Dual RNA Sequencing Meta-analysis in *Plasmodium* Infection Identifies Host-Parasite Interactions

Parnika Mukherjee,[a,b,c] Gaétan Burgio,[c] Emanuel Heitlinger[a,b]

aDepartment of Molecular Parasitology, Humboldt University, Berlin, Germany
bResearch Group Ecology and Evolution of Molecular Parasite-Host Interactions, Leibniz-Institute for Zoo and Wildlife Research (IZW), Berlin, Germany
cDepartment of Immunology and Infectious Diseases, John Curtin School of Medical Research, Australian National University, Canberra, ACT, Australia

ABSTRACT Dual RNA sequencing (RNA-Seq) is the simultaneous transcriptomic analysis of interacting symbionts, for example, in malaria. Potential cross-species interactions identified by correlated gene expression might highlight interlinked signaling, metabolic, or gene regulatory pathways in addition to physically interacting proteins. Often, malaria studies address one of the interacting organisms—host or parasite—rendering the other "contamination." Here we perform a meta-analysis using such studies for cross-species expression analysis. We screened experiments for gene expression from host and *Plasmodium*. Out of 171 studies in *Homo sapiens*, *Macaca mulatta*, and *Mus musculus*, we identified 63 potential studies containing host and parasite data. While 16 studies (1,950 samples) explicitly performed dual RNA-Seq, 47 (1,398 samples) originally focused on one organism. We found 915 experimental replicates from 20 blood studies to be suitable for coexpression analysis and used orthologs for meta-analysis across different host-parasite systems. Centrality metrics from the derived gene expression networks correlated with gene essentiality in the parasites. We found indications of host immune response to elements of the *Plasmodium* protein degradation system, an antimalarial drug target. We identified well-studied immune responses in the host with our coexpression networks, as our approach recovers known broad processes interlinked between hosts and parasites in addition to individual host and parasite protein associations. The set of core interactions represents commonalities between human malaria and its model systems for prioritization in laboratory experiments. Our approach might also allow insights into the transferability of model systems for different pathways in malaria studies.

IMPORTANCE Malaria still causes about 400,000 deaths a year and is one of the most studied infectious diseases. The disease is studied in mice and monkeys as lab models to derive potential therapeutic intervention in human malaria. Interactions between *Plasmodium* spp. and its hosts are either conserved across different host-parasite systems or idiosyncratic to those systems. Here we use correlation of gene expression from different RNA-Seq studies to infer common host-parasite interactions across human, mouse, and monkey studies. First, we find a set of very conserved interactors, worth further scrutiny in focused laboratory experiments. Second, this work might help assess to which extent experiments and knowledge on different pathways can be transferred from models to humans for potential therapy.

KEYWORDS dual RNA-Seq, malaria, meta-analysis, transcriptomics

Transcriptomes are often analyzed in a first attempt to understand molecular, cellular, and organismic events. A comprehensive profile of RNA expression can be obtained using high-throughput sequencing of cDNA from reverse transcribed expressed RNA. Such

RNA sequencing (RNA-Seq) provides high technical accuracy at a reasonable cost, making it the current method of choice for transcriptomics (1, 2).

In an infection experiment, RNA-Seq can assess host and pathogen transcriptomes simultaneously if RNA from both organisms is contained in a sample. On the one hand, it has been proposed to analyze transcriptomes of both organisms involved in an infection for a more complete understanding of the disease (3–5), such as virulence of a pathogen resulting from interlinked processes of both host and pathogen (host-pathogen interactions). This approach is called dual RNA-Seq. Some recent studies on malaria pathogenesis make use of dual RNA-Seq to study the host and the parasite simultaneously. On the other hand, researchers intending to study one of the two organisms, the target, might consider transcripts from the nontarget organism, "contamination." Malaria research is indeed traditionally designed to target one organism, either the host or the parasite. Nevertheless, expression of "contaminant" transcripts potentially corresponds to response to stimuli.

Malaria is the most thoroughly investigated disease caused by a eukaryotic organism and the accumulation of these two types of studies, RNA-Seq with "contaminants" and intentional dual RNA-Seq, provides a rich resource for meta-analysis. In the case of malaria, unlike in bacterial or viral infections, both the parasite, *Plasmodium* spp. and the host are eukaryotic organisms with similar transcriptomes (4). Their mRNAs have a long poly(A) tail at the 3′ end; therefore, host mRNA and parasite mRNA are selected simultaneously when poly(dT) priming is used to amplify polyadenylated mRNA transcripts (4, 5). This makes most malaria transcriptome data sets potentially suitable for dual RNA-Seq analysis.

Such a meta-analysis can use coregulated gene expression to infer host-parasite interactions. In the mammalian intermediate host, *Plasmodium* spp. first multiply in the liver and then invade red blood cells (RBCs) for development and asexual expansion. While the nuclear machinery from both host and parasite cells produces mRNA in the liver, RBCs are enucleated and transcriptionally inactive in the mammalian host. In the blood stage of the infection, leukocytes or white blood cells (WBCs) provide most of the transcriptomic response and are, thus, the source of host mRNA. Correlation of mRNA expression can be indicative of different types of biological interactions: protein products could be directly involved in the formation of complexes and might therefore be produced at quantities varying similarly under altered conditions. Alternatively, involvement in the same biological pathways can result in coregulated gene expression without physical interactions. This broad concept of interaction has long been exploited in single organisms (e.g., references 6, 7, and 8). We (and others before [9, 10]) propose to extrapolate this to interactions between the host and the parasite. It can be expected that a stimulus presented by the parasite to a host causes host immune response and the parasite in turn tries to evade this response, creating a cascade of genes coregulated at different experimental conditions.

Here we showed that existing raw sequencing read data sets collectively present the potential to answer questions that have not been investigated by individual studies on malaria. We explored this potential by conducting a comparative meta-analysis of dual RNA-Seq transcriptomes of *Plasmodium* and its hosts in the blood stage of infection. Since rodent and simian malaria are often used as laboratory models for human malaria, we demonstrated the availability and suitability of mRNA sequencing data from three evolutionarily close hosts—*Homo sapiens*, *Macaca mulatta*, and *Mus musculus*—and their respective *Plasmodium* species. We summarized available data and conceived an approach to elucidate host-parasite interactions using orthologs across different host-parasite pairs. We demonstrated that this approach provides meaningful results by cross-validation of information content from networks from different host-parasite systems. We found that these networks highlight known functional determinants of host-parasite interactions in broad categories such as invasion and calcium ion homeostasis for the parasite and host immune response and cadherin-mediated

TABLE 1 Potentially available host-parasite systems, number of studies, and number runs in different tissues after database querying

Host-bacterium system	No. of studies/no. of runs in tissue:					
	Blood	Spleen	Liver	Brain	Bone marrow	Lungs
Human-*P. falciparum*	14/855	0/0	0/0	0/0	0/0	0/0
Human-*P. vivax*	6/141	0/0	1/4	0/0	0/0	0/0
Human-*P. berghei*	0/0	0/0	6/145	0/0	0/0	0/0
Mouse-*P. yoelii*	3/19	2/10	3/20	0/0	1/4	1/6
Mouse-*P. chabaudi*	4/204	5/795	2/46	0/0	0/0	0/0
Mouse-*P. berghei*	3/44	0/0	0/0	3/83	0/0	0/0
Monkey-*P. cynomolgi*	5/811	0/0	2/59	0/0	0/0	0/0
Monkey-*P. coatneyi*	1/35	0/0	0/0	0/0	2/66	0/0

cell-cell adhesion on the host side. We also found novel interactions at a finer scale potentially worth further investigation.

RESULTS AND DISCUSSION

Potentially suitable studies for human, mouse, and simian malaria. We found 63 potentially suitable studies (3, 11–44; SRA accession no. SRP059851, ERP017542, ERP023982, ERP105548, ERP107298, SRP059851, SRP071199, SRP083918, SRP143373, SRP116117, SRP116593, SRP116793, SRP118503, SRP118827, SRP118996, ERP002116, ERP002273, ERP004042, ERP005730, ERP106769, SRP109709, SRP112213, SRP164767, SRP164768, SRP261098, ERP108490, and ERP114892) (see Table S1 in the supplemental material) on querying SRA and performing a literature review. The host organism for 27 studies was *Homo sapiens*, for 26, *Mus musculus*, and for 10, *Macaca mulatta*. The infecting parasites were *Plasmodium falciparum*, *Plasmodium vivax*, and *Plasmodium berghei* in human studies (including artificial infections of human hepatocyte culture with *P. berghei*), *Plasmodium yoelii*, *Plasmodium chabaudi*, and *P. berghei* in mouse studies and *Plasmodium cynomolgi* and *Plasmodium coatneyi* in macaque studies (Table 1). For 16 out of the 63 studies, the authors state that they intended to simultaneously study host and parasite transcriptomes. This includes eight studies from MaHPIC (Malaria Host-Pathogen Interaction Center), based at Emory University, which made extensive omics measurements in simian malaria. The original focus of the remaining 47 studies was on the parasite in 23 cases and on the host in 24 cases. Our collection of studies comprises data derived from blood and liver for all three host organisms, human, mouse, and macaque. Experiments performed on mouse blood focus on the parasite instead of the host (12 versus 0). Studies on human blood infection focus more often on the host immune response than on the parasite (10 versus 6). Liver and spleen studies focus on host and parasite almost equally as often, with sources for host tissue in this case being either mice (*in vivo*) or hepatoma cultures (*in vitro*).

We note that 21 of the 63 studies depleted or enriched specific classes of cells from their samples. Seventeen blood-stage studies depleted or enriched host WBCs to focus expression analysis on *Plasmodium* or host immunity, respectively. Assuming this depletion was imperfect, we tested whether such samples present mRNA from both organisms. The physiologically asymptomatic liver stage (20) is an appealing target for sporozoite-derived vaccines (45–47), but low parasite numbers make it difficult to study *Plasmodium* transcriptomes in liver. To reduce overwhelming host RNA levels, 10 out of the 14 liver studies sorted infected hepatoma cells from uninfected cells. Three out of the remaining four studies were among the studies interested in the host expression during the infection and one study enriched for host cells.

Blood and liver samples from different studies and host parasite systems are potentially suitable for dual RNA-Seq analysis. A sample or run must provide sufficient gene expression from both host and parasite to be suitable for dual RNA-Seq coexpression analysis. Analyzing the proportions of sequencing reads mapping to the
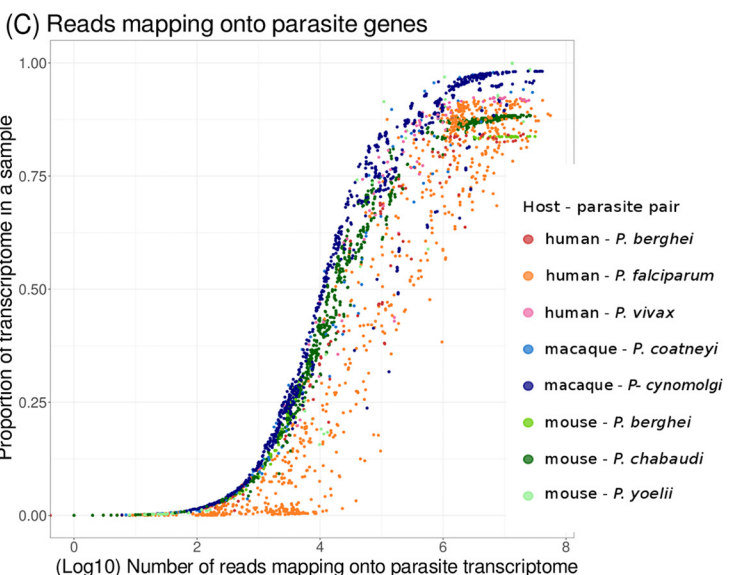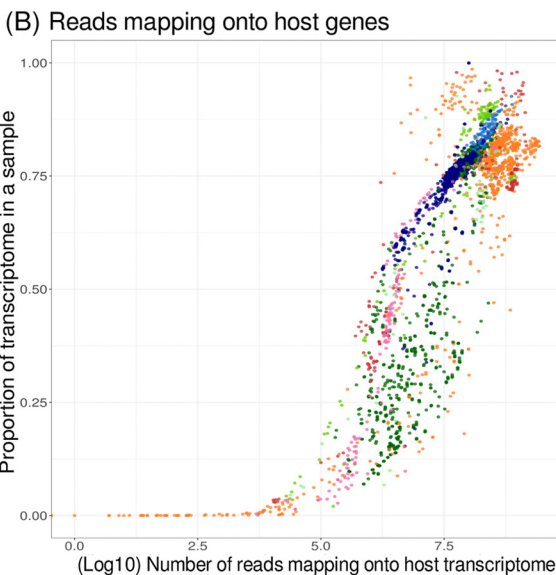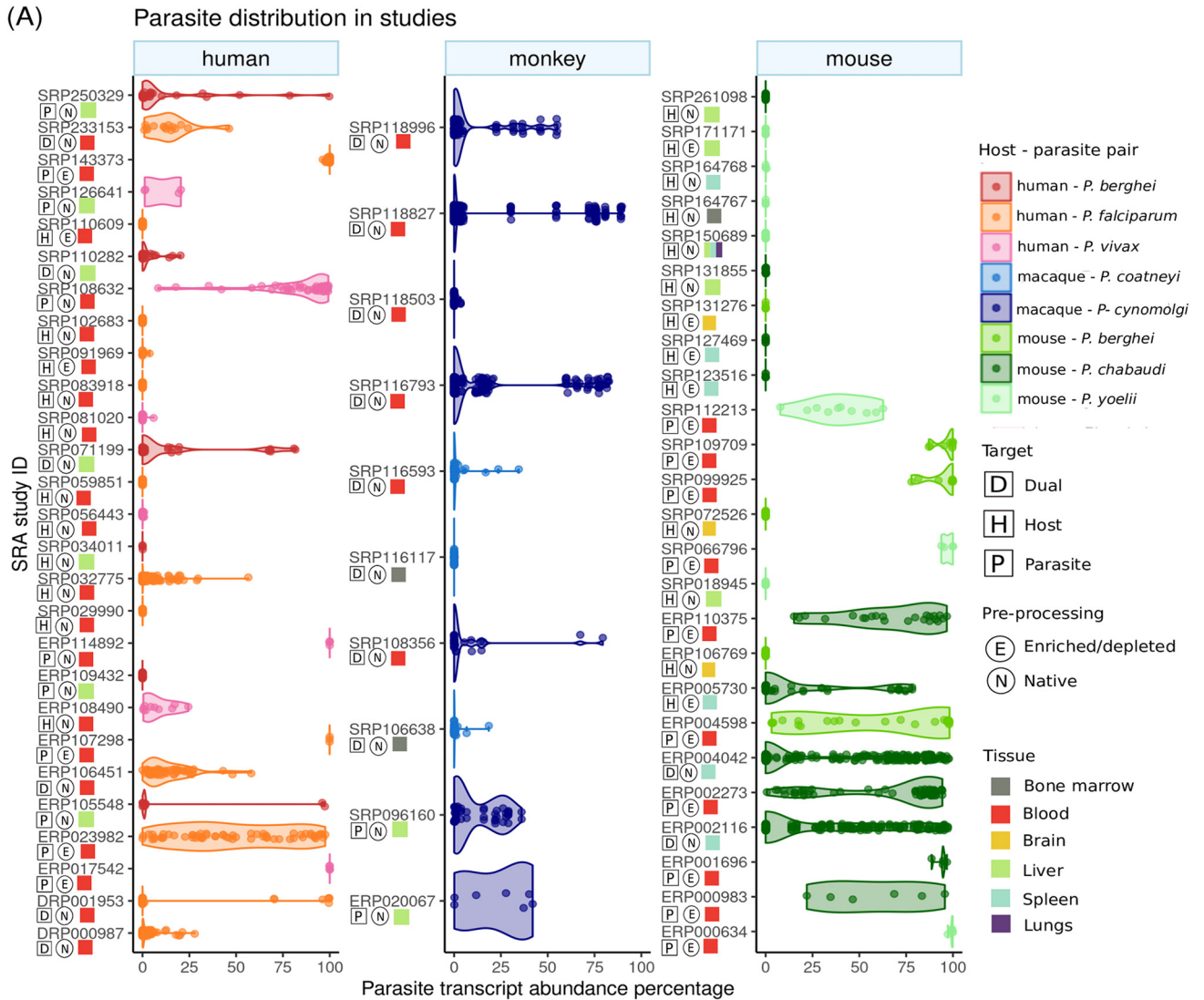
FIG 1 Proportion and number of sequencing reads and expressed genes from parasite and host in selected malaria RNA-Seq studies. We mapped sequencing reads from studies selected for their potential to provide both host and parasite gene expression data (total number of studies = 63; total

