

Genes involved in host–parasite interactions can be revealed by their correlated expression

Adam James Reid* and Matthew Berriman

Parasite genomics group, Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK

Received May 30, 2012; Revised November 26, 2012; Accepted December 1, 2012

ABSTRACT

Molecular interactions between a parasite and its host are key to the ability of the parasite to enter the host and persist. Our understanding of the genes and proteins involved in these interactions is limited. To better understand these processes it would be advantageous to have a range of methods to predict pairs of genes involved in such interactions. Correlated gene expression profiles can be used to identify molecular interactions within a species. Here we have extended the concept to different species, showing that genes with correlated expression are more likely to encode proteins, which directly or indirectly participate in host–parasite interaction. We go on to examine our predictions of molecular interactions between the malaria parasite and both its mammalian host and insect vector. Our approach could be applied to study any interaction between species, for example, between a host and its parasites or pathogens, but also symbiotic and commensal pairings.

INTRODUCTION

Understanding the molecular mechanisms by which hosts and parasites interact is key to understanding how parasites subvert host immune defenses and cause disease. Unfortunately, we currently know little about which gene products interact between hosts and parasites. Large interaction datasets and a range of prediction methods exist for examining protein–protein interactions (PPIs) within a particular organism (1); however, few data exist on PPIs between interacting species such as parasites and their hosts. To fill this gap, there have been several bioinformatic attempts to identify or predict host–parasite protein–protein interactions (HPPPIs). These approaches are all based in some way on inferring HPPPIs from known intraspecific PPIs by homology (2–4). This interologues (interaction homologue) approach infers

that if a pair of proteins is known to interact, then homologues of these proteins are also likely to interact. This approach is limited because it is unable to make inferences about interactions between species-specific gene families. These include many well-known HPPPIs, such as those involving parasite- and host-specific cell-surface receptors, for example, PfEMP1 and ICAM-1 or Rh5 and basigin, in human malaria (5,6).

PPI prediction methods are prone to high false-positive rates, and a common approach to dealing with this is to integrate multiple lines of evidence (7–9). Likewise, multiple independent approaches to predicting HPPPIs are needed. There are a variety of existing approaches to predict interactions between proteins within an organism, in addition to those that use interologues (10). One that could feasibly be applied to HPPPIs is that introduced by Grigoriev (11) and Ge *et al.* (12) using correlated gene expression. The products of genes with highly correlated expression profiles are more likely to be involved in the same protein complex or pathway than expected by chance (12,13). This work was initially done in yeast, but it has since been used across a wide variety of species (10). Within an organism, gene products involved in the same complex or pathway must be present at the same time, and thus the genes can be reasonably expected to have correlated expression profiles. We hypothesized that this approach would also work between species. One might expect that, commonly, expression changes in a parasite, which lead to an insult to its host, will result in expression changes in the host to counter the challenge or vice versa. This approach has previously been applied to the detection of correlated expression profiles between macaque and pathogenic *Streptococcus spp.* but was not formally tested (14). A recent approach used simultaneous gene expression data and regulatory network modelling to examine host–parasite interactions between the yeast *Candida albicans* and mouse (15). Although this approach was effective in identifying a limited number of verifiable interactions, it was restricted to a limited set of genes with previous knowledge of involvement in host–parasite interaction.

*To whom correspondence should be addressed. Tel: +44 1223 494810; Fax: +44 1223 494919; Email: ar11@sanger.ac.uk

Using gene expression data from a murine malaria parasite species, its host and vector, we found that pairs of host–parasite genes with correlated expression profiles were enriched for those involved in HPPPIs. Furthermore, we found evidence that although many well-correlated pairs may not interact physically, they tend to act upstream of HPPPIs. This suggests that the detected relationships encompass interacting molecular subsystems and not simply directly interacting proteins. We have named this approach Inter-Species Interactions using Gene Expression Measurements (ISIGEM). Unlike previous methods used to predict HPPPIs, our approach does not rely on known interactions between homologues in other organisms. This is key because it is often the products of species-specific genes (surface antigens, effectors etc) that are at the forefront of host–parasite interactions. This approach could be applied to any parasitic, pathogenic, commensal or symbiotic relationship for which simultaneous gene expression measurements can be taken.

METHODS

Reanalysis of Lovegrove microarray dataset

The probe sequences from the Lovegrove *et al.* (16) hybrid microarray were remapped to more recent versions of the *Plasmodium berghei* (July 2011; <http://www.genedb.org>) and mouse (NCBIM37; <http://ensembl.org>) genome assemblies using SMALT (Ponstingl, unpublished; <http://www.sanger.ac.uk/resources/software/smalt/>).

Reads mapping to multiple locations were excluded, including those mapping to both genomes. The probe sequences were downloaded from the Lovegrove *et al.* project website (http://hugheslab.ccb.utoronto.ca/supplementary-data/malaria/extra_material/GEO_info_allProbes.txt). Of the 42 034 probe sequences, 7295 reads were reliably mapped to 3849 *P. berghei* genes, whereas 4077 reads were mapped to 3480 mouse genes. Where multiple probes mapped to a single gene, the mean intensity was taken for each time point. We used per-probe normalized BALB/c mouse intensity profiles as used in (16) and kindly provided by L. Peña-Castillo.

Gene Ontology (GO) terms from GeneDB (17) were identified for 1551 *P. berghei* genes and from Ensembl (18) for 3211 genes. Pfam annotations (19) were also extracted from GeneDB for *P. berghei* and from Ensembl for mouse.

Reanalysis of Xu microarray dataset

The microarray used by Xu *et al.* (20) was designed using probes based on several Expressed Sequence Tag (EST) libraries that we downloaded from Genbank: ookinete library, *P. berghei* (LIBEST_012759); early-oocyst library, *Anopheles stephensi*/*P. berghei* mixed EST library (LIBEST_012760); mid-oocyst subtraction library, *A. stephensi*/*P. berghei* mixed EST library (LIBEST_012900); and late-oocyst subtraction library, *A. stephensi*/*P. berghei* mixed EST library (LIBEST_012901). We remapped these ESTs to current versions of the *P. berghei* and *Anopheles gambiae* genomes (the *A. stephensi* genome is currently not available). To get

the best hit to either genome, and account for the evolutionary distance between *A. stephensi* and *A. gambiae*, we used BLASTX to map the ESTs to predicted protein sequences (ESTs corresponding to the untranslated regions of *A. stephensi* genes were therefore excluded). Individual ESTs with E-values $<10^{-20}$ were accepted, except for those with multiple hits with equivalent E-values. Of the 4402 gene expression profiles produced by Xu *et al.*, not all had been associated with sequenced probes and therefore could not be included in our analysis. We were able to associate 316 profiles with 287 genes: 67 from *Plasmodium* and 220 from *Anopheles*. Where multiple profiles matched to a single gene, the mean value was taken for each time point.

We were able to identify GO terms for 40 of the *Plasmodium* genes and 113 of the *Anopheles* genes. Pfam domain annotation and GO terms for *A. gambiae* were retrieved from Ensembl Metazoa (21).

