptional data set that reflects functional relationships between *P. falciparum* genes. We use this data set to construct a gene interaction network that predicts the functions of 2,545 *P. falciparum* hypothetical proteins with confidence levels comparable to those of similar

[1]Division of Genetics and Genomics, School of Biological Sciences, Nanyang Technological University, Singapore. [2]Bernhard Nocht Institute for Tropical Medicine, Department of Molecular Parasitology, Hamburg, Germany. [3]These authors contributed equally to this work. Correspondence should be addressed to T.-W.G. (gilberger@bni.uni-hamburg.de) or Z.B. (zbozdech@ntu.edu.sg).

Received 8 September; accepted 6 December; published online 27 December 2009; doi:10.1038/nbt.1597

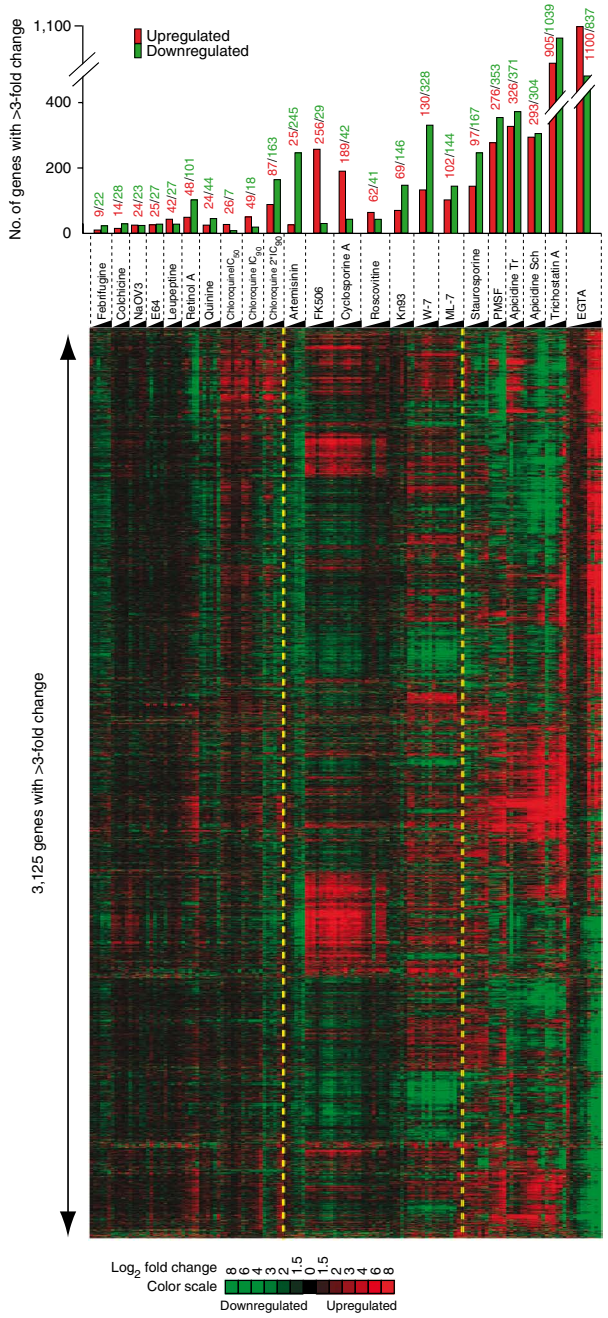


Figure 1 Overview of the gene expression responses of *P. falciparum* to growth perturbation induced by drug or inhibitor treatments. The heatmap summarizes global transcriptional responses to 20 compounds conducted in 23 time-course experiments with a total of 144 microarrays. A total of 3,125 genes that show at least a threefold change in mRNA abundance in at least one experiment are included in the overview data set. The color scale indicates upregulation or downregulation of each individual mRNA transcript compared to the corresponding time point in control untreated cells (**Supplementary Table 1**). The bar diagram (top) indicates the total number of genes that show more than threefold upregulation (red bar) or downregulation (green bar) in each treatment experiment. The number of up- and downregulated genes is also indicated. The treatment experiments were ordered according to the total number of genes with altered expression (more than threefold) and grouped (yellow dashed lines) according to the number of genes with altered levels of their mRNA levels (see text). The treatment experiments were conducted in the time courses indicated along the horizontal axis and genes were arranged using hierarchical clustering.

approaches applied for well-studied model organisms[21,22]. We focused mainly on the late stage (schizont) of the *P. falciparum* intraerythrocytic developmental cycle (IDC) to target the key process of parasite invasion and identified a subnetwork that encompasses 416 genes likely to participate in this process. Using a green fluorescent protein (GFP)-tagging approach, we demonstrate that 31 of 42 genes selected from the subnetwork localize within cellular compartments directly associated with host-cell invasion.

RESULTS

Transcriptional profiling of growth perturbations

We carried out microarray measurements of *P. falciparum* global transcriptional responses to 20 growth-inhibiting compounds (**Fig. 1** and **Supplementary Table 1**). For each compound, synchronized *P. falciparum* cells were exposed to inhibitory concentrations (IC) of 50 (IC_{50}) or 90 (IC_{90}) determined individually for each drug and RNA samples were collected from multiple time points (**Supplementary Table 1**).

A total of 3,125 genes exhibited at least a threefold increase or decrease in transcript level after exposure to at least one chemical stimulus for at least one of the time points after initiating growth perturbation (**Fig. 1** and **Supplementary Table 2**). Using a threefold change in transcript abundance as a cutoff for transcriptional modulation, we loosely classify the transcriptional responses into three compound classes.

The first class induced <50 genes (~1% of the genome) and had an overall transcriptional effect on <250 genes (~5% of the genome). This includes compounds like colchicine, Na_3VO_4, E64, leupeptin and two of the three tested antimalarial drugs, chloroquine and quinine (**Fig. 1**). These results are reminiscent of those in reports that revealed unusually low levels of transcriptional responses to highly toxic antimalarial drugs[16,17]. Despite their low amplitudes, these responses were, however, highly reproducible and specific to each compound[16,17]. In agreement with this, we observed highly reproducible responses of *P. falciparum* to chloroquine (data not shown) that were also dose dependent (26, 49 and 87 genes were induced more than threefold and 194, 257 and 330 genes, more than twofold with IC_{50}, IC_{90} and 2*IC_{90} concentrations, respectively).

We found only moderate overlap between our results and previously published data[17,19]. Compared with these studies, only 12.5% and 10% of the genes whose expression was altered by chloroquine and artemisinin, respectively, were also found to be differentially expressed. Differences in experimental design that might account for these dissimilarities may relate to the considerably higher drug concentrations used previously, different representations of the developmental stages in starting cultures (e.g., asynchronized parasites for chloroquine studies[17]) and different approaches to data analyses (e.g., filtering of genes with stage-specific expression in the artesunate study[19]). Despite these discrepancies, our experiments and the previous published work showed genes with highly reproducible and dose-dependent responses to these malaria drugs. This suggests that, despite their low amplitudes and broad gene representations, transcriptional changes in response to chemical stimuli may reflect physiologically relevant processes involving functionally related genes.