(Continued on next page)

host and parasite transcriptomes, we found that the original focus of the study was not always reflected (Fig. 1). In many native samples (not enriched or depleted for host material), the number of host reads was found to be overwhelming. However, probably when parasitemia was very high, parasite transcriptomes were still recovered. Some examples are runs in the studies with NCBI SRA accession no. SRP032775, SRP029990, and ERP106769. Similarly, many studies using depletion or enrichment of a certain cell type prior to RNA sequencing show considerable expression of the nontarget organism (enriched/depleted "E" in Fig. 1A). Examples are the studies ERP023982, ERP002273, ERP004598, ERP005730, ERP110375, and SRP112213. Studies depleting whole blood from leukocytes to focus on parasite transcriptomes still show considerable host gene expression and provide potentially suitable runs for the analysis of blood infection at lower intensities, although we note that this might come with the caveat that host expression might be biased by unequal depletion of particular cell types.

The number of sequencing reads mapping onto a transcriptome is a major determinant of the proportion of the organism's transcriptome detected as expressed (transcripts with at least one read mapped; Fig. 1B and C). As expected, only runs that were sequenced deeply were able to capture the expression of a high proportion of the transcriptome. For both host and parasite, this plateaus at the total number of genes expressed in the transcriptome of the respective tissue in these organisms. Runs with an acceptable proportion of the transcriptome detected as expressed were chosen in our analysis of blood-stage malaria below. We propose that thresholds based on the proportion of the transcriptome detected as expressed are most suitable for this selection of runs. Table 2 provides an overview of the number of studies and runs available for each of these analyses.

Evaluation of thresholds on transcriptome representation improve the analysis of coregulated gene expression. Deciding on the stringency of the thresholds applied to a dual RNA-Seq meta-analysis requires additional analysis. Previous studies have reported that estimated proportions of the transcriptionally active parts of the *P. falciparum* genome intraerythrocytic stages range from 60 to 90% (48–50). Here we test different thresholds that maximize the signal common between blood studies—the number of edges shared between correlation networks for different studies (Fig. 2).

Thirty-six studies investigated the blood stages of *Plasmodium*. Thirteen of them provided more than five runs at all thresholds (schematic in Fig. 2A) and were analyzed as independent data sets. Similarly, 11 studies could be analyzed separately with selection of runs at a "stringent" threshold of 70% transcriptome coverage. To make use of runs not meeting thresholds from studies that could not be analyzed separately, 6 runs from two studies were pooled into a combined "humanPvivax" data set and 10 runs from five studies into a combined mouse data set at intermediate thresholds. The combined mouse data set comprised four runs from two *P. berghei* studies, five runs from two *P. chabaudi* studies, and one run from a *P. yoelii* study. Sixteen studies did not provide any runs meeting the thresholds and were therefore excluded from the analysis. Based on the sum of Jaccard indices (see Materials and Methods) of the data sets, we selected a total of 15 sub-data sets maximizing overlap between individual study networks. We concatenated them to construct the "overall" data set: four studies without thresholds ("all"), seven at intermediate (50%; "int"), and four at stringent (70% "str") thresholds (marked with an asterisk in Fig. 2B to D). This means that a total of 915 runs were included in the "overall" data set.

We can conclude from this analysis that the edges shared between different studies within and among host-parasite systems (see below for an exception regarding human-*P.*

FIG 1 Legend (Continued)
number of runs = 3,351) against appropriate host and parasite genomes. (A) The percentage of parasite reads (*x* axis) is plotted for runs in each study (host and parasite add to 100%). The studies are categorized according to the host organisms studied and "enriched/depleted (E)" to indicate enrichment of infected hepatocytes or depletion of leukocytes from blood. Studies labeled "dual" were originally intended to simultaneously assess host and parasite transcriptomes. We also plot the number of reads mapped against the number of expressed genes for host (B) and parasite (C). The proportion of transcriptome detected as expressed increases with sequencing depth toward the maximum of all genes expressed in the transcriptome.
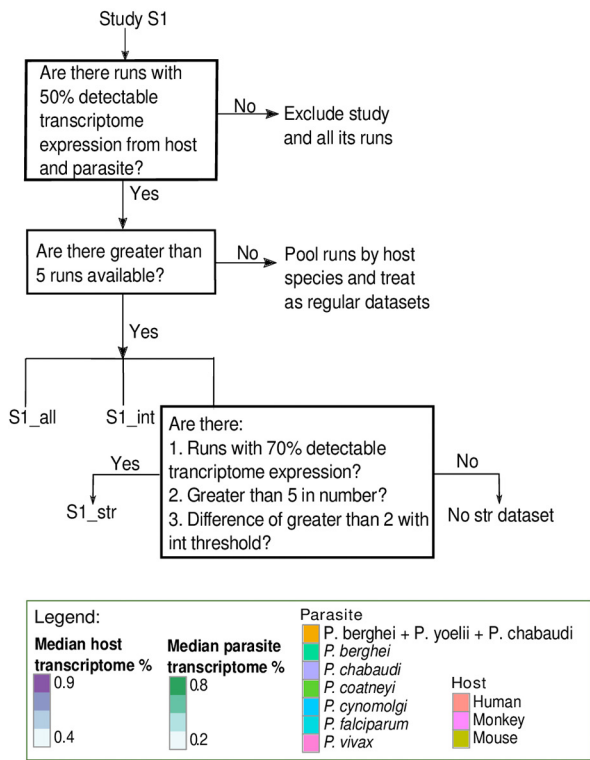
TABLE 2 Number of blood studies for each host-parasite pair and suitability analysis of their experimental replicates (runs)[a]

Host-parasite	# studies	# runs	Threshold: % transcriptome expressed	# studies with runs meeting threshold	# runs meeting threshold
Human – *P. falciparum*	13	621	30	8	180
			40	8	152
			50	6	131
			60	5	112
			70	4	101
Human – *P. vivax*	6	105	30	3	11
			40	3	10
			50	2	6
			60	2	5
			70	1	2
Monkey – *P. cynomolgi*	5	811	30	5	535
			40	5	500
			50	5	478
			60	5	370
			70	5	237
Monkey – *P. coatneyi*	1	34	30	1	21
			40	1	20
			50	1	18
			60	1	16
			70	1	13
Mouse – *P. berghei*	3	44	30	3	28
			40	3	23
			50	3	20
			60	1	13
			70	1	1
Mouse – *P. chabaudi*	4	91	30	4	31
			40	4	20
			50	3	8
			60	1	2
			70	0	0
Mouse – *P. yoelii*	3	14	30	1	5
			40	1	4
			50	1	1

[a]The intermediate threshold (yellow) and stringent threshold (green) are indicated by color.

falciparum studies) outnumber random expectations and are highly significantly ($P < 0.001$, Fisher's exact test; between all selected studies), indicating a biological signal shared between data sets (Fig. 2D). We suggest analyzing the optimal version (selection of runs) of each study instead of setting a fixed (e.g., intermediate) threshold. Figure 2 indicates
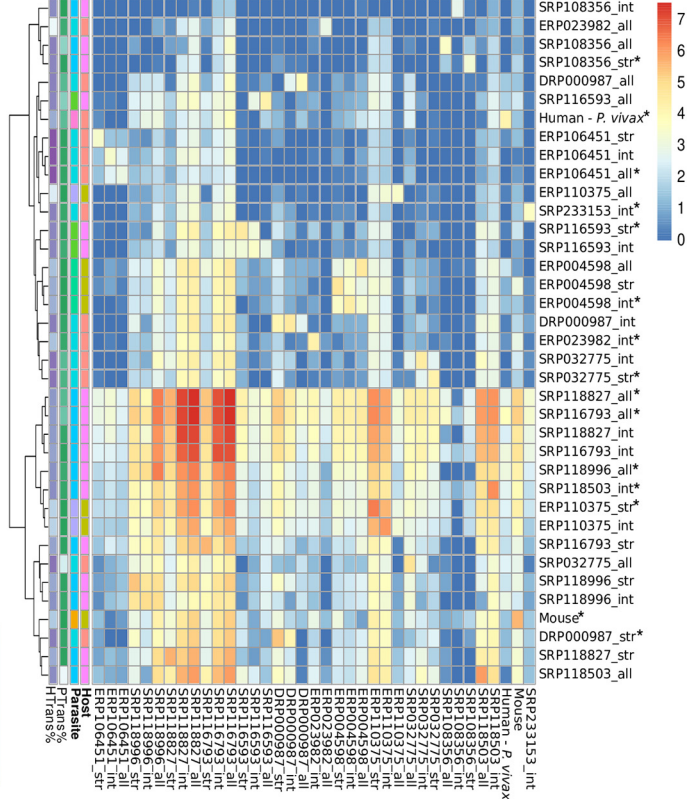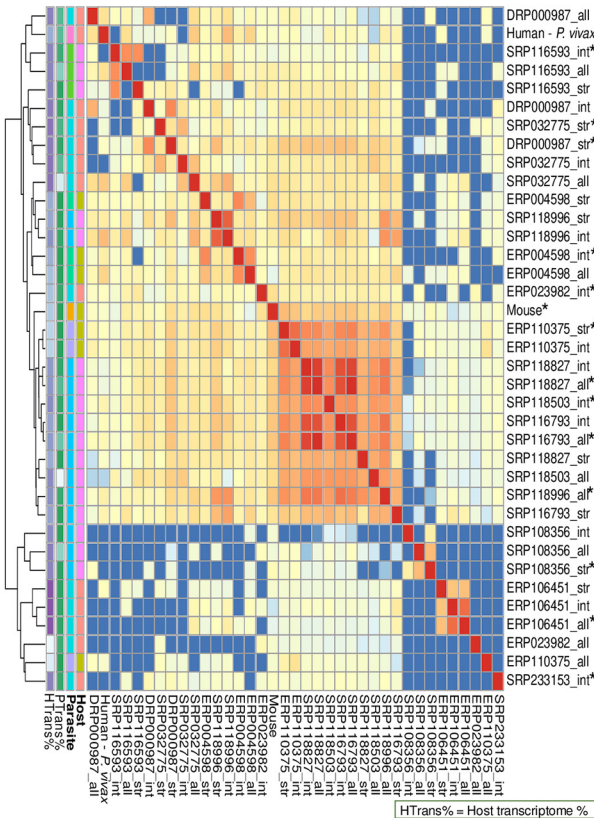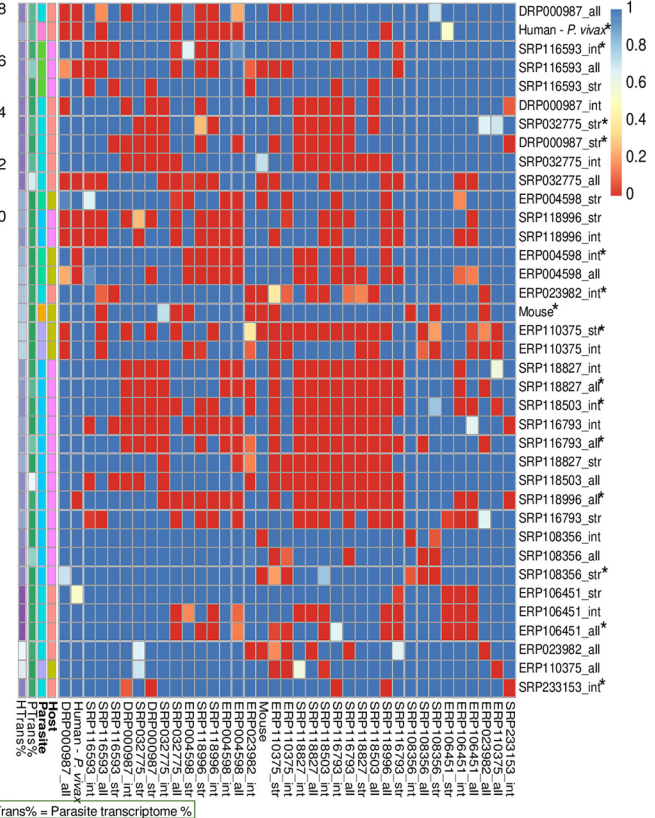
FIG 2 Overlapping bipartite edges between data sets across host-parasite systems. Based on the proportion of transcriptome expressed in each study, three thresholds were implemented on the selection of runs for coexpression analysis. Without thresholds, "all" includes all runs. The intermediate ("int")

which different thresholds we deem most suitable for different studies and our selection for downstream analyses.