An integrated benchmark dataset of HPPPIs

Correlated sequence signature benchmark datasets

Dyer *et al.* (3) adapted a correlated sequence signature (CSS) approach, developed by Sprinzak and Margalit (22), to predict PPIs between human and *Plasmodium* using protein domains overrepresented in known intraspecific interactions. They identified protein domain pairs, which tend to occur in experimentally identified PPIs within mouse or *Plasmodium*, and used these domain pairs to predict PPIs between mouse and *Plasmodium*. Using domains rather than sequence similarity between whole proteins is more powerful because more distant evolutionary relationships can be detected. Furthermore, domain-mediated interactions between multidomain protein families not present in the known interaction dataset can be revealed. Dyer *et al.* showed using several tests that these predicted HPPPIs were enriched in likely true interactions. They showed that where a host protein was predicted to interact with two parasite proteins, the two parasite proteins tended to be close to each other in the parasite protein interaction network and furthermore that the genes relating to those proteins tended to be co-expressed. This suggests that the method identifies consistent relationships between a host protein and a particular subsystem within the parasite. Furthermore, they showed that those parasite proteins predicted to interact with host proteins were enriched for functional annotation relevant to host–parasite interaction. Both the CSS approach and the tests of the method used by Dyer *et al.* are independent from those used by us, making it a suitable benchmark. Importantly, it can be applied to any host–parasite system, producing a relatively large set of predictions to use as true-positives.

To generate such a benchmark dataset for assessing the accuracy of mouse–malaria HPPPI predictions, we trained our implementation of the Dyer algorithm using 3928 *Plasmodium falciparum* ‘direct complex’ interactions from Reactome (23) among 275 proteins and Pfam domain predictions from GeneDB (24). These pairs of interacting Pfam domains and their associated scores were used to predict interacting pairs of genes between

malaria and mouse (Lovegrove dataset) and malaria and mosquito (Xu dataset). Using a cutoff of $\text{Pr} \geq 0.5$, as suggested by Dyer *et al.* (3), we identified 52 083 interactions in the Lovegrove dataset and 12 192 such interactions in the Xu dataset.

This dataset is large and likely to be rich in false-positives. Therefore, to improve its quality, we wanted to exclude CSS interactions that were unlikely to be real. To achieve this, we excluded interactions where the proteins were not predicted to be present outside of the cell in a similar manner to Krishnadev and Srinivasan (25). A list of *P. berghei* genes predicted to be exported, to have signal peptides or transmembrane domains was downloaded from PlasmoDB (26). For mouse, we used the LOCATE subcellular localization database (27) and extracted all proteins predicted to be secreted or membrane-bound. This reduced an initial 52 083 CSS interactions to 241. We refer to this as the filtered CSS benchmark dataset. It was not possible to build a similarly filtered CSS interaction benchmark for the Xu dataset, owing to its small size.

Vignali yeast two-hybrid dataset

Vignali *et al.* (28) used a modified yeast two-hybrid methodology to identify interactions between *P. falciparum* proteins and those of its human host. They identified 456 interactions after filtering their data to reduce likely false-positives. This is the only published dataset that describes PPIs between *Plasmodium* and a host using high-throughput methods. We used *Plasmodium* orthology information from PlasmoDB (26) and human–mouse orthology information from Ensembl (18) to convert *P. falciparum*–human interactions into their *P. berghei*–mouse interologues. Of the 456 interactions identified, we were able to determine 274 interologues between *P. berghei* and mouse genes, 51 of which were represented by gene pairs on the Lovegrove microarray.

An integrated benchmark dataset

For benchmarking the performance of ISIGEM on the Lovegrove dataset, we used a combination of the Vignali interactions and the filtered CSS interactions. We refer to this as the integrated benchmark dataset.

Testing

To determine whether there was a link between HPPPIs and gene expression profile correlations, we used the Kolmogorov–Smirnov (KS) test, implemented in *R*. This non-parametric test is used to determine whether a sample distribution is likely to have been drawn from a reference distribution. We compared either the distribution of Pearson r for randomized profiles with that of the true-positive profile pairs or the distribution of the ISIGEM score with that of true-positive profile pairs. For Pearson r comparisons, we used the alternative hypothesis that the truly interacting genes have greater values of r than those of randomized profiles. For ISIGEM score comparisons, we used the alternative hypothesis that the truly interacting genes have lower scores than those of randomized profiles.

Identification of correlated expression patterns

Initially, we wanted to determine whether gene expression profiles between species were better correlated than expected by chance. We compared the distribution of Pearson r between *P. berghei* and mouse profiles with those for the same set of profiles where the order of the time points within a profile was randomized. We chose Pearson correlation to compare gene expression profiles rather than Euclidean distance because magnitude is unimportant with Pearson. This allows genes expressed at low and high levels to correlate if they follow the same pattern. Spearman rank correlation was not used, as integer ranks are less sensitive than continuous numbers.

We found that some probe pairs were highly correlated for trivial reasons, for example, where two genes were constitutively expressed. Therefore, it was not sufficient to score the gene pairs based simply on Pearson r . Instead, we determined the significance of each correlation using empirically derived P -values. Genes expected to correlate well by chance, due to uncomplicated profiles, should have high P -values, whereas genes that are unlikely to correlate by chance should have low P -values. To achieve this, we randomized all profiles and calculated the Pearson correlation coefficient between every interspecific pair of genes. We repeated this 10^5 times and calculated the P -value as the number of times a randomized pair of genes profiles were correlated at least as well as the real profiles, divided by 10^5 . We did not correct the P -values for multiple hypothesis testing, as, owing to the large number of tests ($\sim 1.3 \times 10^7$ host–parasite gene pairs), it was not possible to perform sufficient numbers of randomizations to achieve significant corrected P -values using the computational resources available. Thus, the P -values we use should not be interpreted as representing statistical probabilities. They remain useful, however, for ranking the host–parasite gene pairs—down-weighting spuriously correlated pairs.

Co-localization of ISIGEM predictions and benchmark interactions in functional clusters

We wanted to determine whether ISIGEM predictions, although often not direct interactions, were likely indirectly interacting gene products. We reconstructed intra-specific networks of functionally associated gene products using the STRING 9.0 database (1). STRING data for *P. berghei* and *A. stephensi* were derived by orthology from that for *P. falciparum* and *A. gambiae*, respectively. We then clustered each network using the MCL (29) with default parameters.

For each of our top-scoring ISIGEM predictions, we then determined whether the genes were found in clusters also containing interactions from the integrated benchmark dataset, for example, in an ISIGEM prediction involving a mouse gene occurring in cluster m1 and a *Plasmodium* gene in cluster p1, we looked for a benchmark interaction with a mouse gene in cluster m1 and a *Plasmodium* gene in cluster p1.

We compared the frequency of co-localizations with that expected from random predictions. We took random pairs of mouse and *Plasmodium* genes (with

replacement) and looked to see whether they co-localized with benchmark interactions. We did this for the same number of interactions as identified by ISIGEM and repeated this 1000 times.

Functional analysis of interacting genes

We performed two types of functional analysis, both using GO functional annotation. Firstly, we wanted to examine functional terms enriched among all genes in one species (e.g. host) predicted to interact with genes from another species (e.g. parasite). To do this, we used TopGO (30) and used the *weight01* algorithm and the Fisher statistic with a *P*-value cutoff of 0.05.