The second class of compounds induced transcription of >50 genes (~1%) and overall involved 250–500 genes (~5–10%). This includes inhibitors of calcium/calmodulin-dependent protein kinases (CDPK; ML-7 and W-7) and the calcineurin pathway (FK506 and cyclosporine A), all of which inhibited the development of the schizont stage (**Supplementary Fig. 1**). We observed striking similarities

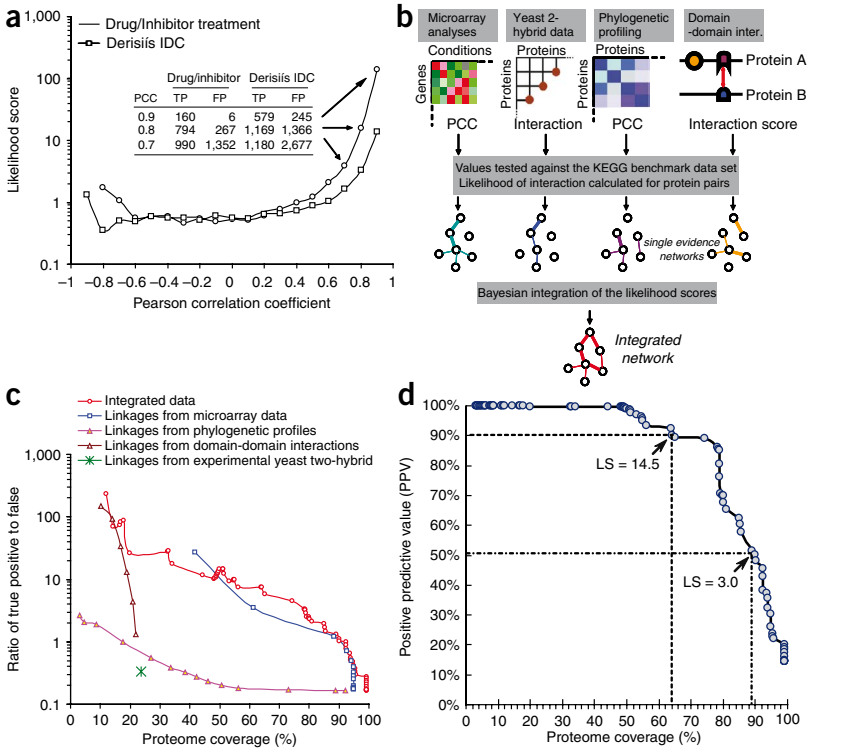


Figure 2 Reconstruction of the PlasmoINT interaction network. (**a**) The plot depicts the likelihood of functional relationships along the correlation of mRNA abundance profiles for all gene pairs in the microarray data. Pearson Correlation Coefficients (PCC) were calculated for every pair of the 492 *P. falciparum* genes with KEGG functional assignments in both perturbation data sets (Drug/inhibitor) and the IDC transcriptome[12]. The numbers of false-positive (FP) and true-positive (TP) gene pairs in the high PCC bins are indicated in the inset table. (**b**) Flow chart describing assembly of the interaction network. The four input data sets were evaluated for protein interaction using a relevant scoring system and score values were tested against the KEGG benchmark to derive the interaction likelihood scores that were used as an input evidence for Bayesian integration. For more details on KEGG benchmark scoring and network building, see **Supplementary Table 3.** (**c**) The relationship between proteome coverage of the individual input data sets (microarray data, phylogenetic profiles, domain-domain interaction and yeast two-hybrid system) and TP/FP ratio thresholds illustrates the contribution of each individual input to the integrated network data set. (**d**) The predictive precision rates (positive predictive value, PPV) at different likelihood score cutoffs were evaluated by tenfold cross-validation and plotted against the proteome coverage. Each dot of the ratio represents an average of ten cross-validations at a particular likelihood score cutoff. The vertical dashed line shows the likelihood score cutoffs and proteome coverage corresponding to the PPV (PPV = TP/(TP + FP)) 50% and 90% (likelihood score thresholds (LS) of 3 and 14.5). At these ratios, TP/FP was equal to 1 (~50% confidence) and 9 (~90% confidence), respectively.

in transcriptional responses induced by inhibitors within each class, which suggests that their inhibitory effect in *P. falciparum* may be very specific (**Fig. 1**). Moreover, there is only a limited overlap between the transcriptional responses induced by the CDPK and calcineurin inhibitors. This suggests that these two types of intracellular signaling pathways play specific, nonoverlapping roles in *P. falciparum* parasites that are both connected to transcriptional regulation.

The third class of compounds was able to induce transcription of >250 genes (~5%) and overall involved >500 genes (~10%). These include EGTA, phenylmethylsulfonyl fluoride, staurosporine, trichostatin A and apicidin (**Fig. 1**). With the exception of apicidin, these responses were compatible with an arrest in IDC development, indicating that the inhibitory effects of these compounds are associated with mechanisms that regulate the *P. falciparum* life cycle (**Supplementary Fig. 2**). In contrast, apicidin and to some degree trichostatin A (both histone deacetylase inhibitors) caused a general deregulation of the IDC transcriptional cascade by derepression of genes that are normally suppressed at both the trophozoite and schizont stages.

Reconstruction of a probabilistic gene functional network

To evaluate co-transcriptional properties of functionally related genes, we calculated the Pearson correlation coefficient (PCC) between transcription profiles of a subset of 492 genes that can be assigned to at least one pathway defined by the Kyoto Encyclopedia of Genes and Genomes (KEGG)[23]. Overall, we observed a disproportionately high number of functionally related genes being transcriptionally co-regulated (PCC > 0.6) (**Fig. 2a** and Online Methods). In comparison with the *P. falciparum* IDC transcriptome[12], the enrichment of functionally related genes was improved by 1.6-, 3.5- and 11-fold for the 0.7, 0.8 and 0.9 PCC thresholds, respectively (**Fig. 2a**). This high occurrence of transcriptional co-regulation among functionally related genes suggests a good potential of the perturbation data set for functional gene predictions. Hence, we used it as a core data set for the assembly of a probabilistic network in which we integrated this data set with additional inputs: (i) phylogenetic profiles with sequence homology values (E-values) of all 5,363 *P. falciparum* protein sequences to their orthologs in 210 sequenced genomes; (ii) domain-domain interactions [24]; and, (iii) yeast two-hybrid interactions[25] (**Fig. 2b** and Online Methods). In addition, the perturbation microarray data were combined with the IDC transcriptomes from three *P. falciparum* laboratory strains[26] and four field isolates[27].

To reconstruct the probabilistic network, we used the KEGG gold standard data set to calculate the likelihood score of protein interaction evidence from all four input data sets (**Supplementary Table 3**) and subsequently integrated these scores into the final score using a Bayesian integration approach (**Fig. 2b** and Online Methods). Overall, we established integrated likelihood scores for 14,168,597 functional linkages between 5,374 *P. falciparum* proteins (99.2% of the proteome). In general, the integrated likelihood scores provided higher proteome coverage than each of the individual input data sets at all probability thresholds (**Fig. 2c**). In contrast to the domain-domain interaction data set, which provides high-accuracy predictions for a small proportion of the proteome (~10%), the transcriptome data and phylogenetic profiles can provide high proteome coverage. However, their predictive values are consistently lower. In our calculations, we observed low accuracy for the protein-protein interaction data set based on the two-hybrid system[25]. This data set therefore provides a low contribution to the final likelihood scores (**Fig. 2c**).