Across different studies, across different host-parasite systems. Knowledge of one-to-one orthologs between different hosts and different parasite species can be used in the next steps to integrate across different host-parasite systems (Fig. 3). Humans share 18,179 1:1 orthologous genes with macaques and 17,089 with mice. A total of 13,986 genes are 1:1:1 orthologs among the three host species. Similarly, 4,005 one-to-one groups of orthologous genes were found among the *Plasmodium* species.

We have identified two different but interlinked workflows to reconstruct a consensus network of expression correlation (Fig. 3). A first approach integrates data from different studies of one host-parasite system by simply appending expression profiles of their runs. Expression data across host-parasite systems are combined in orthogroups, and correlations of gene expression are computed for all samples simultaneously. We refer to the results of this approach as the "overall" network (Fig. 3C). It has the benefits that it uses and weighs all the information sources in one concise process.

Alternatively, we search for consensus networks by comparison of individual networks from different studies (Fig. 3C). Similar to the approach of optimizing expression thresholds (Fig. 2 and 3B), we compare overlapping edges in a multilayer network analysis. Sets of overlapping edges from multiple studies and host-parasite systems offer more control when querying for similar correlation in different layers representing different host-parasite systems. Similarly, correlations from different types of tissues (blood and liver) could be combined as multilayer networks in future work.

Correlating expression on the "overall" data set resulted in a larger network (3.64 million edges, 12,652 host genes, 3,996 parasite genes) than on 13 out of 15 individual study data sets—two macaque data sets were exceptions (Fig. 4). A total of 528,883 bipartite edges are found only in the "overall" data set and not in any individual data set. Thirteen to 43% of edges from the "overall" data set are recovered in each individual data set (median, 18.5%), indicating a substantial contribution from each individual data set to the "overall" network. The most dominant study (presenting 43% of the edges in the "overall" network in its own network) was a mouse study (ERP004598 [29]) in which parasite gene expression was the primary focus. Ninety-one percent of this study's bipartite edges are shared with at least one other individual study in addition to the "overall" network. The second most influential study was a dual RNA-Seq study (on monkey, SRP118827 [11]) that shares 89% of its bipartite edges with at least one other study (Fig. 4A). This not only illustrates the contributions of studies that initially focused on one organism ("with contaminants") in dual RNA-Seq analysis but also that information can be pooled across different studies and host-parasite systems with our approach. Surprisingly, there were no common edges found between two human studies of Indonesian patients (12) and Gambian patients (3) (both sets of patients infected with *P. falciparum*), possibly indicating differences in sampling from patients. Alternatively, this might hint toward idiosyncrasies in human blood studies using native samples with high parasitemia. The potential difficulty to apply RNA-Seq metaanalysis to these human samples highlights the need to transfer such data across model systems while controlling for concordance with the human system.

FIG 2 Legend (Continued)
and stringent ("str") thresholds, include runs in which 50% ("int") and 70% ("str") of the transcriptome of both host and the parasite are detected as expressed (covered), respectively. (A) Schematic to show how samples were selected for coexpression analysis. (B) For each data set, the median host and parasite transcriptome coverage is indicated. The heatmap shows the $\log_{10}$-transformed number of bipartite edges for each study on the diagonal. The set sizes of the overlapping bipartite edges between corresponding data sets are displayed in the remaining fields. The data sets are clustered based on euclidean distances between these set sizes. The size of the overlapping set of bipartite edges is not determined by the median parasite or host transcriptome coverage, even studies with low coverage of the parasite transcriptome provide data sets leading to high number of bipartite edges found in common with other data sets. The optimal overlap set size does not suggest a single threshold for the selection of sequencing runs—all three thresholds are found to have a high number of overlapping edges with other data sets. (C) Jaccard indices (size of intersection as a ratio of the size of union) for each data set pair are displayed as $-\log_{10}$-transformed indices. The versions of the data sets used for further analysis are marked with an asterisk. (D) The size of the intersection was tested for significance using Fisher's exact test. The heatmap shows that all data sets led to correlation networks that overlapped with other networks from other data sets more than expected by chance. We can conclude that they contain biological signals suitable for combined analysis.
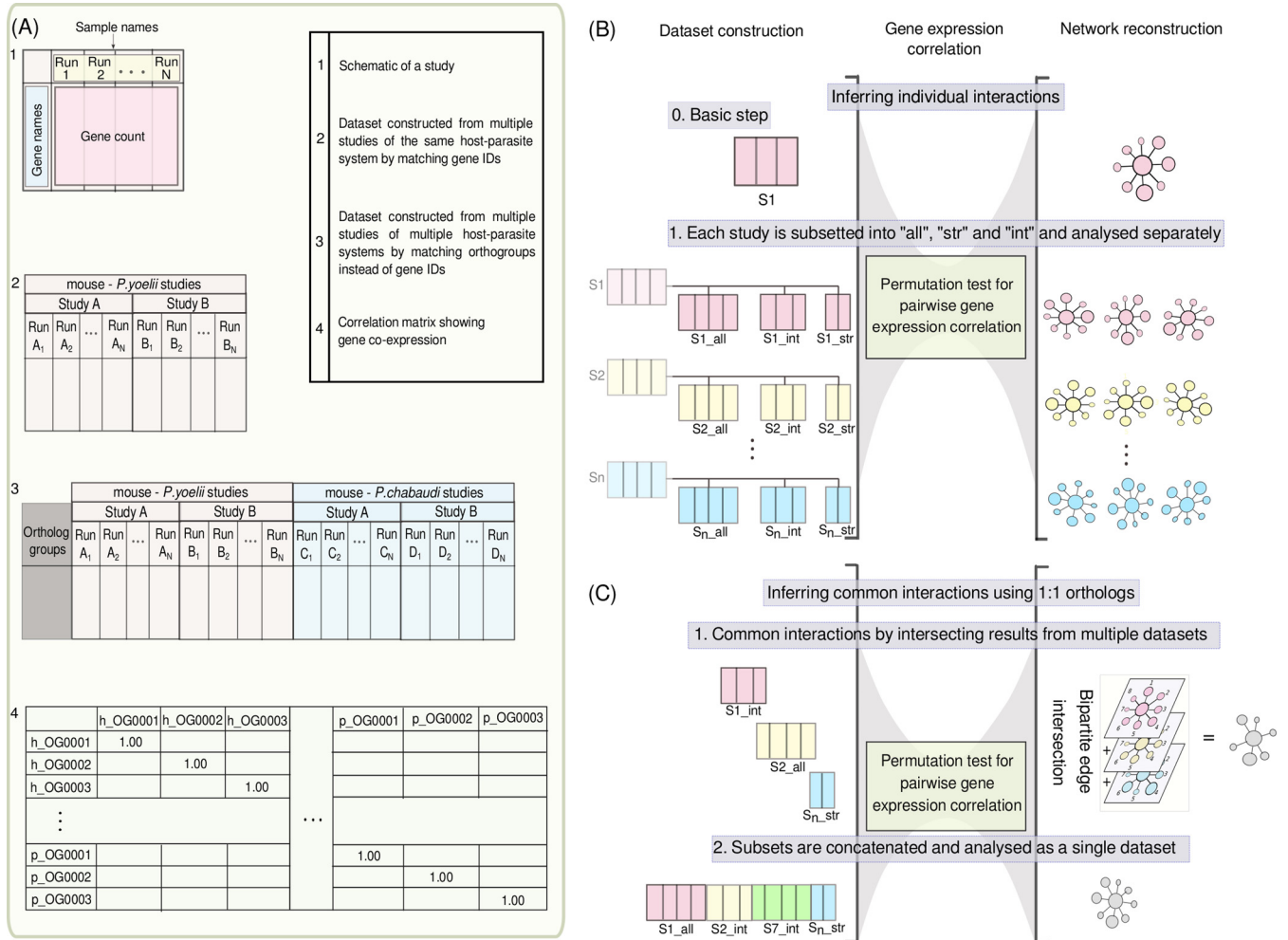
FIG 3 Alternative and interlinked strategies to reconstruct host-parasite interaction networks. We designed two interlinked approaches to obtain a consensus network involving multiple host-parasite systems. (A) An implementation of permutation tests derives "empirical *P* values" to test correlation coefficients for significance and infers interactions. h, host; p, parasite. (B) The methodology used to compare transcriptome coverage thresholds (50%, intermediate ["int"]; 70%, stringent ["str"]) is using different subsets of runs from each individual study ("sub-data sets" S1, S2, and S3 in this illustration). As in Fig. 2, without thresholds, "all" includes all runs. Each subset is analyzed separately to infer interactions, allowing the subset (and threshold) leading to highest information overlap with other studies to be selected. (C) Inference of common interactions across host-parasite systems using the knowledge of single-copy orthologs across the hosts and the parasites. Here, we can take two approaches: the first approach uses intersections in a multilayer network approach to come to a consensus of common interactions. In the second approach, we concatenate runs and correlate gene expression on those as a single "overall" data set. In the analysis for our final results, we combined these approaches mapping networks from each study onto the common network from an "overall" data set.