In the second analysis, we wanted to identify pairs of functions, which were enriched in predicted interactions, one term from each species. To do this, we developed a method to identify GO term pairs found in interacting genes, which occur more often than expected by chance in a similar fashion to Dyer *et al.* (3). In each case, we considered only those genes for which expression could be detected, which also had GO terms associated with them. For each term associated with a gene, we added all its ancestral terms. We then counted the number of times a pair of terms *a*, *b* occurred such that *a* was associated with a host gene and *b* was associated with a parasite gene with which it was predicted to interact. We then randomized the interactions using the Fisher–Yates shuffle and counted the number of occurrences of *a* with *b* among these randomized gene pairs. The randomization step was repeated 1000 times, and *P*-values for each GO term pair were calculated empirically. Many of the GO term pairs thus identified were redundant in the sense that other pairs were more specific, although both pairs might have the same *P*-value. For example, where pairs *a*₁, *b*₁ and *a*₂, *b*₂ are identified and *a*₁ is an ancestor of *a*₂ and *b*₁ is an ancestor of *b*₂, then *a*₂, *b*₂ is a more specific pair of terms. To reduce such redundancy within the results, less specific term-pairs were removed based on the product of their depths in the GO graph. Depth in the graph is roughly correlated with specificity of terms, as the most general node is the root and each child term is a more specific instance of its parent, for example, *nucleotide triphosphate biosynthetic process* is a child of *nucleotide biosynthetic process*. Because a node may have multiple parents, there are often multiple pathways between a node and the root, and therefore, there may be different numbers of ancestral nodes depending on the pathway chosen. We therefore calculated the depth of a particular node as the mean of all possible paths to the root. The GO term pairs were sorted by *P*-value from lowest to highest and then by the product of the depth of the GO terms. Thus, for a particular *P*-value, we considered the most specific pair of terms first. We then excluded a pair if either parents or children of both terms had already been seen together. *P*-values were corrected for multiple hypothesis testing using the Benjamini–Hochberg method implemented in the R program *p.adjust*.

RESULTS

Pairs of host–parasite genes with correlated expression are enriched for those involved in host–parasite interactions

Within a single organism, the products of gene pairs with well-correlated gene expression profiles are known to be enriched in interacting proteins (12,13,31). We investigated the extent to which correlated gene expression is also predictive of PPIs between genomes, such as a parasite and its host. To do this, we required a simultaneous gene expression time-course experiment where gene expression is measured at the same time points in interacting species. A small number of existing studies have generated such data, encompassing several different host–pathogen pairs (16,20,32–35). Two of these studies investigated the rodent malaria model *P. berghei* using a combined microarray to examine both the parasite and murine host (16), or parasite and mosquito vector (20), respectively.

We initially used the gene expression data for *P. berghei* and its host *Mus musculus* generated by Lovegrove *et al.* (16). The dataset comprised four independent time-courses, each measuring gene expression in a different mouse tissue over three time points. We concatenated these four experiments to produce a series with 12 data points for each host/parasite gene and calculated correlation coefficients between the expression profiles for all host–parasite gene pairs. We hypothesized that well-correlated pairs were more likely to represent genes whose products interact.

Figure 1A shows the distribution of expression profile correlation values between pairs of mouse and *Plasmodium* genes (in black). These tend to be higher than expected by chance (KS test $D^+ = 0.0415$; *P*-value $< 2.2e-16$), as shown by the distribution of correlations between pairs of profiles with randomized time points (in grey). This suggests that there is a signal in the correlated profiles. We noticed, however, that there were many spurious positive correlations between simple profiles, for example, genes constitutively expressed in both species. To account for these spurious correlations, we calculated empirical *P*-values for gene pairs based on the likelihood of observing a correlation between their profiles by chance. Thus, instead of using the correlation coefficient to score pairs of profiles, we instead used a score that expressed the chance of observing the correlation. We refer to this value as the ISIGEM score, where 0 represents the best match and 1 the worst.

To examine whether this signal relates to HPPPIs, a large number of confirmed HPPPIs between *P. berghei* and mouse would ideally be required. In the absence of such benchmarking data being available, we combined multiple sources of inferred interactions. Firstly, we examined a dataset of 456 experimentally determined host–parasite interactions between *P. falciparum* and human (28). To use this dataset, it was necessary to infer orthologous interactions (interologues) between *P. berghei* and mouse. We were able to infer 279 interologues; however, only 51 involved gene pairs where both genes were found in the expression dataset. To enrich this