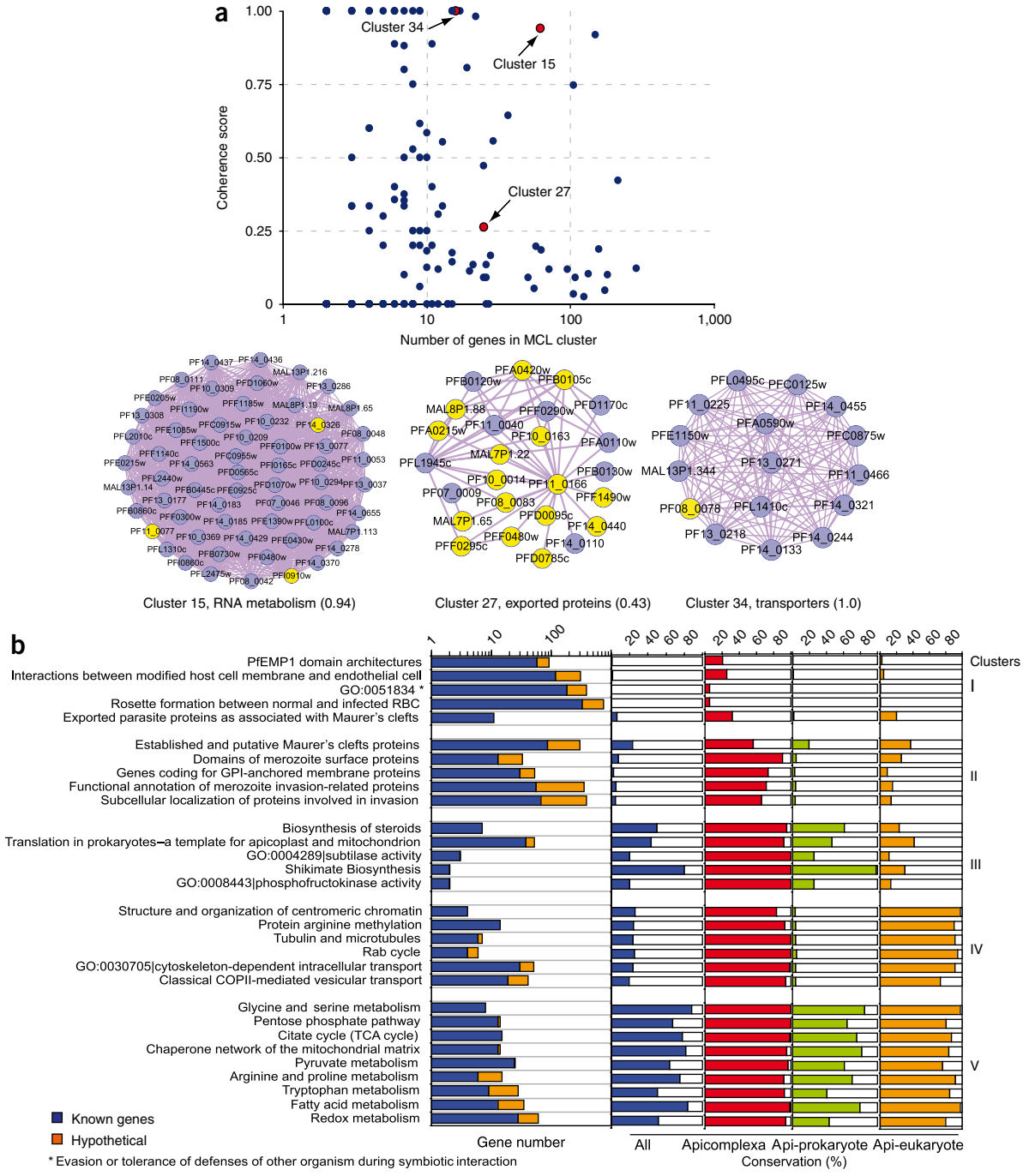


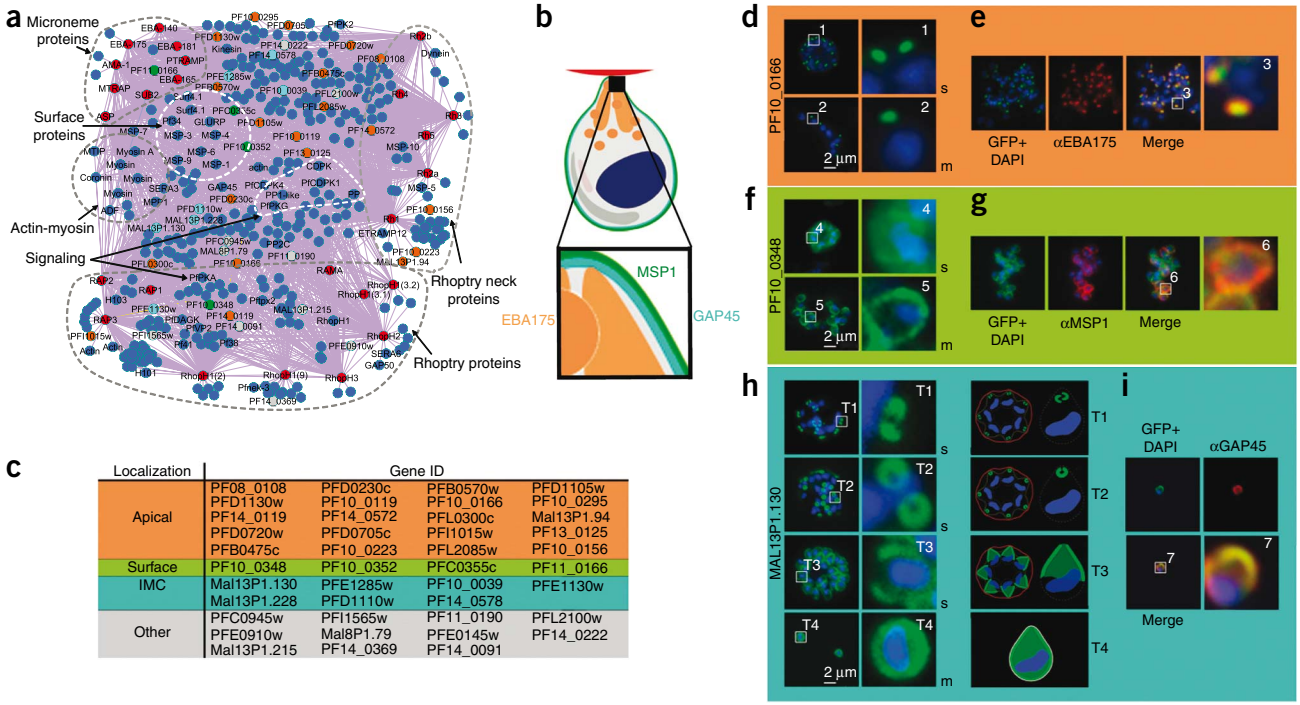
Figure 3 MCL- and WNC-based functional predictions and their functional categorizations. (**a**) Summary of the 208 MCL clusters depicted as a scatter plot with the number of genes plotted against their coherence score. Coherence score 0 corresponds to MCL cluster without any functionally characterized proteins. Examples of three clusters with high- and medium-coherence scores are indicated in the scatter plot and also drawn (below) with the functionally characterized (purple) and hypothetical proteins (yellow) linked by edges that correspond to functional links with >90% confidence. (**b**) The conservation of different functional pathways across 210 genomes including 155 prokaryotes, 6 apicomplexa and 49 other eukaryotes is summarized and indicated for selected functional gene groups (for the full list, see **Supplementary Table 5**). The conservation of each pathway is calculated independently as the fraction of the number of species containing potential homologs (reciprocal BLASTP hit, E-value ≤ 10^{-10}) according to four categories: total 210 genomes (the second panel, blue bar), apicomplexa (third panel, red bar), prokaryotes plus apicomplexa (fourth panel, green bar) and eukaryote plus apicomplexa (right panel, orange bar). Pathways were classified into five categories: genes specific to *P. falciparum* (cluster I), genes conserved in apicomplexa (II), genes conserved in apicomplexa and prokaryotes (III), genes conserved in apicomplexa and other eukaryotes (IV) and genes conserved in all 210 genomes (V). The total number of functionally characterized and hypothetical genes in each category are displayed similarly to **a**. Api-eukaryote, genes conserved in apicomplexans and eukaryotes; api-prokaryote, genes conserved in apicomplexans and prokaryotes.