A "core" network of evolutionarily conserved interactions. The "overall" network is a highly connected graph with a total of ~3.64 million edges (Table S2). Even though this network provides some resolution relative to the ~56 million edges possible in a network of 13,986 host and 4,005 parasite genes, the resolution of this network might still be improved. We therefore defined a "core" network: using the "overall" network as a scaffold, we extracted edges that were recovered in at least one human study and at least one model organism. Using this definition, the resultant "core" network has 1,876 host genes and 2,050 parasite genes connected by 15,324 edges (Table S3). A list of Gene Ontology (GO) terms enriched or depleted in the genes of the "core" network are provided in Table S4. As expected, many GO terms were the same as in the "overall" network. Most GO terms in the "overall" network were enriched more strongly because of the higher number of genes in that network. However, many GO terms in the "overall" network were broader than in the "core" network. The recovery of more specific functions in the "core" network indicates higher resolution in this network (Table S4).
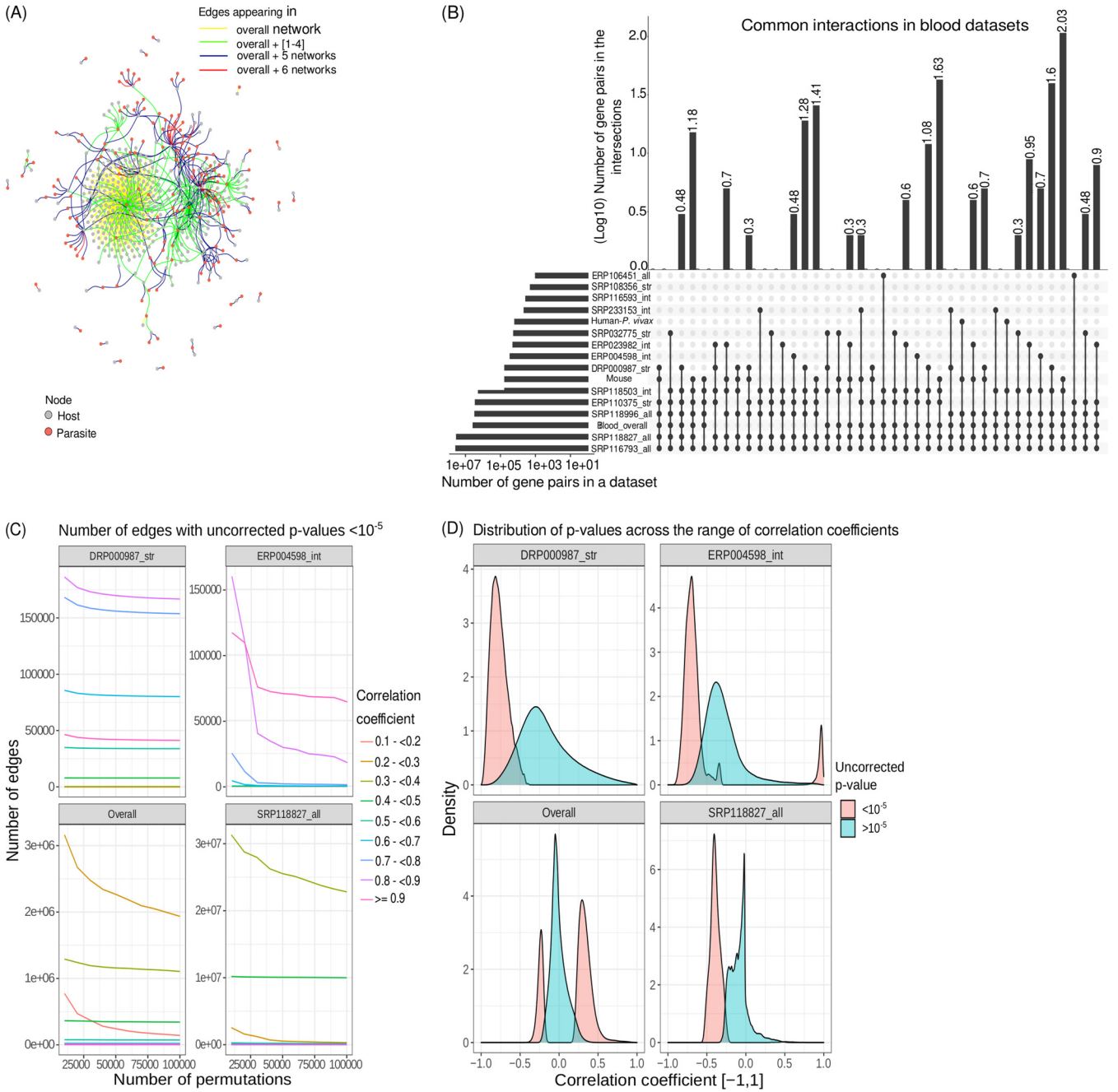
FIG 4 Multilayer networks of host-parasite interactions across host-parasite systems increase resolution over networks from single studies. We performed pairwise gene correlation tests on individual studies for gene expression in malaria and on an "overall" data set constructed by concatenating data from 15 data sets. (A) A small part of the "overall" network shows those overlaps. We selected this part of the network starting from a set of "core edges" found in most networks (the "overall" network and networks from six individual studies). Edges in the neighborhood to this were randomly selected in equal numbers, where available, based on how many networks, out of 15, they were found in. The number of networks an edge is found in is represented by the edge color. Several edges were found common to multiple data sets across host-parasite systems, a result later used to derive a "core" network. (B) The ($\log_{10}$-transformed) number of edges in networks from individual studies (horizontal bars) and shared (vertical bars) among networks shows the large overlaps between individual studies but also the large differences in the sizes for those overlaps. (C) Effect of performing an increasing number of permutations on the P value of an edge within a single network (data sets ERP004598_all, SRP118827_int, DRP000987_str, and "overall" as examples). As the number of permutations increases, the increase in resolution (decrease in the number of significantly correlated edges recovered) slows, warranting a different approach to increase resolution, such as the multilayer analysis depicted in previous panels. (D) Distribution of P values across the range of correlation coefficients. The same studies as in panel C are used. While individual data sets might represent different aspects of malaria infection and have a reduced number of interactions in common, we here show the benefit of analyzing the "overall" data set: the "overall" analysis identifies both positive and negative correlations (for a histogram, see Fig. S1 in the supplemental material). In the results of the "overall" data set, we reach a consensus of all the data sets included.

TABLE 3 Overview of beta regression models to explain relative growth rate (RGR) and mutagenesis index score (MIS) using eigenvector centrality measures of five networks

Network centrality measure or parameter[a]	Effect size with the following dependent variable in the indicated network[b]:									
	RGR					MIS				
	1	2	3	4	5	1	2	3	4	5
P. berghei EC	−1.445*** (0.089)					−0.754*** (0.087)				
Overall EC		−1.279*** (0.128)					−0.418*** (0.128)			
Core EC			−0.984*** (0.073)					−0.511*** (0.071)		
P. falciparum (I) EC				−0.833*** (0.267)					−0.952*** (0.272)	
P. falciparum (G) EC					−0.746*** (0.070)					−0.454*** (0.069)
Constant	1.092*** (0.048)	1.505*** (0.107)	0.880*** (0.042)	0.486*** (0.030)	0.767*** (0.041)	1.146*** (0.049)	1.159*** (0.108)	1.032*** (0.043)	0.836*** (0.032)	1.001*** (0.042)
No. of observations	2,169	2,169	2,169	2,169	2,169	2,169	2,169	2,169	2,169	2,169
R_2	0.083	0.038	0.053	0.003	0.033	0.03	0.004	0.018	0.004	0.016
Log likelihood	2,847.969	2,766.102	2,789.414	2,718.350	2,718.52	6,065.9	6,032.86	6,050.41	6,033.30	6,019.49

[a](I), Indonesia; (G), Gambia.
[b]The five networks are as follows: 1, a single study on *P. berghei* infection in mice; 2, overall; 3, core; 4, a single study on human *P. falciparum* infection in Indonesia; 5, another single study on human *P. falciparum* infection in Gambia. Values are effect sizes with errors in parentheses. *, $P < 0.1$; **, $P < 0.05$; ***, $P < 0.01$.

Gene coexpression explains gene essentiality. A node is important in a network if it is central, i.e., highly connected to other nodes in the network. The metrics node degree (DG), betweenness (BW,) and eigenvector centrality (EC) are network centrality measures that can quantify this. Similarly, an essential gene is classically described as a gene performing a function necessary for the viability of a cell or organism. We thus hypothesized central nodes in our parasite networks would tend to be more essential genes of *Plasmodium*.

In quantitative terms, more essential genes are more important for cell growth, and thus, the disruption of those genes leads to larger growth defects. Such data are available for *P. berghei* in mice, for which one study (51) reported "relative growth rates" (RGR) of mutant parasites reflecting the essentiality of a gene. A second study (52) reported *P. falciparum* growth rates impacted by genome-wide mutagenesis covering 5,399 protein-coding genes. Here, the mutagenesis index score (MIS) quantifies essentiality of a gene.

First, we tested the different network metrics for predicting gene essentiality (RGR and MIS) as response variables in beta-regression models. EC generally resulted in the best models, and combination of the centrality measures did not improve the models in most cases (Table S5). This might be explained by DG and EC being tightly correlated. We therefore report models with EC as a single predictor.

We then compared how well EC from different networks explained gene essentiality (Table 3 and Fig. 5): EC in the network derived from human-*P. falciparum* studies did significantly predict *P. falciparum* MIS and *P. berghei* RGR. EC from *P. berghei* networks was also a significant predictor for both *P. falciparum* MIS and *P. berghei* RGR. Surprisingly, *P. berghei* network centrality was a better predictor of MIS in *P. falciparum* than the centrality in *P. falciparum* expression networks. EC from both the "overall" network and the "core" network was a significant predictor for gene essentiality. Considering effect sizes, *P* values, and Akaike information criterion (AIC), we conclude
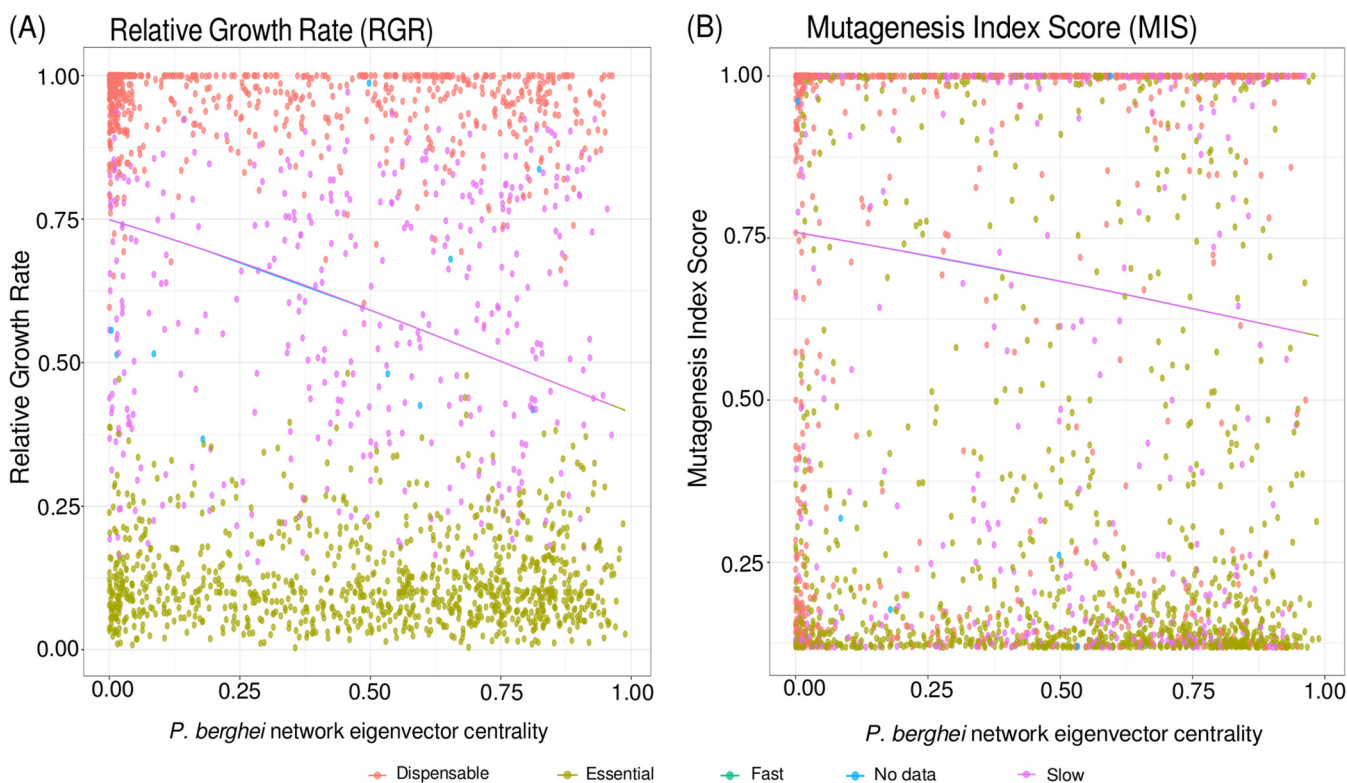
FIG 5 Explanation of gene essentiality with gene coexpression networks. (A and B) Eigenvector centrality in a *P. berghei* network of correlated parasite-parasite gene expression is plotted against relative growth rate (RGR) (A) and mutagenesis index score (MIS) (B). Lines depict predictions of the essentiality scores from the centrality measure in a beta-regression: the higher the centrality, the lower is the essentiality score (RGR and MIS), meaning that genes central in the network are more essential to the parasite's growth and survival. Color for both figures indicates phenotypes as categorized in reference 51.

that centrality in the *P. berghei* network best explains both RGR and MIS, followed by the metric from the "core" network. Given the much smaller size of the "core" network, the predictive power of centrality within this network is striking.

We can conclude from this analysis, that (i) our coexpression networks capture biological characteristics independently measured for the parasite and (ii) that parasite genes with a central position in the "core" network are important for the parasite's growth and survival. Highly connected genes in the intraparasitic "core" network, however, might control the reaction to host signals without necessarily directly interacting with host genes.