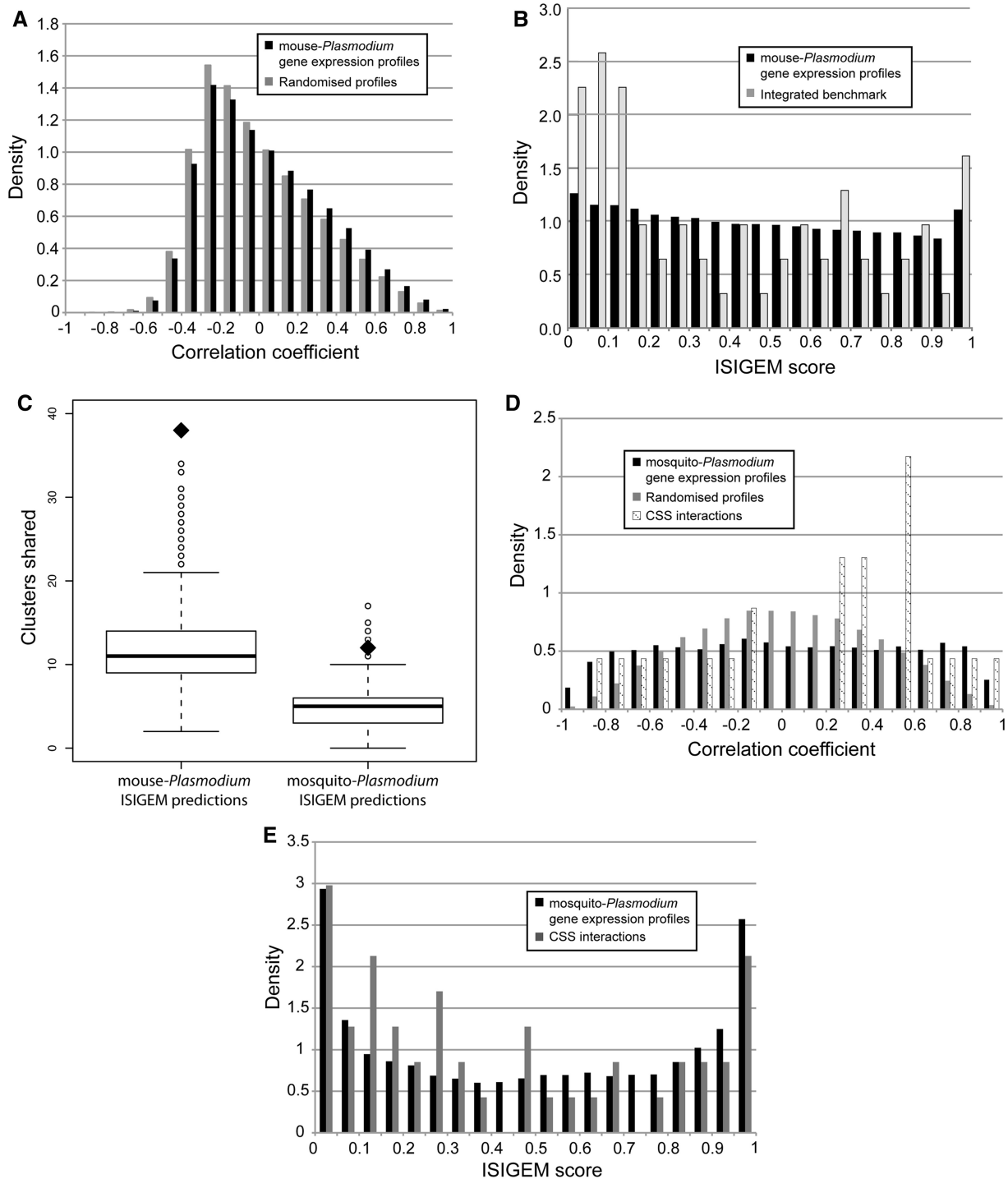


Figure 1. ISIGEM predictions with low *P*-values are enriched in host-parasite protein-protein interactions. (A) Pairs of host-parasite genes in the mouse-*Plasmodium* dataset are better correlated than expected by chance. (B) Using the integrated benchmark dataset, we reveal a clear enrichment of HPPPIs among mouse-*Plasmodium* gene pairs with low ISIGEM scores. (C) ISIGEM predictions with the lowest scores lie closer to HPPPIs in functional association networks than expected by chance for both mouse-*Plasmodium* and mosquito-*Plasmodium* datasets. Filled diamonds represent observed number of predictions, which share STRING clusters with benchmark interactions. Box and whisker plots represent the distribution of expected values calculated using randomized predictions. (D) Host-parasite gene pairs between malaria and its insect vector are better correlated than expected by chance in both positive and negative directions. There is a shift towards a more positive correlation coefficient among true-positive (CSS) gene pairs, although this is not statistically significant. (E) Mosquito-*Plasmodium* gene pairs with low ISIGEM *P*-values are not significantly enriched in HPPPIs.

dataset, we used an interologue-based HPPPI prediction method, which we then filtered to improve reliability.

The CSS approach has been used previously to identify HPPPIs by inferring interaction through homology (3). Firstly, protein domain pairs are identified that are overrepresented in known intraspecific PPIs. These are then used to infer HPPPIs between proteins with homologous sequence signatures (in this case, protein domains). This approach was shown to be successful in identifying HPPPIs in malaria (3). We identified Pfam domain pairs that occurred more often than expected by chance in interacting protein pairs from *P. falciparum* (23). We then looked for pairs of proteins, one each from *P. berghei* and mouse, which contained one each of these domains. These 52083 CSS interactions were then filtered to improve their reliability; we retained only those interactions where both gene products were predicted to be exposed, that is, exported or membrane-bound. This resulted in a reduced set of 241 more reliable interactions. We combined the interologues from the Vignali dataset with these filtered CSS interactions to produce our integrated benchmark dataset of 520 interactions. Of these, 62 pairs of genes were found on the Lovegrove microarray and could be used to benchmark our approach.

We then tested whether host–parasite gene pairs from the integrated benchmark dataset tended to have lower ISIGEM scores than other pairs from the mouse–*Plasmodium* dataset. We found that they did (KS test $D^- = 0.19$, P -value = 0.01; Figure 1B). This suggests that the ISIGEM score and therefore gene expression correlation is associated with pairs of genes involved in host–parasite interaction in rodent malaria.

We found however that the best-correlated gene pairs (ISIGEM score = 0, $n = 741$) were not found in our benchmark dataset of HPPPIs. This might be because our relatively small benchmark dataset simply does not overlap well with ISIGEM, that is, the interactions identified by CSS and yeast two-hybrid in *P. falciparum* are different from those predicted by ISIGEM in *P. berghei*. However, we hypothesized that ISIGEM would often identify indirectly interacting proteins, because genes that are co-expressed between host and parasite might be upstream of those whose products directly interact at the host–parasite interface. Indeed, genes with correlated expression within a species are often not direct interactors but are involved in common pathways. To examine this, we used functional association networks from the STRING database for mouse and *Plasmodium*. These networks were clustered into highly connected subnetworks using MCL (29). We then asked whether the genes involved in ISIGEM predictions occurred in subnetworks along with gold-standard interaction partners. We found that the highest-scoring ISIGEM predictions (with score = 0) were more often in clusters containing benchmark interactions than expected by chance (P -value ≤ 0.001 ; Figure 1C). Thus, we reasoned that our predictions represent indirect, or upstream, interactions. Although these may still be direct HPPPIs, which are simply not in our benchmark dataset, this supports our general finding that correlated

gene expression between species is predictive of gene pairs involved in host–parasite interaction, be that directly or indirectly.

To determine the accuracy of our results, we calculated the expected number of false-positives based on the aforementioned analysis of proximity to known interactions. Many of our ISIGEM predictions could not be assessed because the genes involved were not represented in the STRING networks. Furthermore, we cannot reliably say that if we do not find an association between an ISIGEM prediction and a gold-standard interaction it is a false-positive, because the STRING networks are incomplete. Therefore, we calculated the expected number of false-positives by generating randomized predictions and determining how often these were found in the same clusters as gold-standard interactions. We found that, on average, over 1000 randomizations, 12.0 (standard error 4.6) interactions co-localized with gold-standard interactions (Figure 1C). Promisingly, 38 ISIGEM interactions co-localized. Thus, we expect that 31.6% (12/38) of predictions will be false-positives. Therefore, we expect that of the 741 interactions predicted by ISIGEM with a score of 0, 507 (68.4%) represent biologically meaningful indirect interactions between host and parasite gene products.