Using the calculated functional linkages, we assembled two interaction networks based on likelihood score thresholds that correspond to 50% (339,721 linkages for 89% of proteome) and 90% confidence precision rates (72,748 linkages for 68% of proteome) (**Fig. 2d** and Online Methods).The connectivity of both the 50% and 90% confidence networks fits a power-law distribution with power (λ) values of 0.93 and 1.14, respectively (**Supplementary Fig. 3**). This distribution represents a typical scale-free network, well-known for protein-protein interaction networks in eukaryotic cells[28]: a small number of highly connected nodes (hubs) are linked to a larger number of less connected nodes and so on.

Modular analysis and network-based functional predictions

In the next step, we used two parallel approaches to explore the assembled network for the prediction of *P. falciparum* hypothetical protein function. First, we used the Markov cluster (MCL) algorithm[29] to define significant clusters of highly interconnected genes in the network. We used a coherence score to test enrichment of every single cluster for genes involved in a particular pathway. This analysis not only tests the quality of the network but also generates functional predictions for hypothetical genes that fall into these clusters (**Fig. 3a**). For this work, we used the 90% confidence network to provide the most conservative assessment of the network quality. Second, we used the weighted neighbor-counting (WNC) method to derive functional prediction for the hypothetical proteins. For this, we explored the 50% confidence network to maximize the number of functional predictions for hypothetical proteins. The confidence of these predictions was assessed by a 'leave-one-out' analysis[30] that is based on the efficiency of recalling functional predictions of previously characterized genes (**Supplementary Fig. 4**).

MCL identified 208 modules in the 90% confidence network, resulting in 3,029 genes being assigned to at least one of the 106 modules with functional assignments (**Fig. 3a** and **Supplementary Table 4**). The MCL modules represent many pathways conserved across the eukaryotic species (e.g., RNA metabolism) or specific to *P. falciparum* (e.g., proteins exported to the host cell cytoplasm, "exported proteins"), as well as coherent functional groups (e.g., transporters) (**Fig. 3a**). The functions of 1,376 hypothetical genes can be predicted



Localization	Gene ID			
Apical	PF08_0108	PFD0230c	PFB0570w	PFD1105w
	PFD1130w	PF10_0119	PF10_0166	PF10_0295
	PF14_0119	PF14_0572	PFL0300c	Mal13P1.94
	PFD0720w	PFD0705c	PFI1015w	PF13_0125
	PFB0475c	PF10_0223	PFL2085w	PF10_0156
Surface	PF10_0348	PF10_0352	PFC0355c	PF11_0166
IMC	Mal13P1.130	PFE1285w	PF10_0039	PFE1130w
	Mal13P1.228	PFD1110w	PF14_0578	
Other	PFC0945w	PFI1565w	PF11_0190	PFL2100w
	PFE0910w	Mal8P1.79	PFE0145w	PF14_0222
	Mal13P1.215	PF14_0369	PF14_0091	

Figure 4 Blueprint of the protein network implicated in merozoite invasion. (**a**) Subnetwork associated with merozoite invasion process. This subnetwork has a total of 2,417 links (purple lines) that are derived from the 90% confidence network and link the 25 reference genes to 25 core apical proteins (marked with red circles) with 418 proteins that include the experimentally validated (colored circles) and other proteins (blue circles). The forty-two proteins whose intracellular localization were studied are represented by a corresponding color; apical proteins (orange), merozoite surface proteins (green), IMC (turquoise) and other localization (gray). The core proteins and other previously characterized proteins were grouped manually based on their functional assignments. The dotted lines outline areas with functionally related proteins previously linked with invasion, such as microneme proteins, actin and myosin. (**b**) Schematic representation of an invasive merozoite. The apical organelles are depicted in orange, the IMC in turquoise and the surface in green. Examples for compartment-specific marker proteins are given. (**c**) Synopsis of subcellular localization of 42 proteins predicted to be involved in invasion. Proteins are grouped into either apical (orange), surface (green), IMC (turquoise) or other (gray; cytosolic, apicoplast or mitochondrial), according to their predominant localization. (**d**–**i**) Representative localization for one member of each group in late schizonts and free merozoites. Boxed regions are numbered and depicted in higher magnification to the right. The nucleus is stained with DAPI (blue). PF10_0166-GFP (green) localized to the apical region of schizonts (s) and free merozoites (m) in unfixed parasites (**d**). PF10_0166-GFP co-localized with the microneme protein EBA175 (red) in fixed parasites (**e**). PF10_0348-GFP (green) localized to the surface of schizonts and free merozoites in unfixed parasites (**f**). PF10_0348-GFP co-localized with the surface protein MSP-1 (red) in fixed parasites (**g**). Dynamics of MAL13P1.130-GFP (green) during schizogony in unfixed parasites. In early schizogony (T1), MAL13P1.130-GFP emerged as a cramp-like-structure at the apical tip of forming merozoites (**h**). This structure develops to be ring-like (T2) before becoming evenly distributed at the periphery of the nascent merozoite (T3-4). The third row shows a schematic representation. For confocal three-dimensional reconstitution, see **Supplementary Movies 1–3**. MAL13P1.130-GFP co-localized with the IMC protein GAP45 (red) in fixed parasites (**i**).

by their association to these modules, whose confidence is represented by the coherence scores. The MCL analysis suggests that the assembled network detects functionally related genes with sufficient precision. The WNC approach allows (functional) explorations of unknown genes even outside of the identified modules and generated predictions for 2,545 hypothetical proteins (95% in the genome) that can be assigned to 216 functional terms (**Supplementary Fig. 5** and **Supplementary Table 5**).