Interacting parasite processes and host immune response. On the basis of functional annotation of our networks, we show that the correlation between host and parasite transcript expression can highlight known processes important in host-parasite interactions (Fig. 6 and Table S4). Biological processes for hosts consistent among almost all networks include "Cell adhesion by cadherin" and "Calcium-dependent cell adhesion." Cadherins are cell adhesion proteins that depend on calcium. *Plasmodium* infection causes systemic endothelial activation of host blood vessels when the infected RBCs (iRBCs) sequester on the lining of blood vessels. This is accompanied with an increased interaction of endothelial cells with WBCs, as cell adhesion molecules on the vessel lining direct WBC trafficking to infected RBC sequestration sites (53). In addition, immune cells like B cells and monocytes express adhesion-related genes found in this GO term, like Fer (tyrosine-protein kinase) and ICAM-1 (intercellular adhesion molecule 1), respectively (54–56). To connect cooccurring host and parasite biological processes, we found associations between enriched (*P* value ≤ 0.05) host and parasite GO terms based on the interactors of the genes in these enriched GO terms. From the "core" network, we identified a GO network of 617 host GO terms and 464 parasite GO terms. This analysis suggests that besides the detailed coexpression of
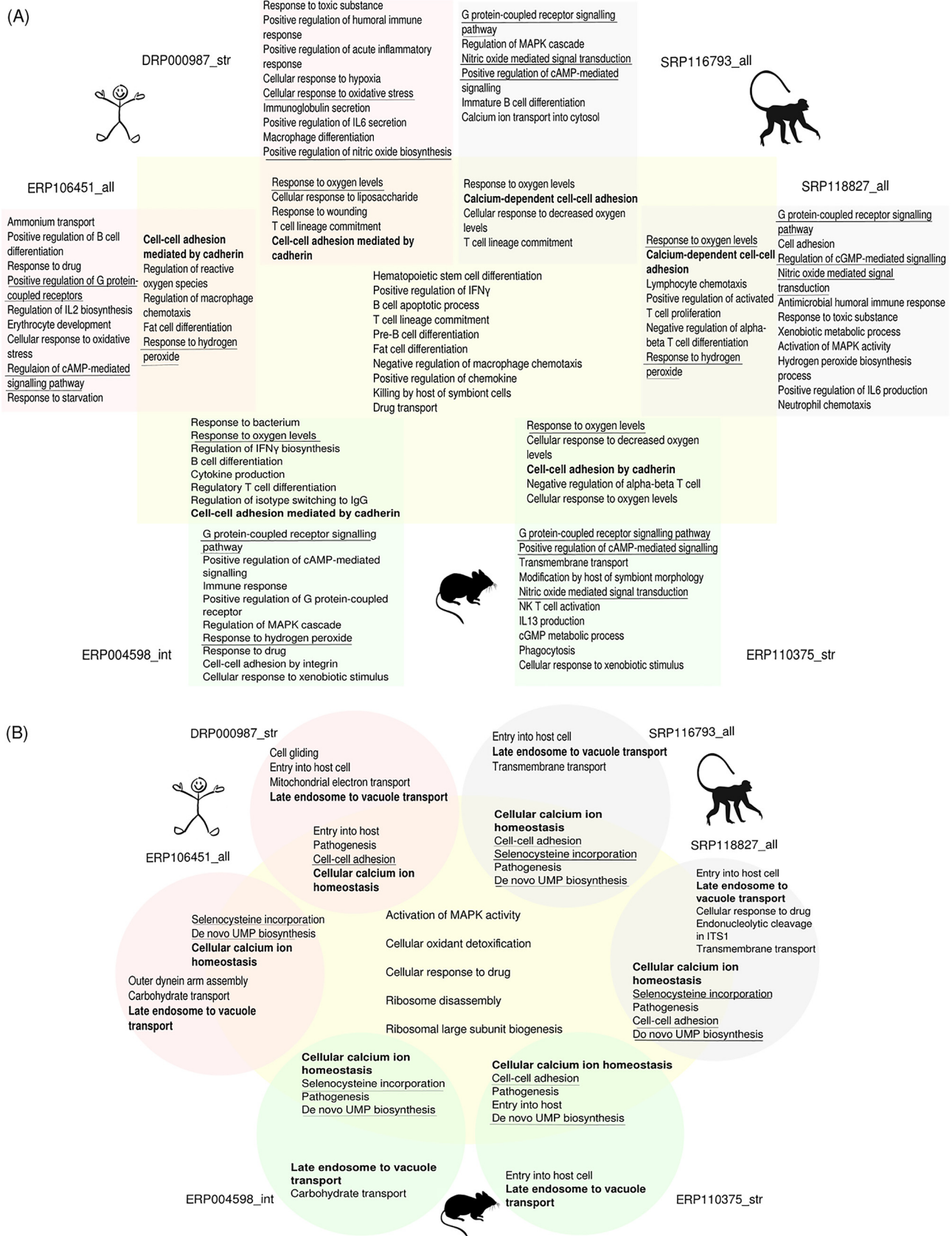
(A)

DRP000987_str

Response to toxic substance
Positive regulation of humoral immune response
Positive regulation of acute inflammatory response
Cellular response to hypoxia
Cellular response to oxidative stress
Immunoglobulin secretion
Positive regulation of IL6 secretion
Macrophage differentiation
Positive regulation of nitric oxide biosynthesis

G protein-coupled receptor signalling pathway
Regulation of MAPK cascade
Nitric oxide mediated signal transduction
Positive regulation of cAMP-mediated signalling
Immature B cell differentiation
Calcium ion transport into cytosol

SRP116793_all

ERP106451_all

Ammonium transport
Positive regulation of B cell differentiation
Response to drug
Positive regulation of G protein-coupled receptors
Regulation of IL2 biosynthesis
Erythrocyte development
Cellular response to oxidative stress
Regulaion of cAMP-mediated signalling pathway
Response to starvation

Cell-cell adhesion mediated by cadherin
Regulation of reactive oxygen species
Regulation of macrophage chemotaxis
Fat cell differentiation
Response to hydrogen peroxide

Response to oxygen levels
Cellular response to liposaccharide
Response to wounding
T cell lineage commitment
Cell-cell adhesion mediated by cadherin

Response to oxygen levels
Calcium-dependent cell-cell adhesion
Cellular response to decreased oxygen levels
T cell lineage commitment

Response to oxygen levels
Calcium-dependent cell-cell adhesion
Lymphocyte chemotaxis
Positive regulation of activated T cell proliferation
Negative regulation of alpha-beta T cell differentiation
Response to hydrogen peroxide

SRP118827_all

G protein-coupled receptor signalling pathway
Cell adhesion
Regulation of cGMP-mediated signalling
Nitric oxide mediated signal transduction
Antimicrobial humoral immune response
Response to toxic substance
Xenobiotic metabolic process
Activation of MAPK activity
Hydrogen peroxide biosynthesis process
Positive regulation of IL6 production
Neutrophil chemotaxis

Hematopoietic stem cell differentiation
Positive regulation of IFNγ
B cell apoptotic process
T cell lineage commitment
Pre-B cell differentiation
Fat cell differentiation
Negative regulation of macrophage chemotaxis
Positive regulation of chemokine
Killing by host of symbiont cells
Drug transport

Response to bacterium
Response to oxygen levels
Regulation of IFNγ biosynthesis
B cell differentiation
Cytokine production
Regulatory T cell differentiation
Regulation of isotype switching to IgG
Cell-cell adhesion mediated by cadherin

Response to oxygen levels
Cellular response to decreased oxygen levels
Cell-cell adhesion by cadherin
Negative regulation of alpha-beta T cell
Cellular response to oxygen levels

G protein-coupled receptor signalling pathway
Positive regulation of cAMP-mediated signalling
Immune response
Positive regulation of G protein-coupled receptor
Regulation of MAPK cascade
Response to hydrogen peroxide
Response to drug
Cell-cell adhesion by integrin
Cellular response to xenobiotic stimulus

G protein-coupled receptor signalling pathway
Positive regulation of cAMP-mediated signalling
Transmembrane transport
Modification by host of symbiont morphology
Nitric oxide mediated signal transduction
NK T cell activation
IL13 production
cGMP metabolic process
Phagocytosis
Cellular response to xenobiotic stimulus

ERP004598_int

ERP110375_str

(B)

DRP000987_str

Cell gliding
Entry into host cell
Mitochondrial electron transport
Late endosome to vacuole transport

Entry into host cell
Late endosome to vacuole transport
Transmembrane transport

SRP116793_all

ERP106451_all

Entry into host
Pathogenesis
Cell-cell adhesion
Cellular calcium ion homeostasis

Cellular calcium ion homeostasis
Cell-cell adhesion
Selenocysteine incorporation
Pathogenesis
De novo UMP biosynthesis

SRP118827_all

Selenocysteine incorporation
De novo UMP biosynthesis
Cellular calcium ion homeostasis

Outer dynein arm assembly
Carbohydrate transport
Late endosome to vacuole transport

Activation of MAPK activity

Cellular oxidant detoxification

Cellular response to drug

Ribosome disassembly

Ribosomal large subunit biogenesis

Entry into host cell
Late endosome to vacuole transport
Cellular response to drug
Endonucleolytic cleavage in ITS1
Transmembrane transport

Cellular calcium ion homeostasis
Selenocysteine incorporation
Pathogenesis
Cell-cell adhesion
Do novo UMP biosynthesis

Cellular calcium ion homeostasis
Selenocysteine incorporation
Pathogenesis
De novo UMP biosynthesis

Cellular calcium ion homeostasis
Cell-cell adhesion
Pathogenesis
Entry into host
De novo UMP biosynthesis

Late endosome to vacuole transport
Carbohydrate transport

Entry into host cell
Late endosome to vacuole transport

ERP004598_int

ERP110375_str

FIG 6 Biological processes (BP) Gene Ontology (GO) terms shared between data sets across host-parasite systems. We performed GO term enrichment analysis for the "overall" data set and its 15 constituent data sets for host (A) and parasite (B) genes. Here we show GO terms for the six largest studies for

gene products, broader enriched host and parasite processes or pathways are likely interacting in malaria infection (Fig. 7; Table S6).

Another host biological process commonly found in our interaction networks is the response to oxidative stress combined with response to hydrogen peroxide and nitric oxide-mediated signaling. It has been found that iRBCs produce twice as many free OH^- radicals and H_2O_2 than uninfected RBCs (29, 57). This mechanism is believed to be a defense mechanism to abate the infection, although the exact mechanism of parasite killing is still unclear. Nitric oxide (NO) production by the host is also believed to promote parasite death in malaria (58). An interesting GO term observed in the results was "Killing by host of symbiont cells." This term included genes that were all involved in defense response toward microbes by invoking immune cells (e.g., neutrophil cytosol factor 1 and cathepsin G), respiratory burst (Ncf1), and blood coagulation (e.g., prothrombin).

Host-parasite interaction networks from all infection systems were consistently enriched for the parasite processes, "Calcium ion homeostasis" and "Late endosome to vacuole transport." Calcium ion homeostasis regulated by calmodulins and Ca^{2+}-dependent kinases (CDPKs) play roles in complex signaling pathways and are important for apicomplexan parasite virulence (59, 60). Vacuolar protein sorting-associated proteins 1, 2, and 46 (VPS1, -2, and -46) along with serine/threonine protein kinase (VPS15) are the underlying signal for the enrichment of "Late endosome to vacuole transport" process and are interlinked with host immune reaction genes in our interaction networks. The endomembrane system is important for the parasite to invade the host cell, to establish the parasitophorous vacuole, and to obtain nutrients. The endosomal sorting complex required for transport (ESCRT) is involved in late endosome formation as endosomes exists at different stages of formation—from early to late endosome, before finally fusing with the lysosome or vacuole (reviewed in reference 61). The interrelation of the immune system and endosomal formation and sorting events should be a future focus of research, including our deeper studies of correlated gene expression networks.

To test whether our networks (gene clusters) are indicative of correlated gene expression originating from specific immune cells, we used immune cell marker genes established in reference 62. These marker genes are specific for six types of immune cells in the event of a *Plasmodium* infection in the blood: neutrophils and monocytes (innate immunity), T cells and B cells (adaptive immunity), and myeloid dendritic cells and NK cells (innate and adaptive). We found (Table S7) that specific markers from neutrophils were significantly overrepresented (Fisher's exact test, P value = 0.0089) in our "overall" network. Specific markers from monocytes were not significant but overrepresented (Fisher's exact test, P value = 0.19). This likely indicates phagocytosis and killing of *Plasmodium* via respiratory burst in which neutrophils and monocytes have a central role (63, 64). Specific markers for myeloid dendritic cells had significant underrepresentation (Fisher's exact test, P value = 0.0027). Marker genes for all other cell types were underrepresented (Fisher's exact test, P value > 0.05). In the "core" network, marker genes were not significantly enriched, but those specific for monocytes, neutrophils, and NK cells were tentatively overrepresented. We then used the presence of these specific cell type markers to annotate network clusters with the cell type indicated (Fig. 8). This resulted in 68 clusters in the "overall" network and 14 clusters in the "core" network annotated with a single cell type and additional 49 and 5 clusters annotated with multiple cell types, respectively. Among the most densely interconnected cluster in the "core" network with strong evidence of specific cell types was a set of 203 host genes (three are specific for T cells, two for neutrophils) correlating in their