Using CSS interactions, we can generate a benchmark dataset for any system. Xu *et al.* performed a simultaneous microarray study of *P. berghei* with its vector, the mosquito *A. stephensi* (20). This dataset allowed us to perform an independent test of our approach and to extend our examination of malaria HPPPIs to this part of the life cycle. Unfortunately, we were unable to use a filtered CSS benchmark, as after filtering only a single interaction relating to genes on the microarray remained. Using the full CSS dataset as a benchmark, we found that correlation coefficients between host–parasite gene pairs were significantly different from those expected by chance (KS test $D = 0.1149$, $P = 2.2e-16$). The distribution of correlation coefficients between mosquito and *Plasmodium* genes (Figure 1D) showed an enrichment of both strong positive and strong negative correlations, unlike the mouse–*Plasmodium* dataset, which only showed a signal of positive correlation (Figure 1A). However, when we examined the distribution of correlation coefficients for CSS predictions between malaria and mosquito, we found that there was no enrichment for HPPPIs among negatively correlated profiles (Figure 1D; striped columns). More high-coverage gene expression datasets might shed light on the signal among the negative correlations. We did see, however, that CSS HPPPIs have a greater tendency towards positive correlations (Figure 1D; striped versus black), although the association is not statistically significant with $\alpha = 0.05$ (KS test; $D^+ = 0.2546$, P -value = 0.05092). When we converted the correlations to ISIGEM scores, we found that there was no enrichment of true-positive interactions below a score of 0.05 (χ^2 test, $P = 0.9672$), or a score of <0.2 (χ^2 test, $P = 0.2466$). We then looked to see whether high-scoring ISIGEM interactions co-localized with CSS interactions in STRING networks as we did for the mouse–*Plasmodium* predictions. We found that they did

Table 1. GO biological process terms enriched among mouse and *Plasmodium* genes predicted to be involved in HPPPIs using ISIGEM

GO Id	Term	Observed	TopGO <i>P</i> -value
Mouse genes			
GO:0006888	ER to Golgi vesicle-mediated transport	3	0.0048
GO:0006903	Vesicle targeting	2	0.0064
GO:0046165	Alcohol biosynthetic process	2	0.0124
GO:0006308	DNA catabolic process	2	0.0124
GO:0005979	Regulation of glycogen biosynthetic process	2	0.0124
GO:0006892	Post-Golgi vesicle-mediated transport	2	0.0124
GO:0006418	tRNA aminoacylation for protein translation	3	0.0127
GO:0015986	ATP synthesis coupled proton transport	2	0.0200
GO:0032434	Regulation of proteasomal ubiquitin-dependent protein catabolic process	2	0.0291
GO:0006396	RNA processing	5	0.0391
GO:0006457	Protein folding	3	0.0426
GO:0043242	Negative regulation of protein complex disassembly	2	0.0467
GO:0044403	Symbiosis, encompassing mutualism through parasitism	2	0.0468
Plasmodium genes			
GO:0006334	Nucleosome assembly	5	0.016
GO:0048193	Golgi vesicle transport	4	0.037
GO:0030163	Protein catabolic process	9	0.044

(*P*-value = 0.02; Figure 1C). We found a similar false-positive rate to that for ISIGEM predictions based on the mouse–*Plasmodium* dataset: 30% (12/40). Thus, we expect that of the 100 top-scoring predictions, 70 (70%) are truly indirect host–parasite interactions.

ISIGEM predictions between malaria and its mammalian host

We wanted to know whether ISIGEM interactions were commonly involved in particular biological processes. We selected 741 gene pairs with an ISIGEM score of 0 from the mouse–*Plasmodium* dataset and looked for GO biological process terms enriched among mouse genes in this set (Table 1). The most striking enriched term was *Symbiosis, encompassing mutualism through parasitism*, suggesting we have recovered genes known to be involved in inter-specific interactions. The first of these genes—*Scd1* or stearoyl-coenzyme A desaturase 1 (ENSMUSP0000036936)—was correlated with a plasmodium ABC transporter (PBANKA_136480), which may be involved with drug resistance and transportation of heavy metals (Figure 2A). *Scd1* encodes an iron-containing enzyme that catalyses a rate-limiting step in the synthesis of unsaturated fatty acids. It has been identified as part of a lipid-based antimicrobial effector pathway in mammals active against gram-positive bacteria (36). This suggests that this effector pathway or one related to it may be involved in *Plasmodium* infection. As we have shown, using the ISIGEM approach, we do not necessarily expect to find directly interacting proteins, but often those upstream. We examined genes functionally related to those identified by ISIGEM, for example, one link away in the STRING database. Here we found genes previously implicated in malaria: leptin and uncoupling protein 2. Leptin is a protein hormone involved in regulating energy intake and expenditure and is also involved in proinflammatory immune responses (37). Serum and urine levels of leptin have been found to be elevated in *P. berghei*-infected mice (38). Additionally, leptin seems to

be involved in placental malaria in humans, where *P. falciparum* infection disrupts the relationship between leptin levels and birth weight (39). Uncoupling protein 2 in mouse has previously been shown to be upregulated in *P. berghei*-infected mice and proposed to play a role in protecting from oxidative stress in the brain (40). We find that the *Plasmodium* ABC transporter is linked to PfMDR2, a membrane transporter putatively involved in removing xenobiotics such as folate from cells. Its homologue PfMDR1 is thought to be involved in resistance to antimalarials such as chloroquine (41).

The second mouse gene known to be involved in inter-specific interaction was Tap1 (ENSMUSP00000128401), part of the *Transporter associated with antigen processing* (TAP) complex. This gene was highly correlated with a minichromosome maintenance complex subunit (PBANKA_113160) from *Plasmodium*, part of the replication licensing factor (Figure 2B). TAP delivers antigenic peptides to the endoplasmic reticulum, where they bind to MHC class I molecules, which then display the antigens on the cell surface so that they can be recognized by T cells. A mutation in the *TAP1* promoter of humans has been associated with hyperparasitaemia and absence of hypoglycaemia in humans infected with malaria (42). A correlated increase in expression of replication licensing factor suggests proliferation of the parasite, perhaps in response to this recognition by the immune system.

The most significant functional term among malaria genes (Table 1) was *nucleosome assembly*, a process that is important in the control of malaria gene expression, especially during the intraerythrocytic development cycle (43). This may relate to a requirement for unpacking DNA before initiation of gene expression in response to signals from the host. Thus, it suggests a biologically meaningful, but non-specific, result, which may be common to further similar studies. It is perhaps not useful for identifying pairwise HPPPIs, although the host signals initiating this process may be of interest.

Table 2. GO term pairs significantly overrepresented in interacting genes of malaria and mosquito predicted using ISIGEM

Malaria term	Mosquito term	Observed	P-value
Cellular macromolecular complex subunit organization	Regulation of transcription, DNA-dependent	2	0.024
Proteolysis involved in cellular protein catabolic process	Chitin catabolic process	1	0.024
Chromatin assembly or disassembly	Proteolysis	1	0.024
Chromatin assembly or disassembly	Phosphate metabolic process	1	0.024
Chromatin assembly or disassembly	Regulation of transcription from RNA polymerase II promoter	1	0.035