Taking advantage of the phylogenetic profiles (see above), we investigated the overall evolutionary conservation of the derived functional groups with the newly assigned genes (**Fig. 3b**). Only a small number of functional gene groups are restricted to *P. falciparum* and exhibit either no or low sequence homology to genes in other organisms, including closely related apicomplexan species. The majority of these represent the subtelomeric gene families encoding several classes of surface antigens, such as *var*, *rifin* and *stevor*, and proteins associated with Maurer's clefts (**Fig. 3b**, cluster I). Parasite invasion dominates the functional cluster that is highly conserved among apicomplexans but diverges from all other eukaryotic and prokaryotic species (**Fig. 3b**, cluster II). Cluster III depicts several *P. falciparum* functions that have a prokaryotic origin such as steroid biosynthesis (a term assigned by KEGG, corresponding to *P. falciparum* isoprenoid synthesis), translation in genes of the mitochondria and apicoplasts (non-photosynthetic plastids found in most Apicomplexa) and three homologs of proteins involved in subtilisin protease activity. Moreover, the WNC analysis assigned many new proteins to the majority of the highly conserved functional groups that are either of eukaryotic (cluster IV) or prokaryotic origin (cluster V). It is possible that many of the newly annotated genes represent evolutionarily diverse factors of these otherwise well-conserved, and thus potentially essential, pathways. The precision rates for these functional terms provide a measure of confidence for these functional predictions and help to identify candidates for previously unrecognized molecular factors that are essential for the growth, development and virulence of *P. falciparum*.

Proteins implicated in *P. falciparum* merozoite invasion

Invasion of the host's red blood cells by a specialized invasive form called the merozoite is a key step in the *P. falciparum* life cycle. To validate the predictive potential of our approach, we explored the utility of our network to identify genes associated with merozoite invasion. Merozoite invasion involves multiple molecular mechanisms ranging from specific ligand-receptor interactions, actin-myosin motility, protease activities, protein translocation and signaling[31–33]. It is mediated by an unknown number of proteins and is of high interest for drug and vaccine development because interference with this crucial biological process holds the potential to disrupt the parasite's life cycle. Although >50 proteins have been previously linked with this process, gaps remain in our understanding of the molecular mechanisms that mediate the entire invasion process. To provide a comprehensive picture of the invasion process, we generated a subnetwork of proteins that are directly connected to 25 previously established invasion-associated proteins in the 90% confidence interaction network (**Fig. 4a**). Overall, this subnetwork contains 418 proteins, including 155 with a predicted function and 263 hypothetical proteins (**Supplementary Table 6**). The subnetwork compiles the majority of proteins previously linked with invasion-like apical organelle proteins, glycosylphosphatidylinisotol-anchored surface proteins, actin-myosin motor components and signal transduction proteins. It also includes 43 out of 56 proteins recently predicted to be associated with cellular compartments of the merozoite invasion machinery[33]. Finally, 230 out of all 263 hypothetical proteins represented in the invasion subnetwork were also predicted by WNC as merozoite invasion factors.

For the functional validations, we initially selected 70 proteins from this invasion process protein subnetwork. For this selection, we prioritized proteins with a high WNC score (**Supplementary Table 5**) and gene length ≤2 kb (to facilitate cloning and expression of these proteins in *P. falciparum* transfection experiments). Open reading frames were fused with GFP and expressed ectopically in *P. falciparum* under the control of an appropriate promoter mimicking the expression profile of the endogenous allele[34]. Of these, 63 proteins could be expressed as GFP-fusion proteins in transgenic parasites, of which 42 resulted in a defined intracellular localization (**Fig. 4** and **Supplementary Fig. 6b**). From the remaining 21 GFP fusions, 11 were not expressed at sufficient levels and 10 were discarded because of retention in the endoplasmic reticulum that might be caused by the bulky GFP moiety, as described previously[34] (data not shown).

The remaining 42 proteins can be grouped according to their localization (**Fig. 4b**,**c**). The largest group consists of 20 proteins that showed a predominantly apical distribution in maturing schizonts and in free merozoites after rupture (**Fig. 4d**,**e** and **Supplementary Fig. 6a**). The second group is represented by four proteins with GFP distributed in the periphery of the parasite (**Fig. 4f**,**g** and **Supplementary Fig. 6a**). The third group (7 proteins) localizes to the inner membrane complex (IMC)[35], a membranous system underlying the plasma membrane and involved in the structural integrity and motility of invasive parasites[35–37]. These proteins display a unique spatial dynamic during schizogony reflecting the biogenesis of this compartment (**Fig. 4h**,**i**, **Supplementary Fig. 6a** and **Supplementary Movies 1–3**). The remaining 11 proteins revealed localizations that are not obviously associated with invasion, although this does not exclude them from playing a role in this process (**Supplementary Fig. 6a**,**b**). Examples are proteins that localize to the cytosol including the putative kinase PFC0945w and the profilin homolog PFI1565w. In summary, 31 out of 42 selected proteins are associated with structures known to be directly involved in invasion. This demonstrates that the functional predictions based on our approach can lead to the identification of new putative targets for malaria intervention strategies.

DISCUSSION

Until now, the potential of using transcriptional profiling of growth perturbation for functional analyses of malaria parasites has been underappreciated. We demonstrate that functionally related genes share similar transcriptional profiles to a diverse panel of chemical perturbations, which suggests that many of these genes share regulatory mechanisms responsive to external stimuli (**Fig. 2a**). This suggests that transcriptional profiling may be a viable approach for functional genomics of human malaria parasites and can provide insights into parasite biology. Although mRNA decay was proposed to make a major contribution to the regulation of gene expression in *P. falciparum*[38], our data suggest that the responses to chemically induced growth perturbations are associated with transcription[39], rather than mRNA stability. We find essentially no relationship between our mRNA profiles and the previously established pattern of mRNA decay (data not shown).

The sensitivity of *P. falciparum* transcription to chemical stimuli has enabled us to make gene-function predictions not included in previous network-based approaches[10,11,40]. Our 90%-confidence network (termed PlasmoINT) contains close to 6 times more linkages and 2.5 times more proteins than PlasmoMAP[10], hitherto the most reliable published *P. falciparum* interaction network. In addition, there are five times as many linkages, which are supported by two or more types of evidence (**Supplementary Table 7**). These additions can be attributed mainly to the extensive transcriptional data and

inclusion of the annotations from the functional genomic database, the Malaria Parasite Metabolic Pathways[41]. This enables us to provide more accurate reconstructions of the majority of metabolic and cellular pathways (**Supplementary Fig. 7**) and thus more confident functional gene predictions. We also compared the Gene Ontology (GO) terms assigned to the *P. falciparum* genes by PlasmoINT with those assigned by the ontology-based pattern identification (OPI) method[40]. There is, however, only a limited congruity between these two studies with only 13%, 22% and 37% of the genes matching the predictions between the OPI and PlasmoINT-assigned GO terms at 4th, 3rd, and 2nd level, respectively. Although the relatively low level of consistency between these two methods is surprising, it is worth noting that the 47% recall precision of PlasmoINT contrasts (Online Methods and **Supplementary Fig. 4**), with only 18% precision for OPI. Similarly, the increased precision of the PlasmoINT prediction may result from the inclusion of the perturbation data set, which captures the finer pattern of transcriptional regulation in response to growth perturbation compared to the development stage–specific expression used by OPI. In addition to the supplementary material, the data presented in this manuscript have been compiled to a searchable database available online (http://zblab.sbs.ntu.edu.sg/), which we plan to update periodically.