FIG 6 Legend (Continued)
each host species. Enriched GO terms (P value $\geq$ 0.05) are presented in sets colored in red, green, and blue for human, mouse, and macaque studies, respectively. The GO terms of the "overall" data set are displayed in the central yellow set. GO terms for the six individual data sets with the highest contribution to the "overall" data set are shown in the perimeter of this. Overlapping between the respective areas shows shared GO terms found in either data set. Set overlaps are not illustrated between the six individual data sets, but respective terms are underlined (shared between four or five studies) or shown in bold type (shared between all six studies).
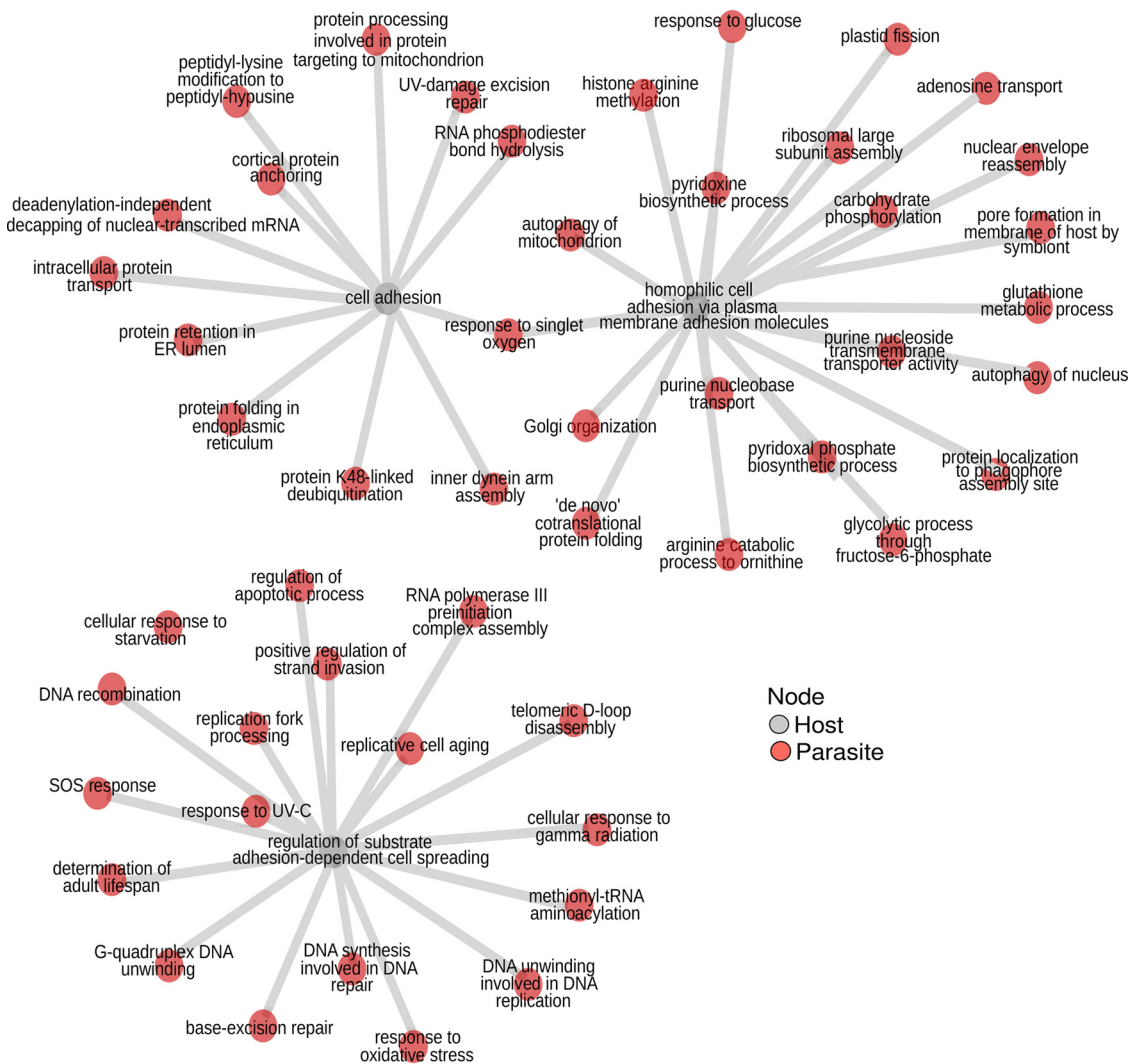
FIG 7 An interaction network of host-parasite GO terms. Host and parasite genes in the "overall" and "core" gene expression correlation networks were analyzed for GO term enrichment. For each individual enriched host GO term, interacting genes from the other species were then analyzed for enrichment more specifically. This gives an understanding of what GO biological processes in one organism, say, the host, are associated with specific processes in the parasite. In this figure, we show all parasite GO terms associated with host GO terms related to "adhesion" in the "core" network.

expression with only two parasite genes. One of them, a ubiquitin regulatory protein (PBANKA_1222400), is connected to 269 host genes in total in the "core" network. It appears to be uniformly expressed in all stages of the *Plasmodium* life cycle (65). WLL-vs (Leu-Leu-Trp vinyl sulfone) and WLW-vs (Trp-Leu-Trp vinyl sulfone) are proteasome inhibitors and likely drug candidates (66). In an experiment to study genetic changes mediating parasite recrudescence in WLL-vs and WLW-vs resistant mutants, this protein's level was amplified in recrudescent lines (67). Thus, it was suggested to confer low-grade resistance to such proteasome inhibitors. The second *Plasmodium* gene is a putative dynamin-like protein (DrpC, PBANKA_1434100), connected to 11 host genes. Dynamins are mechano-chemical enzymes with a GTPase domain and one of the functions of dynamin-related proteins (Drp) is to traffic vesicles (68). DrpC is conserved in apicomplexans and was shown to be localized at the base and periphery of the mitochondrion in *Toxoplasma gondii*, implying a probable role in mitochondrial fission (69).

Another module of three host genes included a NK cell marker correlated with two parasite genes—a 26S proteasome subunit (PBANKA_1206600) and ubiquitin fusion degradation protein UFD1 (PBANKA_1024700). The parasite 26S protease subunit
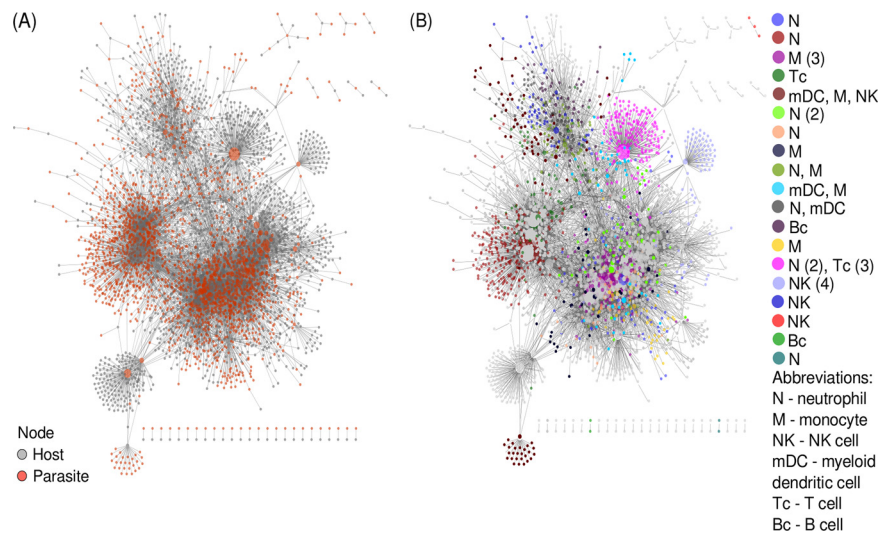
FIG 8 A "core" network highlights parasite interaction with specific immune cells. We derived a "core" network containing interacting genes supported by correlated gene expression in at least one human and one model organism study. (A) This complete "core" network is displayed with gray nodes for host and red nodes for parasite genes. The size of a node is based on its connectivity—the higher its degree, the larger the node. (B) identification of node clusters in the "core" network using an edge betweenness algorithm from igraph. Clusters with immune-cell-specific marker genes are colored (see legend, including a count for the number of marker genes in parentheses). The remaining clusters are left gray.

(70–72) was reported as an essential protein in *P. berghei* but not in *P. falciparum* (51, 52). In general, the ubiquitin-proteasome system (UPS) is essential for quality control of proteins in eukaryotes. In *Plasmodium*, the UPS is expressed across all life cycle stages and is speculated to be a promising drug target. It helps adapt the parasite to stress, such as changes in oxidative environment and temperature differences, ensuring the survival and virulence of the parasite. The *P. falciparum* ortholog of UFD1 (PF3D7_1418000) was reported to be a pathogenesis-related protein in an *in silico* module-based subnetwork alignment approach using protein-protein interactions of *Escherichia coli* as references (73).

One module had genes specific to three different immune cells—monocytes, myeloid dendritic cells (DC), and NK cells—the highest number of distinct immune cell types represented in a "core" network cluster. Among the 46 coexpressed parasite genes in this module were genes related to apicoplast biogenesis, mitochondrial fission, transcription, vacuolar transport from Golgi apparatus to the endoplasmic reticulum, and Fe-S cluster assembly proteins.

Overall, these results (Table S7) give an indication that certain gene expression clusters in our networks might be associated with specific cells of the innate immune response (neutrophils and monocytes). In addition, we show that parasite processes likely invoking this immune response include the expression of genes involved in drug resistance and in vesicular transport.

We next looked at overlaps between networks in our multilayer network analysis. Twenty interactions of specific gene pairs were conserved across six individual networks and the "overall" network (Fig. 4A): this includes (i) negatively correlated (Pearson's rho = −0.26) Kelch13 in *Plasmodium* with laminin subunit beta-2 (LAMB2) in the host and (ii) negatively correlated (Pearson's rho = −0.33) parasite 26S protease subunit and host LAMB2, both recovered only in mouse and monkey studies. LAMB2 is an extracellular high-affinity receptor and is associated with GO term "substrate adhesion-dependent cell spreading" to which infected erythrocytes (IE) have been reported to bind, along with other adhesion molecules such as ICAM-1 and vascular cell adhesion molecule (VECAM). Burgmann et al. (74) and Wenisch et al. (75) found that laminin levels in serum were increased in severe malaria. Mahamar et al. (76) later contradicted

this, finding that the binding of infected erythrocytes to endothelial receptors, including laminin, was the same in severe and nonsevere malaria. Kelch13 is a well-studied protein in which a C580Y mutation in *P. falciparum* confers resistance to artemisinin, a frontline antimalarial drug. It was found to be essential in both *P. falciparum* and *P. berghei* intraerythrocytic stages. Associated proteins and Kelch13 form an endocytic compartment and are essential for feeding on host hemoglobin. Kelch13 is also suggested to be a ubiquitin ligase with a role in the ubiquitin-proteasome system by labeling proteins for degradation (77). Artemisinin and its derivatives (ART) are known to be activated by the products of hemoglobin degradation and promote cell death by increasing endoplasmic reticulum (ER) stress facilitated by the accumulation of polyubiquitinated proteins. Resistance-conferring mutations on Kelch13 reduce host cell and hemoglobin endocytosis and along with 26S proteasome system (which includes the 26S protease subunit, our other parasite gene in discussion here), maintain the normal proteasomal degradation pathway preventing cell death, and thus, resulting in resistance (78–82). Our work here is, to our knowledge, the first suggestion that Kelch13 and LAMB2 might be involved in interacting host-parasite processes.

The gene pairs discussed above were recovered in seven data sets altogether (six individual data sets plus "overall") but were not recovered by any of the human-*Plasmodium* studies and are therefore not present in the "core" network. Of the 20 interactions that were found in seven data sets, five were also found in the "core" network and are indicated as such in Table S3. Of these five, one was an interaction between host protein odorant receptor ODR4 homolog (ENSG00000157181) and parasite thioredoxin-like associated protein 2 (TLAP2; PBANKA_0518100). TLAP2 is a microtubule-associated complex. It was found to be dispensable in both *P. berghei* and *P. falciparum* (51, 52). TLAP2 is conserved in *Toxoplasma gondii* and *Plasmodium* spp. (83). Along with other TLAP proteins, it is associated with protein TrxL1 (thioredoxin-like protein 1) and as a complex, coats cortical microtubules (83). These coating proteins, as an ensemble, stabilize the cortical microtubules (84). Host ODR4 is involved in trafficking of G-protein-coupled receptor (GPCR) proteins (85) and in protein localization (86).

A second interaction was between host biotinidase (ENSG00000169814) and parasite with protein kinase (PBANKA_1016200). In the host, biotinidase makes biotin available from dietary sources. Host biotin was found to be essential for *Plasmodium* survival in the liver stages but not in the blood stages (87). In general, several mutations in biotinidase reduce biotin availability without causing severe diseases. There have been suggestions that biotinidase treatment could be used as antimalarial therapies, following evidence that there are biotinidase mutations in Somali populations (88). Putative parasite protein kinase PBANKA_1016200 is a membrane protein involved in protein phosphorylation and ATP binding. Even though it is dispensable in both *P. berghei* and *P. falciparum*, in general, protein kinases have been an attractive subset of proteins to be used as antimalarial targets (reviewed in reference 89).

We propose that interactions between these proteins (physical protein-protein interaction) or interlinkage of associated pathways may be worth further scrutiny in mechanistic investigations.

Conclusion. Our analysis recovers both very broad and well-known processes involved in host-parasite interactions as well as strong interaction signals for narrower processes, including specific interacting gene pairs. This indicates that our analysis of interaction networks might uncover novel links between host and parasite processes helping to focus further research into detailed mechanisms of host-parasite interactions. Further analyses building on the methodology and concepts conceived here will involve networks derived from liver studies curated here, in which host and parasite cells interact more directly than in the blood. Methodologically, we show that analyses of expression from different host-parasite systems as multilayer networks can link host and parasite network modules. Meta-analysis of coexpression networks can thus assess whether host-parasite interactions are shared among different studies and treatments,

different tissues, and different host-parasite systems. This allows the identification of processes at different scales of molecular pathway organization as depicted above.

Moreover, the sharing of pathways among coexpression networks from different host-parasite systems might indicate how easily results on those pathways can be translated between different host-parasite systems—here presented in a "core" network. Our results can thus provide insights into how easily and for which pathways observations made in malaria models can be translated to human malaria.

MATERIALS AND METHODS

Data review and curation of potentially suitable studies. Sequence data generated in biological experiments is submitted to one of the three mirroring databases of the International Nucleotide Sequence Database Collaboration (INSDC): NCBI Sequence Read Archive (SRA) (90), EBI European Nucleotide Archive (ENA) (91), and DDBJ Sequence Read Archive (DRA) (92). We used SRAdb v1.36.0 (93), a Bioconductor/R package (94), to query SRA (90) for malaria RNA-Seq studies with the potential to provide host and *Plasmodium* reads for our meta-analysis. We first selected studies with library_strategy "RNA-Seq" and with "Plasmodium" in the study title, abstract, or sample attributes fields using the function dbGetQuery(). Then we used the getSRA() function with the query "(malaria OR Plasmodium) AND RNA-Seq." This function searches all fields. We manually curated the combined results and added studies based on a literature review using the terms described for the getSRA() function in PubMed and Google Scholar. We disregarded studies on vectors and nontarget hosts and studies derived from cell culture conditions devoid of transcriptionally active host cells. In these databases, all experiments submitted under a single accession number are given a single "study accession number" and are collectively referred to as a "study" here onwards. We used prefetch and fastq-dump functions from SRAtoolkit v2.8.2 to download all replicate samples (called runs in the databases) of the selected studies. The curation and the download of the studies were performed on 21 January 2019 and updated on 24 July 2020.