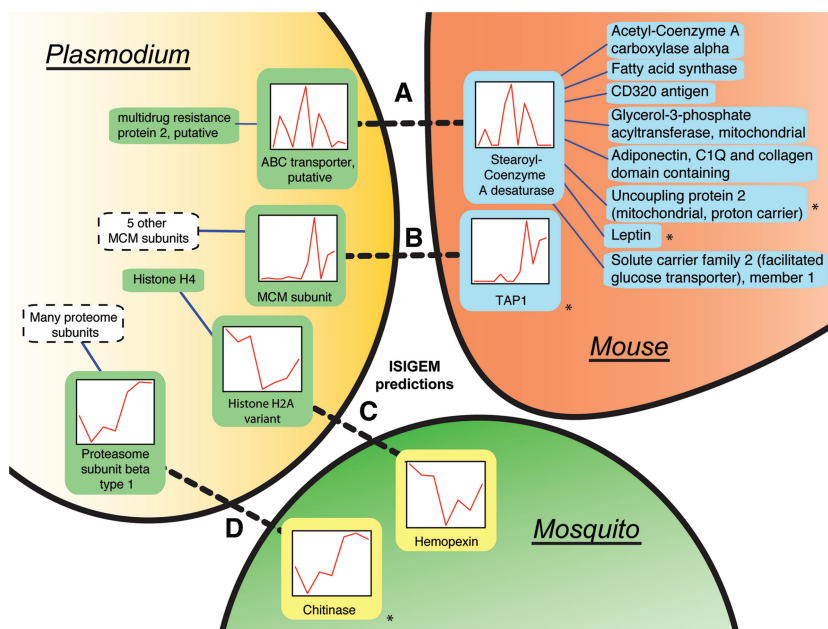


Figure 2. ISIGEM predictions between *Plasmodium*, its mammalian host and insect vector. We highlight here some of the interactions identified from the ISIGEM results using GO terms enrichment. Thick dashed lines represent associations identified by ISIGEM. Solid blue lines indicate intraspecific functional associations from the STRING database (score ≥ 750). Asterisks highlight genes previously identified as involved in malaria. Multiple similar interactors are collapsed into boxes with dashed outlines. Expression levels for genes involved in ISIGEM interactions are shown as red traces, the y-axis numbers are not shown, although the y-axis is normalized expression intensity. For **A** and **B**, the x-axis is 0, 3 and 6 days after infection (dpi) in brain tissue, and then 0, 3 and 6 dpi in liver; 0, 3 and 6 dpi in lung; and 0, 3 and 6 dpi in spleen. For **C** and **D**, the x-axis is 6, 20 and 40 hours after infection, and then 4, 8, 14 and 20 days after infection.

Several terms enriched in both mouse and malaria genes were related to the Golgi apparatus and vesicle targeting (Table 1). This suggests that we have detected interactions involving secretion systems. Secretion is particularly important for host–parasite interactions. To interact directly with the host in the blood stage, *Plasmodium* proteins are trafficked from the endoplasmic reticulum, perhaps through the Golgi to the parasite membrane. They then pass through the parasitophorous vacuole and are subsequently trafficked to the erythrocyte membrane (44). Similarly, in mammals, antibodies are secreted from B cells and membrane proteins are trafficked through similar routes. Although we see genes related to the Golgi apparatus among mouse interactors and among *Plasmodium* genes, we do not see pairs of correlated genes that are both involved in Golgi function. This suggests that secretion events in host and parasite are not closely associated; rather, secretion in one organism results in a different sort of response in the other.

ISIGEM predictions between malaria and its insect vector

At the stage of the malarial life cycle examined in the Xu dataset, the parasite, having been ingested in the insect blood meal, must invade the midgut epithelium. The peritrophic matrix, a mesh of chitin microfibrils, protects this epithelium. We examined 100 interactions between 45 *P. berghei* genes and 58 *A. stephensi* genes from the Xu dataset with score ≤ 0.001 . Examining overrepresented GO terms in these mosquito genes, we found only *translocation* to be enriched. No GO terms were enriched among *Plasmodium* interacting genes, although this may be due to few genes with GO terms being considered in the analysis.

If we consider pairs of GO terms occurring between correlated gene pairs (i.e. interacting GO terms), we identify five pairs of significantly enriched terms (Table 2). Three of these correlations are related to chromatin assembly on the malarial side, reinforcing the role of nucleosome remodelling in gene regulation for host–parasite interactions in malaria. One of these correlations

is between histone H2A variant from *P. berghei* and a haemopexin from mosquito (Figure 2C). Haemopexins are involved in scavenging haeme and recycling iron from it. This might relate to a reaction by *Plasmodium* to arrival in the gut where haemopexins are scavenging haeme from the blood meal. Additionally, we identified a chitinase gene (A0NGD1) correlated with a proteasome subunit in malaria (PBANKA_123310) (Figure 2D). It has been shown that midgut chitinase from the mosquito is required for invasion by the parasite (45). As the bloodmeal is digested the peritrophic matrix is slowly degraded by this chitinase. The proteasome response in *Plasmodium* might be a regulatory response to the presence of the chitinase, readying the parasite for crossing the midgut epithelium.

An integrated dataset of rodent malaria interactions

We have produced an integrated dataset comprising the high-confidence ISIGEM interactions, yeast two-hybrid interologues and CSS interactions (filtered for malaria and mouse, not filtered for malaria and mosquito). We have made this available as Supplementary Dataset S1. This dataset is the first network of rodent/mosquito malaria host–parasite interactions and provides a basis for future experiments and understanding of malarial host–parasite interaction. Furthermore, it can act as a benchmarking dataset for future approaches to prediction of host–parasite interactions.

DISCUSSION

We have shown that signatures of the molecular interactions between host and parasite are detectable in their transcriptomes. It is unclear to what extent this relates to direct interactions, but it is perhaps mechanistically more likely that, in general, correlated genes act upstream of direct molecular interactions. Although we identified a signal of HPPPI gene pairs between mouse and *Plasmodium*, the most high-scoring pairs were not found in our benchmark dataset. Furthermore, there was not a reliable signal of HPPPIs in the mosquito–*Plasmodium* dataset. However, in both cases, we found that the most high-scoring predictions were more closely associated with HPPPIs in functional networks than expected by chance, with a true-positive rate of ~70%. Although it is not currently clear whether the ISIGEM approach will be useful in predicting direct HPPPIs, it appears to be accurate in identifying individual pairs of genes from functional modules that interact between species.

Currently, this approach is limited in application by the paucity of available simultaneous host–pathogen gene expression datasets. Additionally, the lack of high-quality HPPPI data makes it difficult to assess the accuracy of our approach in identifying truly interacting proteins. This is a problem, however, for all approaches aiming to predict HPPPIs. The clear advantage of the ISIGEM approach compared with those previously proposed is that it does not rely on inferring interologues.

Approaches to HPPPI prediction using interologues are limited by the extent to which intraspecific and

interspecific interactions are conserved as well as the specificity of intraspecific interactions. Some proteins interact with a large variety of other proteins, whereas others may interact with only a single member of a large family, adding noise to the inference of interactions by homology. Although gene expression correlation is limited to those genes that are expressed in synchrony, it is adaptable and experiments can be tailored to identify interactions relating to particular aspects of host–parasite biology. Biophysical techniques such as yeast two-hybrid are able to identify pairs of host–parasite proteins that will physically interact. However, our approach can be tuned to look at groups of functionally related genes at particular stages of infection or in particular tissues by changing the experimental design.

Examining the predictions made by ISIGEM, we found that the genes identified were enriched for functional terms related to host–parasite interaction. We found several host genes known to be involved in various aspects of malaria infection. More generally, we confirmed that chromatin remodelling is important for malaria in interacting with its host, presumably in controlling the timing of gene expression (46). We also found that genes involved in vesicle transport to the Golgi are important in host–parasite interactions for both *Plasmodium* and mouse. For the parasite, this finding may relate to the modified secretion system, which *Plasmodium* uses to export proteins to the host cell surface (47). We provide an integrated dataset of interaction predictions between mouse and *Plasmodium* based on ISIGEM, CSS and interologues of experimentally determined interactions.