As invasion of the host cell is essential for survival of *P. falciparum* and is a key target for new malaria intervention strategies, we used the functional annotations obtained from our interactome to experimentally validate proteins predicted to be associated with the invasion process. Of the 42 proteins that could be localized in the parasite, 31 were predominantly targeted either to the apical organelles, the parasite periphery or the IMC (**Fig. 4** and **Supplementary Fig. 6**): all key compartments for host cell invasion. Interestingly, 11 out of the 31 proteins contain neither a predicted signal peptide nor a transmembrane domain. Both of these are characteristic for proteins previously associated with the invasion machinery, highlighting the power of this approach. For instance, network prediction enabled us to identify novel proteins associated with the IMC such as MAL13P1.228, PF14_0578 or PFE1130w. This notion is further supported by the identification and localization of PFB0570w and PFD1105w, two proteins previously associated with the rhoptries, (exocytotic organelles containing many proteins with adhesive functions[42,43]), PF10_0348 and PF10_0352, two proteins of the merozoite surface protein super-family[44,45], and MAL13.P1.130 and PFD1110w, two newly localized IMC proteins[46,47]. Further confirmation of the utility of this study came from the identification and localization of PFD0230c. This unique serine protease was recently identified in a forward chemical genetic screen as one of the key regulators for merozoite egress[48]. Although it will be crucial to further validate these novel proteins and to extend their characterization, this subnetwork of proteins predicted to be involved in invasion offers a comprehensive blueprint of this process at the molecular level. These results may be useful for functional studies of each identified protein and rational drug and vaccine development.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturebiotechnology/.

Accession codes. Gene Expression Omnibus: GSE19468.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

This project was funded by the Academic Research Council of the Singapore Ministry of Education (grant no. ARC 11/05 M45080011), Singaporean National Medical Research Council (grant # IRG07Nov030) and the Deutsche Forschungsgemeinschaft (GI312 and GRK1459). The authors also thank B.D. Wastuwidyaningtyas and S. Tan for excellent technical assistance with the microarray experiments, K. Jurries for graphical assistance and A. Law, R. Stanway, T. Voss and H. Hoppe for critical reading of the manuscript. We are grateful to M. Blackman (National Institute for Medical Research, London) for providing the MSP-1 antibody, to P. Sharma (National Institute of Immunology, New Delhi) for providing the GAP45 antibody and to Jacobus Pharmaceuticals for providing WR99210.

AUTHOR CONTRIBUTIONS

G.H. and Z.B. performed the computation and data analysis. G.H., A.C., P.R.P., T.S., T.-W.G. and Z.B. drafted the paper. S.M., G.H., S.C. and B.K.C. performed the microarray experiments. A.C., M.K., S.H., K.E. and T.S. cloned the genes, generated the transgenic parasites and carried out the microscopy. All authors read and approved the final manuscript.

Published online at http://www.nature.com/naturebiotechnology/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Vitoria, M. *et al.* The global fight against HIV/AIDS, tuberculosis, and malaria: current status and future perspectives. *Am. J. Clin. Pathol.* **131**, 844–848 (2009).
2. Hay, S.I. *et al.* A world malaria map: Plasmodium falciparum endemicity in 2007. *PLoS Med.* **6**, e1000048 (2009).
3. Molyneux, D.H. Control of human parasitic diseases: Context and overview. *Adv. Parasitol.* **61**, 1–45 (2006).
4. Nwaka, S. & Hudson, A. Innovative lead discovery strategies for tropical diseases. *Nat. Rev. Drug Discov.* **5**, 941–955 (2006).
5. Gardner, M.J. *et al.* Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511 (2002).
6. Balu, B., Shoue, D.A., Fraser, M.J., Jr. & Adams, J.H. High-efficiency transformation of Plasmodium falciparum by the lepidopteran transposable element piggyBac. *Proc. Natl. Acad. Sci. USA* **102**, 16391–16396 (2005).
7. Maier, A.G. *et al.* Exported proteins required for virulence and rigidity of *Plasmodium falciparum*-infected human erythrocytes. *Cell* **134**, 48–61 (2008).
8. Winzeler, E.A. Applied systems biology and malaria. *Nat. Rev. Microbiol.* **4**, 145–151 (2006).
9. Kato, N. *et al.* Gene expression signatures and small-molecule compounds link a protein kinase to Plasmodium falciparum motility. *Nat. Chem. Biol.* **4**, 347–356 (2008).
10. Date, S.V. & Stoeckert, C.J. Jr. Computational modeling of the Plasmodium falciparum interactome reveals protein function on a genome-wide scale. *Genome Res.* **16**, 542–549 (2006).
11. Wuchty, S. & Ipsaro, J.J. A draft of protein interactions in the malaria parasite *P. falciparum*. *J. Proteome Res.* **6**, 1461–1470 (2007).
12. Bozdech, Z. *et al.* The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. *PLoS Biol.* **1**, E5 (2003).
13. Le Roch, K.G. *et al.* Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* **301**, 1503–1508 (2003).
14. Hughes, T.R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).
15. MacCarthy, T., Pomiankowski, A. & Seymour, R. Using large-scale perturbations in gene network reconstruction. *BMC Bioinformatics* **6**, 11 (2005).
16. Ganesan, K. *et al.* A genetically hard-wired metabolic transcriptome in *Plasmodium falciparum* fails to mount protective responses to lethal antifolates. *PLoS Pathog.* **4**, e1000214 (2008).
17. Gunasekera, A.M., Myrick, A., Le Roch, K., Winzeler, E. & Wirth, D.F. *Plasmodium falciparum*: genome wide perturbations in transcript profiles among mixed stage cultures after chloroquine treatment. *Exp. Parasitol.* **117**, 87–92 (2007).
18. Oakley, M.S. *et al.* Molecular factors and biochemical pathways induced by febrile temperature in intraerythrocytic Plasmodium falciparum parasites. *Infect. Immun.* **75**, 2012–2025 (2007).
19. Natalang, O. *et al.* Dynamic RNA profiling in *Plasmodium falciparum* synchronized blood stages exposed to lethal doses of artesunate. *BMC Genomics* **9**, 388 (2008).
20. Tamez, P.A. *et al.* An erythrocyte vesicle protein exported by the malaria parasite promotes tubovesicular lipid import from the host cell surface. *PLoS Pathog.* **4**, e1000118 (2008).
21. Groth, P., Weiss, B., Pohlenz, H.D. & Leser, U. Mining phenotypes for gene function prediction. *BMC Bioinformatics* **9**, 136 (2008).
22. Kim, W.K., Krumpelman, C. & Marcotte, E.M. Inferring mouse gene functions from genomic-scale data using a combined functional network/classification strategy. *Genome Biol.* **9** Suppl 1, S5 (2008).
23. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**, D277–D280 (2004).
24. Lee, H., Deng, M., Sun, F. & Chen, T. An integrated approach to the prediction of domain-domain interactions. *BMC Bioinformatics* **7**, 269 (2006).
25. LaCount, D.J. *et al.* A protein interaction network of the malaria parasite Plasmodium falciparum. *Nature* **438**, 103–107 (2005).