Mapping and quantification of gene expression. We mapped sequencing reads onto concatenated host and parasite reference genomes using STAR v2.6.0c (95, 96) with default parameters. Only runs with at least 70% uniquely mapping reads were considered for further analysis. Host and *Plasmodium* sp. genome assemblies and gene annotation files were downloaded from Ensembl version 43. Simultaneous mapping against both genomes should avoid nonspecific mapping of reads in regions conserved between host and parasites. We quantified the sequencing reads mapped to exons using the countOverlaps function of the GenomicRanges package v3.7 (97).

Identification of coexpressed genes via correlation techniques. Reid and Berriman (9) recommended using empirical P values for the analysis of gene coexpression. This allows us not only to scrutinize housekeeping genes likely to show almost uniform expression under different experimental conditions but also requires fewer assumptions about the quality of the input data. We computed correlation indices for each gene pair and obtained empirical P values by comparison against null distributions computed using permutations of the given data, instead of assuming a theoretical null distribution. This is a robust way to estimate whether gene pairs are correlated because of specific events (treatment condition, time point) and not by chance (e.g., housekeeping genes) (98, 99). To obtain P values corrected for multiple testing, since host and parasite genomes total at nearly 30,000 genes, the number of permutations would have to be around 1,012 for a resolution of 0.1% false discovery rate (FDR). As computational costs for these permutations are too high, we used uncorrected P values to rank genes as initially proposed by Reid and Berriman (9). Here we consider uncorrected P values of 10^{-5} (in 100,000 permutations) as evidence of coregulation.

Selection of runs for analysis. The construction of host-parasite gene coexpression networks require RNA-Seq runs to have both host and parasite transcript expression. To fulfil these criteria, we implemented thresholds based on host and parasite transcript expression on the selection of runs. If a study had at least five runs with 50% detectable transcriptome expression from the host and from the parasite, we defined two thresholds for runs: (i) "intermediate" (int) with at least 50% detectable host and parasite transcriptome expression and (ii) "stringent" (str) with 70% detectable host and parasite transcriptome expression. If a single study had less than five runs in the intermediate threshold, these runs were pooled with runs from other such studies.

To compare sub-data sets with runs selected at different thresholds and including all runs without thresholds, we calculated the Jaccard index for every pairwise combination of these data sets.

The Jaccard index is defined as $\frac{|A \cap B|}{|A \cup B|}$ (100), where A and B are the set of bipartite edges in or from two data sets. To include the maximum amount of host and parasite data and to use the best data set from each study, we calculated the sum of all Jaccard indices from each data set. We chose the sub-data set with the highest Jaccard index for a given study for further analysis and concatenated these representative studies into an "overall" data set to compute possible conserved interactions from all host-parasite systems.

Identification of orthologs. Orthologs are genes in different species that have been derived from a single gene in the last common ancestor of those species by vertical descent. This means that orthologous genes from different species, in general, are likely to perform the same function (101).

Whole-genome-based predictions of the proteome for host species were downloaded from Ensembl release 43 and for *Plasmodium* from PlasmoDB release 37. Orthologs were identified using OrthoFinder v2.2.7 (102) wherein blastp (103) results for all-versus-all protein comparisons are clustered using the MCL clustering algorithm (104) (tools bear default parameters for OrthoFinder). One-to-one ortholog clusters, or orthogroups, were recovered for hosts and for parasites.

Network and functional analyses. Bipartite networks/graphs are graphs in which the nodes of the graph can be divided into two independent sets, and the nodes from one set connect only with nodes in the other set (11, 105). To visualize host-parasite interactions, bipartite networks were constructed with R package igraph v1.2.5 (106) and Cytoscape v3.8.0 (107) where the nodes were the genes and their interaction were represented with an edge. To cluster nodes into network modules, we used the edge-betweenness algorithm [function edge_betweenness()] from igraph. The UpSetR package (v1.4.0) (108) in R was used to visualize intersection sizes of overlapping edges between networks.

For functional analysis of the genes found in host-parasite interactions from coexpression analysis, their Gene Ontology (109) terms were found using the package topGO v2.36.0 and enriched using Kolmogorov-Smirnov test (110). PlasmoDB (111) and Malaria.tools (65) databases were used for looking up gene functions and stage-associated transcript expression.

Statistical modeling. Using the R package fitdistrplus v1.1-1, we determined how relative growth rate (RGR) (51) and mutagenesis index score (MIS) (52) are distributed. We used the betareg package in R to model RGR and MIS using centrality properties of our networks.

The centrality properties measured for our networks were node degree (DG), eigenvector centrality (EC), and betweenness (BW) using their dedicated functions in igraph package in R: "degree()," "betweenness()," and "eigen_centrality()." RGR and MIS (as response variables) were modeled first with a single centrality measure as a predictor and then with combinations of two centrality measures: DG with EC and DG with BW. "Nested" models were compared based on likelihood ratio tests and the more complex model was accepted when delta likelihood exceeded 2. In addition to this, the Akaike information criterion (AIC) was computed for each model to compare models, and again models were considered differing in explanatory power at a delta-AIC of 2. Models on different data sets (and different response variables) were compared without explicit statistical testing discussing differences in P values for variables in question.

In general, R (v3.4.3 - 4.0.0) (112) was used for analysis.

Data availability. All code is available at https://doi.org/10.5281/zenodo.4535898 (as used in this article) and GitHub repository https://github.com/parnika91/CompBio-Dual-RNAseq-Malaria (under further development).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.
FIG S1, DOCX file, 0.1 MB.
TABLE S1, XLSX file, 0.1 MB.
TABLE S2, TXT file, 34.4 MB.
TABLE S3, XLSX file, 0.2 MB.
TABLE S4, XLSX file, 0.4 MB.
TABLE S5, XLSX file, 0.01 MB.
TABLE S6, XLSX file, 0.1 MB.
TABLE S7, XLSX file, 0.4 MB.

REFERENCES

1. Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63. https://doi.org/10.1038/nrg2484.
2. Stark R, Grzelak M, Hadfield J. 2019. RNA sequencing: the teenage years. Nat Rev Genet 20:631–656. https://doi.org/10.1038/s41576-019-0150-2.
3. Lee HJ, Georgiadou A, Walther M, Nwakanma D, Stewart LB, Levin M, Otto TD, Conway DJ, Coin LJ, Cunnington AJ. 2018. Integrated pathogen load and dual transcriptome analysis of systemic host-pathogen interactions in severe malaria. Sci Transl Med 10:eaar3619. https://doi.org/10.1126/scitranslmed.aar3619.
4. Westermann AJ, Gorski SA, Vogel J. 2012. Dual RNA-seq of pathogen and host. Nat Rev Microbiol 10:618–630. https://doi.org/10.1038/nrmicro2852.
5. Westermann AJ, Barquist L, Vogel J. 2017. Resolving host–pathogen interactions by dual RNA-seq. PLoS Pathog 13:e1006033. https://doi.org/10.1371/journal.ppat.1006033.

6. Stuart JM, Segal E, Koller D, Kim SK. 2003. A gene-coexpression network for global discovery of conserved genetic modules. Science 302:249–255. https://doi.org/10.1126/science.1087447.

7. Bono H, Okazaki Y. 2002. Functional transcriptomes: comparative analysis of biological pathways and processes in eukaryotes to infer genetic networks among transcripts. Curr Opin Struct Biol 12:355–361. https://doi.org/10.1016/S0959-440X(02)00335-4.

8. Lee HK, Hsu AK, Sajdak J, Qin J, Pavlidis P. 2004. Coexpression analysis of human genes across many microarray data sets. Genome Res 14:1085–1094. https://doi.org/10.1101/gr.1910904.

9. Reid AJ, Berriman M. 2013. Genes involved in host–parasite interactions can be revealed by their correlated expression. Nucleic Acids Res 41:1508–1518. https://doi.org/10.1093/nar/gks1340.

10. Musungu BM, Bhatnagar D, Brown RL, Payne GA, OBrian G, Fakhoury AM, Geisler M. 2016. A network approach of gene co-expression in the Zea mays/Aspergillus flavus pathosystem to map host/pathogen interaction pathways. Front Genet 7:206–214. https://doi.org/10.3389/fgene.2016.00206.

11. Tang Y, Anuj G, Swetha G, Galinski MR, Styczynski MP, Fonseca LL, Voit EO. 2018. Metabolic modeling helps interpret transcriptomic changes during malaria. Biochim Biophys Acta 1864:2329–2340. https://doi.org/10.1016/j.bbadis.2017.10.023.

12. Yamagishi J, Anna N, Tolba MEM, Mongan AE, Sugimoto C, Katayama T, Kawashima S, Makalowski W, Maeda R, Eshita Y, Tuda J, Suzuki Y. 2014. Interactive transcriptome analysis of malaria patients and infecting Plasmodium falciparum. Genome Res 24:1433 https://doi.org/10.1101/gr.158980.113.

13. Tran TM, Jones MB, Ongoiba A, Bijker EM, Schats R, Venepally P, Skinner J, Doumbo S, Quinten E, Visser LG, Whalen E, Presnell S, O'Connell EM, Kayentao K, Chaussabel D, Lorenzi H, Nutman TB, Ottenhoff THM, Haks MC, Traore B, Kirkness EF, Sauer-Wein RW, P D C. 2016. Transcriptomic evidence for modulation of host inflammatory responses during febrile Plasmodium falciparum malaria. Sci Rep 6:31291. https://doi.org/10.1038/srep31291.

14. Jaijyan DK, Singh H, Singh AP. 2015. A sporozoite-and liver stage-expressed tryptophan-rich protein plays an auxiliary role in Plasmodium liver stage development and is a potential vaccine candidate. J Biol Chem 290:19496–19511. https://doi.org/10.1074/jbc.M114.588129.

15. Rojas-Peña ML, Vallejo A, Herrera S, Gibson G, Arévalo-Herrera M. 2015. Transcription profiling of malaria-naïve and semi-immune Colombian volunteers in a Plasmodium vivax sporozoite challenge. PLoS Negl Trop Dis 9:e0003978. https://doi.org/10.1371/journal.pntd.0003978.

16. Rojas-Pena MC, Arafat DC, Velasquez JM, Garimalla SC, Arevalo-Herrera M, Herrera S, Gibson G. 2018. Profiling gene expression of the host response to a Plasmodium vivax irradiated sporozoite immunization and infectious challenge. bioRxiv https://doi.org/10.1101/501957.

17. Götz A, Tang MS, Ty MC, Arama C, Ongoiba A, Doumtabe D, Traore B, Crompton PD, Loke P, Rodriguez A. 2017. Atypical activation of dendritic cells by Plasmodium falciparum. Proc Natl Acad Sci U S A 114:E10568–E10577. https://doi.org/10.1073/pnas.1708383114.

18. Rothen J, Murie C, Carnes J, Anupama A, Abdulla S, Chemba M, Mpina M, Tanner M, Lee Sim BK, Hoffman SL, Gottardo R, Daubenberger C, Stuart K. 2018. Whole blood transcriptome changes following controlled human malaria infection in malaria pre-exposed volunteers correlate with parasite prepatent period. PLoS One 13:e0199392. https://doi.org/10.1371/journal.pone.0199392.

19. Kim A, Popovici J, Vantaux A, Samreth R, Bin S, Kim S, Roesch C, Liang L, Davies H, Felgner P, Herrera S, Arévalo-Herrera M, Méuard D, Serre D. 2017. Characterization of P. vivax blood stage transcriptomes from field isolates reveals similarities among infections and complex gene isoforms. Sci Rep 7:7761. https://doi.org/10.1038/s41598-017-07275-9.

20. LaMonte G, Orjuela-Sanchez P, Wang L, Li S, Swann J, Cowell A, Zou BY, Abel-Haleem Mohamed AM, Villa-Galarce ZH, Moreno M, Tong-Rios C, Vinetz J, Lewis N, Winzeler EA. 2019. Dual RNA-seq identifies human mucosal immunity protein mucin-13 as a hallmark of Plasmodium exoerythrocytic infection. Nat Commun 10:488. https://doi.org/10.1038/s41467-019-08349-0.

21. Quin JE, Bujila I, Chérif M, Sanou GS, Qu Y, Vafa Homann M, Rolicka A, Sirima SB, O'Connell MA, Lennartsson A, Troye-Blomberg M, Nebie I, Östlund Farrants A-K. 2017. Major transcriptional changes observed in the fulani, an ethnic group less susceptible to malaria. Elife 6:e29156. https://doi.org/10.7554/eLife.29156.

22. Gural N, Mancio-Silva L, Miller AB, Galstian A, Butty VL, Levine SS, Patrapuvich R, Desai SP, Mikolajczak SA, Kappe SH, Fleming HE, March S, Sattabongkot J, Bhatia SN. 2018. In vitro culture, drug sensitivity, and transcriptome of Plasmodium vivax hypnozoites. Cell Host Microbe 23:395–406.e4. https://doi.org/10.1016/j.chom.2018.01.002.

23. Cubi R, Vembar SS, Biton A, Franetich J-F, Bordessoulles M, Sossau D, Zanghi G, Bosson-Vanga H, Benard M, Moreno A, Dereuddre-Bosquet N, Le Grand R, Scherf A, Mazier D. 2017. Laser capture microdissection enables transcriptomic analysis of dividing and quiescent liver stages of Plasmodium relapsing species. Cell Microbiol 19:e12735. https://doi.org/10.1111/cmi.12735.