This is the first time that a genome-wide correlation has been shown between host–parasite gene expression and functional interactions between genes. This approach can be applied to study intimate interaction between any species, for example, between an organism and its pathogens, as well as symbiotic and commensal pairings, such as in the human gut microbiome. This work demonstrates the need for more simultaneous expression datasets, designed specifically to provide more power to HPPPI prediction. We believe that experiments, which are designed with this method in mind, will be more successful in extracting HPPPIs with fewer false-positives. Important things to consider are the collection of a sufficient number of time points and replicates and using RNA sequencing over microarrays to improve specificity of profiles.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary dataset 1.

ACKNOWLEDGEMENTS

The authors thank Lourdes Peña Castillo for supplying the by-probe intensities used in the original Lovegrove *et al.* study, and Thomas Dan Otto, James Cotton, Bernardo Foth and Chris Newbold for ideas and comments.

FUNDING

Wellcome Trust [098051]. Funding for open access charge: Wellcome Trust.

Conflict of interest statement. None declared.

REFERENCES

- Szklarczyk,D., Franceschini,A., Kuhn,M., Simonovic,M., Roth,A., Minguéz,P., Doerks,T., Stark,M., Müller,J., Bork,P. *et al.* (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.*, **39**, D561–D568.
- Uetz,P., Dong,Y.A., Zeretzke,C., Atzler,C., Baiker,A., Berger,B., Rajagopala,S.V., Roupelieva,M., Rose,D., Fossum,E. *et al.* (2006) Herpesviral protein networks and their interaction with the human proteome. *Science*, **311**, 239–242.
- Dyer,M.D., Murali,T.M. and Sobral,B.W. (2007) Computational prediction of host-pathogen protein-protein interactions. *Bioinformatics*, **23**, i159–i166.
- Davis,F.P., Barkan,D.T., Eswar,N., McKerrow,J.H. and Sali,A. (2007) Host pathogen protein interactions predicted by comparative modeling. *Protein Sci.*, **16**, 2585–2596.
- Crosnier,C., Bustamante,L.Y., Bartholdson,S.J., Bei,A.K., Theron,M., Uchikawa,M., Mboup,S., Ndir,O., Kwiatkowski,D.P., Duraisingh,M.T. *et al.* (2011) Basigin is a receptor essential for erythrocyte invasion by Plasmodium falciparum. *Nature*, **480**, 534–537.
- Miller,L.H., Baruch,D.I., Marsh,K. and Doumbo,O.K. (2002) The pathogenic basis of malaria. *Nature*, **415**, 673–679.
- Lin,N., Wu,B., Jansen,R., Gerstein,M. and Zhao,H. (2004) Information assessment on predicting protein-protein interactions. *BMC Bioinformatics*, **5**, 154.
- Lu,L.J., Xia,Y., Paccanaro,A., Yu,H. and Gerstein,M. (2005) Assessing the limits of genomic data integration for predicting protein networks. *Genome Res.*, **15**, 945–953.
- Patil,A. and Nakamura,H. (2005) Filtering high-throughput protein-protein interaction data using a combination of genomic features. *BMC Bioinformatics*, **6**, 100.
- Szklarczyk,D., Franceschini,A., Kuhn,M., Simonovic,M., Roth,A., Minguéz,P., Doerks,T., Stark,M., Müller,J., Bork,P. *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.*, **39**, D561–D568.
- Grigoriev,A. (2001) A relationship between gene expression and protein interactions on the proteome scale: analysis of the bacteriophage T7 and the yeast *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **29**, 3513–3519.
- Ge,H., Liu,Z., Church,G.M. and Vidal,M. (2001) Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nat. Genet.*, **29**, 482–486.
- Mrowka,R., Liebermeister,W. and Holste,D. (2003) Does mapping reveal correlation between gene expression and protein-protein interaction? *Nature Genet.*, **33**, 15–16, author reply 16–17.
- Shea,P.R., Virtaneva,K., Kupko,J.J. 3rd, Porcella,S.F., Barry,W.T., Wright,F.A., Kobayashi,S.D., Carmody,A., Ireland,R.M., Sturdevant,D.E. *et al.* (2010) Interactome analysis of longitudinal pharyngeal infection of cynomolgus macaques by group A *Streptococcus*. *Proc. Natl Acad. Sci. USA*, **107**, 4693–4698.
- Tierney,L., Linde,J., Müller,S., Brunke,S., Molina,J.C., Hube,B., Schock,U., Guthke,R. and Kuchler,K. (2012) An interspecies regulatory network inferred from simultaneous RNA-seq of *Candida albicans* invading innate immune cells. *Front Microbiol.*, **3**, 85.
- Lovegrove,F.E., Pena-Castillo,L., Mohammad,N., Liles,W.C., Hughes,T.R. and Kain,K.C. (2006) Simultaneous host and parasite expression profiling identifies tissue-specific transcriptional programs associated with susceptibility or resistance to experimental cerebral malaria. *BMC Genomics*, **7**, 295.
- Logan-Klumpler,F.J., De Silva,N., Boehme,U., Rogers,M.B., Velarde,G., McQuillan,J.A., Carver,T., Aslett,M., Olsen,C., Subramanian,S. *et al.* (2012) GeneDB—an annotation database for pathogens. *Nucleic Acids Res.*, **40**, D98–D108.
- Flicek,P., Amode,M.R., Barrell,D., Beal,K., Brent,S., Carvalho-Silva,D., Clapham,P., Coates,G., Fairley,S., Fitzgerald,S. *et al.* (2012) Ensembl 2012. *Nucleic Acids Res.*, **40**, D84–D90.
- Punta,M., Coghill,P.C., Eberhardt,R.Y., Mistry,J., Tate,J., Bournsell,C., Pang,N., Forslund,K., Ceric,G., Clements,J. *et al.* (2012) The Pfam protein families database. *Nucleic Acids Res.*, **40**, D290–D301.
- Xu,X., Dong,Y., Abraham,E.G., Kocan,A., Srinivasan,P., Ghosh,A.K., Sinden,R.E., Ribeiro,J.M., Jacobs-Lorena,M., Kafatos,F.C. *et al.* (2005) Transcriptome analysis of *Anopheles stephensi*-*Plasmodium berghei* interactions. *Mol. Biochem. Parasitol.*, **142**, 76–87.
- Kersey,P.J., Staines,D.M., Lawson,D., Kulesha,E., Derwent,P., Humphrey,J.C., Hughes,D.S., Keenan,S., Kerhornou,A., Koscielny,G. *et al.* (2012) Ensembl Genomes: an integrative resource for genome-scale data from non-vertebrate species. *Nucleic Acids Res.*, **40**, D91–D97.
- Sprinzak,E. and Margalit,H. (2001) Correlated sequence-signatures as markers of protein-protein interaction. *J. Mol. Biol.*, **311**, 681–692.
- Croft,D., O’Kelly,G., Wu,G., Haw,R., Gillespie,M., Matthews,L., Caudy,M., Garapati,P., Gopinath,G., Jassal,B. *et al.* (2011) Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.*, **39**, D691–D697.
- Hertz-Fowler,C., Peacock,C.S., Wood,V., Aslett,M., Kerhornou,A., Mooney,P., Tivey,A., Berriman,M., Hall,N., Rutherford,K. *et al.* (2004) GeneDB: a resource for prokaryotic and eukaryotic organisms. *Nucleic Acids Res.*, **32**, D339–D343.
- Krishnadev,O. and Srinivasan,N. (2008) A data integration approach to predict host-pathogen protein-protein interactions: application to recognize protein interactions between human and a malarial parasite. *In Silico Biol.*, **8**, 235–250.
- Aurrecochea,C., Brestelli,J., Brunk,B.P., Dommer,J., Fischer,S., Gajria,B., Gao,X., Gingle,A., Grant,G., Harb,O.S. *et al.* (2009) PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res.*, **37**, D539–D543.
- Sprenger,J., Lynn Fink,J., Karunarathne,S., Hanson,K., Hamilton,N.A. and Teasdale,R.D. (2008) LOCATE: a mammalian protein subcellular localization database. *Nucleic Acids Res.*, **36**, D230–D233.
- Vignali,M., McKinlay,A., LaCount,D.J., Chettier,R., Bell,R., Sahasrabudhe,S., Hughes,R.E. and Fields,S. (2008) Interaction of an atypical *Plasmodium falciparum* ETRAMP with human apolipoproteins. *Malaria J.*, **7**, 211.
- van Dongen,S. (2000) Graph Clustering by Flow Simulation, PhD thesis, University of Utrecht, May.
- Alexa,A., Rahnenfuhrer,J. and Lengauer,T. (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, **22**, 1600–1607.
- Jansen,R., Greenbaum,D. and Gerstein,M. (2002) Relating whole-genome expression data with protein-protein interactions. *Genome Res.*, **12**, 37–46.
- Motley,S.T., Morrow,B.J., Liu,X., Dodge,I.L., Vitiello,A., Ward,C.K. and Shaw,K.J. (2004) Simultaneous analysis of host and pathogen interactions during an in vivo infection reveals local induction of host acute phase response proteins, a novel bacterial stress response, and evidence of a host-imposed metal ion limited environment. *Cell Microbiol.*, **6**, 849–865.
- Guerfali,F.Z., Laouini,D., Guizani-Tabbane,L., Ottones,F., Ben-Aissa,K., Benkahla,A., Manchon,L., Piquemal,D., Smandi,S., Mghirbi,O. *et al.* (2008) Simultaneous gene expression profiling in human macrophages infected with *Leishmania major* parasites using SAGE. *BMC Genomics*, **9**, 238.
- Virtaneva,K., Porcella,S.F., Graham,M.R., Ireland,R.M., Johnson,C.A., Ricklefs,S.M., Babar,I., Parkins,L.D., Romero,R.A., Corn,G.J. *et al.* (2005) Longitudinal analysis of the group A *Streptococcus* transcriptome in experimental pharyngitis in cynomolgus macaques. *Proc. Natl Acad. Sci. USA*, **102**, 9014–9019.