26. Llinas, M., Bozdech, Z., Wong, E.D., Adai, A.T. & DeRisi, J.L. Comparative whole genome transcriptome analysis of three Plasmodium falciparum strains. *Nucleic Acids Res.* **34**, 1166–1173 (2006).
27. Mackinnon, M.J. *et al.* Comparative transcriptional and genomic analysis of Plasmodium falciparum field isolates. *PLoS Pathog.* **5**, e1000644 (2009).
28. Barabasi, A.L. & Oltvai, Z.N. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004).
29. Krogan, N.J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).
30. Deng, M., Zhang, K., Mehta, S., Chen, T. & Sun, F. Prediction of protein function using protein-protein interaction data. *J. Comput. Biol.* **10**, 947–960 (2003).
31. Cowman, A.F. & Crabb, B.S. Invasion of red blood cells by malaria parasites. *Cell* **124**, 755–766 (2006).
32. Soldati, D., Foth, B.J. & Cowman, A.F. Molecular and functional aspects of parasite invasion. *Trends Parasitol.* **20**, 567–574 (2004).
33. Haase, S. *et al.* Characterization of a conserved rhoptry-associated leucine zipper-like protein in the malaria parasite *Plasmodium falciparum*. *Infect. Immun.* **76**, 879–887 (2008).
34. Treeck, M. *et al.* A conserved region in the EBL proteins is implicated in microneme targeting of the malaria parasite *Plasmodium falciparum*. *J. Biol. Chem.* **281**, 31995–32003 (2006).
35. Baum, J. *et al.* A conserved molecular motor drives cell invasion and gliding motility across malaria life cycle stages and other apicomplexan parasites. *J. Biol. Chem.* **281**, 5197–5208 (2006).
36. Baum, J. *et al.* A malaria parasite formin regulates actin polymerization and localizes to the parasite-erythrocyte moving junction during invasion. *Cell Host Microbe* **3**, 188–198 (2008).
37. Morrissette, N.S. & Sibley, L.D. Disruption of microtubules uncouples budding and nuclear division in Toxoplasma gondii. *J. Cell Sci.* **115**, 1017–1025 (2002).
38. Shock, J.L., Fischer, K.F. & DeRisi, J.L. Whole-genome analysis of mRNA decay in *Plasmodium falciparum* reveals a global lengthening of mRNA half-life during the intra-erythrocytic development cycle. *Genome Biol.* **8**, R134 (2007).
39. De Silva, E.K. *et al.* Specific DNA-binding by apicomplexan AP2 transcription factors. *Proc. Natl. Acad. Sci. USA* **105**, 8393–8398 (2008).
40. Zhou, Y. *et al.* Evidence-based annotation of the malaria parasite's genome using comparative expression profiling. *PLoS One* **3**, e1570 (2008).
41. Ginsburg, H. Caveat emptor: limitations of the automated reconstruction of metabolic pathways in *Plasmodium*. *Trends Parasitol.* **25**, 37–43 (2009).
42. Chattopadhyay, R. *et al.* PfSPATR, a *Plasmodium falciparum* protein containing an altered thrombospondin type I repeat domain is expressed at several stages of the parasite life cycle and is the target of inhibitory antibodies. *J. Biol. Chem.* **278**, 25977–25981 (2003).
43. Wickramarachchi, T., Devi, Y.S., Mohmmed, A. & Chauhan, V.S. Identification and characterization of a novel *Plasmodium falciparum* merozoite apical protein involved in erythrocyte binding and invasion. *PLoS ONE* **3**, e1732 (2008).
44. Pearce, J.A., Mills, K., Triglia, T., Cowman, A.F. & Anders, R.F. Characterisation of two novel proteins from the asexual stage of *Plasmodium falciparum*, H101 and H103. *Mol. Biochem. Parasitol.* **139**, 141–151 (2005).
45. Wickramarachchi, T. *et al.* A novel *Plasmodium falciparum* erythrocyte binding protein associated with the merozoite surface, PfDBLMSP. *Int. J. Parasitol.* **39**, 763–773 (2009).
46. Bullen, H.E. *et al.* A novel family Of apicomplexan glideosome associated proteins With an inner-membrane anchoring role. *J. Biol. Chem.* **284**, 25353–25363 (2009).
47. Rayavara, K. *et al.* A complex of three related membrane proteins is conserved on malarial merozoites. *Mol. Biochem. Parasitol.* **167**, 135–143 (2009).
48. Arastu-Kapur, S. *et al.* Identification of proteases that regulate erythrocyte rupture by the malaria parasite *Plasmodium falciparum*. *Nat. Chem. Biol.* **4**, 203–213 (2008).

ONLINE METHODS

Parasite culture, treatment and microarray. The perturbation time courses were performed with 2% hematocrit and 5% parasitemia cultures. Parasites were treated with appropriate drug or compound concentrations and collected at 5–8 time points taken at regular time intervals (30–120 min). A total of 247 microarray experiments were carried out, including 29 drug treatment time courses with 20 compounds and corresponding untreated controls from different drug or inhibitor treatment (**Supplementary Table 1**). Genome-wide gene expression profiling was conducted using long oligonucleotides representing all 5,363 *P. falciparum* genes as previously described[49]. The expression data were normalized using linear normalization and background filtering as implemented by the NOMAD database (http://derisilab.ucsf.edu) and described[12]. Subsequently each gene profile was represented by an average expression value calculated as an average of all oligonucleotides representing a particular gene. For the final data set we considered only the genes for which at least 80% of time points in each time course yielded a positive expression signal.

For the final microarray input data sets for the reconstruction of the gene functional network, we incorporated the perturbation data set with the IDC transcriptome of laboratory strains (3D7, Dd2 and HB3, 148 microarray experiments)[12] and four lab isolates[27]. To indicate the strength of functional association of each gene pair by gene expression profiles, PCCs were calculated independently across each data set first and intergraded by a new technique that we term the "optional average" method. Briefly, Fisher's z-transform[50] was used to average two PCCs from two independent IDC transcriptomes and compared to the PCC from perturbation data. If the latter is smaller, the final PCC is the PPC from perturbation data. Otherwise, the final PCC is equal to the average PCC from two tested data sets defined by the Fisher's z-transform.

The input data sets for the network construction. For the network assembly we incorporate the microarray data set (above) with three additional inputs. (i) The phylogenetic profiles were calculated for all *P. falciparum* genes obtained from the PlasmoDB version 5.4 (http://www.plasmodb.org/download/). Using BLASTP, the protein sequences of *P. falciparum* were compared with 210 reference organisms, including 155 prokaryotes and 55 eukaryotes available from the NCBI and the ENSEMBL. For each protein a vector was generated with elements p_{ij} where $p_{ij} = -1/\log E_{ij}$ where Eij represents the E-value of the gene (i) ortholog in the genome (j). As a metric of phylogenetic profile similarity, the mutual information was calculated with the histograms of p_{ij} values, binned in 0.01 intervals, as previously described[51]. The mutual information scores were divided into 15 bins for the KEGG benchmark test (**Supplementary Table 3**). (ii) For the domain-domain interaction evidence, we carried out Hidden Markov Model–based predictions of all functional domains defined by the PFAM database in all 5,363 *P. falciparum* proteins. For this we use the set of domain-domain interactions as defined previously[24]. Based on the confidence scores provided by the Lee database[24], the gene pairs were subsequently divided into six bins and tested against the KEGG benchmark. (iii) From the yeast two-hybrid system protein-protein interactions were obtained from the previous publication[25] and all 2,811 interactions among 1,308 *P. falciparum* proteins were tested against the KEGG benchmark as one bin (**Fig. 2b**).