35. Tailleux,L., Waddell,S.J., Pelizzola,M., Mortellaro,A., Withers,M., Tanne,A., Castagnoli,P.R., Gicquel,B., Stoker,N.G., Butcher,P.D. *et al.* (2008) Probing host pathogen cross-talk by transcriptional profiling of both *Mycobacterium tuberculosis* and infected human dendritic cells and macrophages. *PLoS One*, **3**, e1403.
36. Georgel,P., Crozat,K., Lauth,X., Makrantonaki,E., Seltmann,H., Sovath,S., Hoebe,K., Du,X., Rutschmann,S., Jiang,Z. *et al.* (2005) A toll-like receptor 2-responsive lipid effector pathway protects mammals against skin infections with gram-positive bacteria. *Infect. Immun.*, **73**, 4512–4521.
37. Loffreda,S., Yang,S.Q., Lin,H.Z., Karp,C.L., Brengman,M.L., Wang,D.J., Klein,A.S., Bulkley,G.B., Bao,C., Noble,P.W. *et al.* (1998) Leptin regulates proinflammatory immune responses. *FASEB J.*, **12**, 57–65.
38. Pulido-Mendez,M., De Sanctis,J. and Rodriguez-Acosta,A. (2002) Leptin and leptin receptors during malaria infection in mice. *Folia Parasitol. (Praha)*, **49**, 249–251.
39. Kabyemela,E.R., Fried,M., Kurtis,J.D., Mutabingwa,T.K. and Duffy,P.E. (2008) Fetal responses during placental malaria modify the risk of low birth weight. *Infect. Immun.*, **76**, 1527–1534.
40. Oakley,M.S., McCutchan,T.F., Anantharaman,V., Ward,J.M., Faucette,L., Erexson,C., Mahajan,B., Zheng,H., Majam,V., Aravind,L. *et al.* (2008) Host biomarkers and biological pathways that are associated with the expression of experimental cerebral malaria in mice. *Infect. Immun.*, **76**, 4518–4529.
41. Raj,D.K., Mu,J., Jiang,H., Kabat,J., Singh,S., Sullivan,M., Fay,M.P., McCutchan,T.F. and Su,X.Z. (2009) Disruption of a *Plasmodium falciparum* multidrug resistance-associated protein (PfMRP) alters its fitness and transport of antimalarial drugs and glutathione. *J. Biol. Chem.*, **284**, 7687–7696.
42. Niesporek,S., Meyer,C.G., Kremsner,P.G. and May,J. (2005) Polymorphisms of transporter associated with antigen processing type 1 (TAP1), proteasome subunit beta type 9 (PSMB9) and their common promoter in African children with different manifestations of malaria. *Int. J. Immunogenet.*, **32**, 7–11.
43. Pons,N., Harris,E.Y., Prudhomme,J., Wick,I., Eckhardt-Ludka,C., Hicks,G.R., Hardiman,G., Lonardi,S. and Le Roch,K.G. (2010) Nucleosome landscape and control of transcription in the human malaria parasite. *Genome Res.*, **20**, 228–238.
44. Crabb,B.S., de Koning-Ward,T.F. and Gilson,P.R. (2010) Protein export in *Plasmodium* parasites: from the endoplasmic reticulum to the vacuolar export machine. *Int. J. Parasitol.*, **40**, 509–513.
45. Shahabuddin,M., Toyoshima,T., Aikawa,M. and Kaslow,D.C. (1993) Transmission-blocking activity of a chitinase inhibitor and activation of malarial parasite chitinase by mosquito protease. *Proc. Natl Acad. Sci. USA*, **90**, 4266–4270.
46. Cui,L. and Miao,J. (2010) Chromatin-mediated epigenetic regulation in the malaria parasite *Plasmodium falciparum*. *Eukaryot. Cell*, **9**, 1138–1149.
47. Taraschi,T.F., O'Donnell,M., Martinez,S., Schneider,T., Trelka,D., Fowler,V.M., Tilley,L. and Moriyama,Y. (2003) Generation of an erythrocyte vesicle transport system by *Plasmodium falciparum* malaria parasites. *Blood*, **102**, 3420–3426.