Calculation of the likelihood scores using the KEGG gold standard benchmark data set. The KEGG 'gold standard' benchmark data set includes 492 annotated *P. falciparum* genes that can be assigned to 71 metabolic or cellular pathways defined by the KEGG database[23]. This defines 11,046 positive pairs of genes that belong to pathways with >3 genes. The negative set includes 61,721 gene pairs that do not fall into a common pathway. **Supplementary Table 3** online shows the parameters of naive Bayesian network of all data sets based on this reference data set. The ratio of true to false positive in **Figure 2c** is calculated using the KEGG benchmark data set and it reflects measure of agreement of the functional relationship of each gene pair as a function of the individual scoring systems (e.g., PCC for microarray data and phylogenetic profiling). The calculated likelihood scores reflect the functional relationships between *P. falciparum* genes and are applicable as input values for assembling a probabilistic interactome network.

Building the interaction network Integration of the data sets by the Bayesian probabilistic model was carried out as previously described[10]. In principle, the final likelihood score is determined as:

$$\text{Likelihood Score (LS)} = LS_{PPC} \times LS_{PHY} \times LS_{PPI} \times LS_{Domain}$$

PPC, microarray input; *PHY*, phylogenetic profile input; *PPI*, yeast two-hybrid input; *Domain*, domain-domain interaction input.

We performed a tenfold cross-validation to evaluate the overall performance of the prediction. Briefly, first the positive and negative benchmarks were randomly divided into ten separate equal sets, and nine of them were used as the training set to calculate the likelihood scores and the remaining one set as the test to identify the positives and negatives. We ran this process ten times so that each of the ten sets was a test set and the remaining nine constituted the training set. Finally, all true positives (TP) and false positives (FP) were summed up under different likelihood score cutoffs to evaluate the ratio of true positives to false positives. The positive predictive values (PPV=TP/(TP+FP)) were calculated as the fraction of true positives to the total number of true positive and false positive (**Fig. 2d**).

The modular analysis and the weighted neighbor counting for network-based gene function prediction. We searched the local modules in the network using the Markov Cluster (MCL) algorithm, which is a fast and scalable unsupervised graph clustering algorithm[52]. To define the parameter of granularity, we followed a previously published method[53] by optimizing the functional coherence and size of the clusters[54]. The networks and subnetworks were designed and visualized using Cytoscape 2.5 (ref. 55).

The neighbor-counting method weighted by the likelihood score was used for the functional gene predictions in which the likelihood score of each linkage could represent the functional similarity between two proteins:

$$f(i,j) = \sum LS(m)\delta(j) / \sum LS(m)$$

where the f(i,j) is the probability of gene i having function j. The LS(m) is the likelihood score of the m^{th} neighbor of gene i. $\delta(j) = 1$ if the gene has function j, else $\delta(j) = 0$. Without threshold, we assigned an unannotated protein with *k* functions having the top *k* statistic scores. The performance of the predictions were evaluated by plotting precision against recall over various thresholds as described[56]. For a given threshold, precision and recall are defined as:

$$\text{Precision} = \sum_i^V k_{i,\beta} / \sum m_{i,\beta} \qquad \text{Recall} = \sum_i^V k_{i,\beta} / \sum n_i$$

where n_i is the number of known functions of protein i; $m_{i,\beta}$ is the number of functions predicted for protein i at threshold β and $k_{i,\beta}$ is the number of functions predicted correctly for protein i . V is the set of all functionally known genes.

DNA constructs, transfection and intracellular localizations. PCR amplification for the GFP constructs was carried out using cDNA with the gene-specific primers summarized in **Supplementary Table 8**. PCR products were digested with KpnI and AvrII and ligated into the transfection vector $pARL_{ama\text{-}1}$-GFP[34]. To avoid cytotoxic effects due to overexpression of the putative proteases, only 1 kb N-terminal fragments of PF08_0108 and PFD0230c were cloned. To ensure late expression, the promoter of the *ama-1* gene was used to drive transcription. *P. falciparum* asexual stages (3D7) were transfected as described previously[57]. Positive selection for transfectants was achieved using 10 nM WR99210.

The western blot analyses were carried out as previously described[58] using the mouse anti-GFP (1:1000, Roche) and sheep anti-mouse IgG horseradish peroxidase (1:3000, Roche). Images of unfixed GFP-expressing parasites were captured using a Zeiss Axioskop 2plus microscope with a Hamamatsu Digital camera (ORCA C4742-95) using Zeiss axiovision software. Immunofluorescence microscopy was performed on 4% formaldehyde/0.0075% glutaraldehyde-fixed parasites incubated for 1 h with primary antibodies in the following dilutions: rabbit anti-MSP-1 (1:2,000), rabbit anti-GAP45 (1:2,000) and rabbit anti-EBA-175 (1:2,000). Subsequently, cells were incubated with Alexa-Fluor 594 goat anti-rabbit IgG or Alexa-Fluor 488 goat anti-mouse IgG antibodies (1:2,000, Molecular Probes) and with DAPI at 1 μg/ml (Roche).

49. Hu, G., Llinas, M., Li, J., Preiser, P.R. & Bozdech, Z. Selection of long oligonucleotides for gene expression microarrays using weighted rank-sum strategy. *BMC Bioinformatics* **8**, 350 (2007).
50. Huttenhower, C., Hibbs, M., Myers, C. & Troyanskaya, O.G. A scalable method for integration and functional analysis of multiple microarray data sets. *Bioinformatics* **22**, 2890–2897 (2006).
51. Date, S.V. & Marcotte, E.M. Discovery of uncharacterized cellular systems by genome-wide analysis of functional linkages. *Nat. Biotechnol.* **21**, 1055–1062 (2003).
52. Krogan, N.J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).
53. Wuchty, S. & Ipsaro, J.J. A draft of protein interactions in the malaria parasite *P. falciparum*. *J. Proteome Res.* **6**, 1461–1470 (2007).
54. Lee, I., Date, S.V., Adai, A.T. & Marcotte, E.M. A probabilistic functional network of yeast genes. *Science* **306**, 1555–1558 (2004).
55. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
56. Deng, M., Zhang, K., Mehta, S., Chen, T. & Sun, F. Prediction of protein function using protein-protein interaction data. *J. Comput. Biol.* **10**, 947–960 (2003).
57. Fidock, D.A. & Wellems, T.E. Transformation with human dihydrofolate reductase renders malaria parasites insensitive to WR99210 but does not affect the intrinsic activity of proguanil. *Proc. Natl. Acad. Sci. USA* **94**, 10931–10936 (1997).
58. Struck, N.S. *et al.* Re-defining the Golgi complex in *Plasmodium falciparum* using the novel Golgi marker PfGRASP. *J. Cell Sci.* **118**, 5603–5613 (2005).

doi:10.1038/nbt.1597