24. Van Der Wel AV, Roma G, Gupta DK, Schuierer S, Nigsch F, Carbone W, Zeeman A-M, Lee BH, Hofman SO, Faber BW, Knehr J, Pasini E, Kinzel B, Bifani P, Bonamy GMC, Bouwmeester T, Kocken CHM, Diagana TT. 2017. A comparative transcriptomic analysis of replicating and dormant liver stages of the relapsing malaria parasite Plasmodium cynomolgi. Elife 6:e29605. https://doi.org/10.7554/eLife.29605.

25. Joyner C, Moreno A, Meyer EVS, Cabrera-Mora M, Kong A, MAHPIC Consortium, Kissinger JC, Barnwell JW, Galinski MR. 2016. Plasmodium cynomolgi infections in rhesus macaques display clinical and parasitological features pertinent to modelling vivax malaria pathology and relapse infections. Malar J 15:451. https://doi.org/10.1186/s12936-016-1480-6.

26. Ogun SA, Tewari R, Otto TD, Howell SA, Knuepfer E, Cunningham DA, Xu Z, Pain A, Holder AA. 2011. Targeted disruption of py235ebp-1: invasion of erythrocytes by Plasmodium yoelii using an alternative Py235 erythrocyte binding protein. PLoS Pathog 7:e1001288. https://doi.org/10.1371/journal.ppat.1001288.

27. Lawton J, Brugat T, Yan YX, Reid AJ, Böhme U, Otto TD, Pain A, Jackson A, Berriman M, Cunningham D, Preiser P, Langhorne J. 2012. Characterization and gene expression analysis of the cir multi-gene family of Plasmodium chabaudi chabaudi (AS). BMC Genomics 13:125. https://doi.org/10.1186/1471-2164-13-125.

28. Spence PJ, Jarra W, Lévy P, Reid AJ, Chappell L, Brugat T, Sanders M, Berriman M, Langhorne J. 2013. Vector transmission regulates immune control of Plasmodium virulence. Nature 498:228−231. https://doi.org/10.1038/nature12231.

29. Guttery DS, Poulin B, Ramaprasad A, Wall RJ, Ferguson DJ, Brady D, Patzewitz E-M, Whipple S, Straschil U, Wright MH, Mohamed AM, Radhakrishnan A, Arold ST, Tate EW, Holder AA, Wickstead B, Pain A, Tewari R. 2014. Genome-wide functional analysis of Plasmodium protein phosphatases reveals key regulators of parasite development and differentiation. Cell Host Microbe 16:128–140. https://doi.org/10.1016/j.chom.2014.05.020.

30. Lin JW, Reid AJ, Cunningham D, Böhme U, Tumwine I, Keller-Mclaughlin S, Sanders M, Berriman M, Langhorne J. 2018. Genomic and transcriptomic comparisons of closely related malaria parasites differing in virulence and sequestration pattern. Wellcome Open Res 3:142. https://doi.org/10.12688/wellcomeopenres.14797.2.

31. Miller JL, Sack BK, Baldwin M, Vaughan AM, Kappe SH. 2014. Interferon-mediated innate immune responses against malaria parasite liver stages. Cell Rep 7:436–447. https://doi.org/10.1016/j.celrep.2014.03.018.

32. Li J, Cai B, Qi Y, Zhao W, Liu J, Xu R, Pang Q, Tao Z, Hong L, Liu S, Leerkes M, Quiñones M, Su X-Z. 2016. UTR introns, antisense RNA and differentially spliced transcripts between Plasmodium yoelii subspecies. Malar J 15:30. https://doi.org/10.1186/s12936-015-1081-9.

33. Torre S, Polyak MJ, Langlais D, Fodil N, Kennedy JM, Radovanovic I, Berghout J, Leiva-Torres GA, Krawczyk CM, Ilangumaran S, Mossman K, Liang C, Knobeloch K-P, Healy LM, Antel J, Arbour N, Prat A, Majewski J, Lathrop M, Vidal SM, Gros P. 2017. USP15 regulates type I interferon response and is required for pathogenesis of neuroinflammation. Nat Immunol 18:54–63. https://doi.org/10.1038/ni.3581.

34. Yeoh LM, Goodman CD, Mollard V, McFadden GI, Ralph SA. 2017. Comparative transcriptomics of female and male gametocytes in Plasmodium berghei and the evolution of sex in alveolates. BMC Genomics 18:734. https://doi.org/10.1186/s12864-017-4100-0.

35. Gabryšová L, Alvarez-Martinez M, Luisier R, Cox LS, Sodenkamp J, Hosking C, Pérez-Mazliah D, Whicher C, Kannan Y, Potempa K, Wu X, Bhaw L, Wende H, Sieweke MH, Elgar G, Wilson M, Briscoe J, Metzis V, Langhorne J, Luscombe NM, O'Garra A. 2018. c-Maf controls immune responses by regulating disease-specific gene networks and repressing IL-2 in CD4+ T cells. Nat Immunol 19:497–507. https://doi.org/10.1038/s41590-018-0083-5.

36. Mamedov MR, Scholzen A, Nair RV, Cumnock K, Kenkel JA, Oliveira JHM, Trujillo DL, Saligrama N, Zhang Y, Rubelt F, Schneider DS, Hsiu Chien Y, Sauerwein RW, Davis MM. 2018. A macrophage colony-stimulating-

factor-producing $\gamma\delta$ T cell subset prevents malarial parasitemic recurrence. Immunity 48:350–363.e7. https://doi.org/10.1016/j.immuni.2018.01.009.

37. Burrack KS, Huggins MA, Taras E, Dougherty P, Henzler CM, Yang R, Alter S, Jeng EK, Wong HC, Felices M, Cichocki F, Miller JS, Hart GT, Johnson AJ, Jameson SC, Hamilton SE. 2018. Interleukin-15 complex treatment protects mice from cerebral malaria by inducing interleukin-10-producing natural killer cells. Immunity 48:760–772.e4. https://doi.org/10.1016/j.immuni.2018.03.012.

38. Hirako IC, Assis PA, Hojo-Souza NS, Reed G, Nakaya H, Golenbock DT, Coimbra RS, Gazzinelli RT. 2018. Daily rhythms of TNFα expression and food intake regulate synchrony of Plasmodium stages with the host circadian cycle. Cell Host Microbe 23:796–808.e6. https://doi.org/10.1016/j.chom.2018.04.016.

39. Lai SM, Sheng J, Gupta P, Renia L, Duan K, Zolezzi F, Karjalainen K, Newell EW, Ruedl C. 2018. Organ-specific fate, recruitment, and refilling dynamics of tissue-resident macrophages during blood-stage malaria. Cell Rep 25:3099–3109.e3. https://doi.org/10.1016/j.celrep.2018.11.059.

40. Toro-Moreno M, Sylvester K, Srivastava T, Posfai D, Derbyshire ER. 2020. RNA-Seq analysis illuminates the early stages of Plasmodium liver infection. mBio 11:e03234-19. https://doi.org/10.1128/mBio.03234-19.

41. Tweedell RE, Tao D, Hamerly T, Robinson TM, Larsen S, Grønning AGB, Norris AM, King JG, Law HCH, Baumbach J, Bergmann-Leitner ES, Dinglasan RR. 2019. The selection of a hepatocyte cell line susceptible to Plasmodium falciparum sporozoite invasion that is associated with expression of glypican-3. Front Microbiol 10:127. https://doi.org/10.3389/fmicb.2019.00127.

42. Bradwell KR, Coulibaly D, Koné AK, Laurens MB, Dembélé A, Tolo Y, Traoré K, Niangaly A, Berry AA, Kouriba B, Plowe CV, Doumbo OK, Lyke KE, Takala-Harrison S, Thera MA, Travassos MA, Serre D. 2020. Host and parasite transcriptomic changes upon successive Plasmodium falciparum infections in early childhood. mSystems 5:e00116-20. https://doi.org/10.1128/mSystems.00116-20.

43. Kurup SP, Anthony SM, Hancox LS, Vijay R, Pewe LL, Moioffer SJ, Sompallae R, Janse CJ, Khan SM, Harty JT. 2019. Monocyte-derived CD11c+ cells acquire Plasmodium from hepatocytes to prime CD8 T cell immunity to liver-stage malaria. Cell Host Microbe 25:565–577.e6. https://doi.org/10.1016/j.chom.2019.02.014.

44. Subudhi AK, O'Donnell AJ, Ramaprasad A, Abkallo HM, Kaushik A, Ansari HR, Abdel-Haleem AM, Ben Rached F, Kaneko O, Culleton R, Reece SE, Pain A. 2020. Malaria parasites regulate intra-erythrocytic development duration via serpentine receptor 10 to coordinate with host rhythms. Nat Commun 11:2763. https://doi.org/10.1038/s41467-020-16593-y.

45. Hollingdale MR, Sedegah M. 2017. Development of whole sporozoite malaria vaccines. Expert Rev Vaccines 16:45–54. https://doi.org/10.1080/14760584.2016.1203784.

46. Seder RA, Chang LJ, Enama ME, Zephir KL, Sarwar UN, Gordon IJ, Holman LA, James ER, Billingsley PF, Gunasekera A, Richman A, Chakravarty S, Manoj A, Velmurugan S, Li M, Ruben AJ, Li T, Eappen AG, Stafford RE, Plummer SH, Hendel CS, Novik L, Costner PJ, Mendoza FH, Saunders JG, Nason MC, Richardson JH, Murphy J, Davidson SA, Richie TL, Sedegah M, Sutamihardja A, Fahle GA, Lyke KE, Laurens MB, Roederer M, Tewari K, Epstein JE, Sim BK, Ledgerwood JE, Graham BS, Hoffman SL, VRC 312 Study Team. 2013. Protection against malaria by intravenous immunization with a nonreplicating sporozoite vaccine. Science 341:1359–1365. https://doi.org/10.1126/science.1241800.

47. Bijker EM, Borrmann S, Kappe SH, Mordmüller B, Sack BK, Khan SM. 2015. Novel approaches to whole sporozoite vaccination against malaria. Vaccine 33:7462–7468. https://doi.org/10.1016/j.vaccine.2015.09.095.

48. Chappell L, Ross P, Orchard L, Russell TJ, Otto TD, Berriman M, Rayner JC, Llinás M. 2020. Refining the transcriptome of the human malaria parasite Plasmodium falciparum using amplification-free RNA-Seq. BMC Genomics 21:395. https://doi.org/10.1186/s12864-020-06787-5.

49. Bozdech Z, Llinás M, Pulliam BL, Wong ED, Zhu J, DeRisi JL. 2003. The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. PLoS Biol 1:e5. https://doi.org/10.1371/journal.pbio.0000005.

50. Otto TD, Wilinski D, Assefa S, Keane TM, Sarry LR, Böhme U, Lemieux J, Barrell B, Pain A, Berriman M, Newbold C, Llinás M. 2010. New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq. Mol Microbiol 76:12–24. https://doi.org/10.1111/j.1365-2958.2009.07026.x.

51. Bushell E, Gomes AR, Sanderson T, Anar B, Girling G, Herd C, Metcalf T, Modrzynska K, Schwach F, Martin RE, Mather MW, McFadden GI, Parts L,

Rutledge GG, Vaidya AB, Wengelnik K, Rayner JC, Billker O. 2017. Functional profiling of a Plasmodium genome reveals an abundance of essential genes. Cell 170:260–272.e8. https://doi.org/10.1016/j.cell.2017.06.030.

52. Zhang M, Wang C, Otto TD, Oberstaller J, Liao X, Adapa SR, Udenze K, Bronner IF, Casandra D, Mayho M, Brown J, Li S, Swanson J, Rayner JC, Jiang RHY, Adams JH. 2018. Uncovering the essential genes of the human malaria parasite Plasmodium falciparum by saturation mutagenesis. Science 360:eaap7847. https://doi.org/10.1126/science.aap7847.

53. Turner GD, Ly VC, Nguyen TH, Tran TH, Nguyen HP, Bethell D, Wyllie S, Louwrier K, Fox SB, Gatter KC, Day NP, Tran TH, White NJ, Berendt AR. 1998. Systemic endothelial activation occurs in both mild and severe malaria. Correlating dermal microvascular endothelial cell phenotype and soluble cell adhesion molecules with disease severity. Am J Pathol 152:1477–1487.

54. Möst J, Schwaeble W, Drach J, Sommerauer A, Dierich MP. 1992. Regulation of the expression of ICAM-1 on human monocytes and monocytic tumor cell lines. J Immunol 148:1635–1642.

55. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CA-K, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist P-H, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Pontén F. 2015. Tissue-based map of the human proteome. Science 347:1260419. https://doi.org/10.1126/science.1260419.

56. Immunological Genome Project. 2020. ImmGen at 15. Nat Immunol 21:700–703. https://doi.org/10.1038/s41590-020-0687-4.

57. Atamna H, Ginsburg H. 1993. Origin of reactive oxygen species in erythrocytes infected with Plasmodium falciparum. Mol Biochem Parasitol 61:231–241. https://doi.org/10.1016/0166-6851(93)90069-A.

58. Percário S, Moreira DR, Gomes BAQ, Ferreira MES, Gonçalves ACM, Laurindo PSOC, Vilhena TC, Dolabela MF, Green MD. 2012. Oxidative stress in malaria. Int J Mol Sci 13:16346–16372. https://doi.org/10.3390/ijms131216346.

59. Garcia CRS. 1999. Calcium homeostasis and signaling in the blood-stage malaria parasite. Parasitol Today 15:488–491. https://doi.org/10.1016/S0169-4758(99)01571-9.

60. Lourido S, Moreno SNJ. 2015. The calcium signaling toolkit of the Apicomplexan parasites Toxoplasma gondii and Plasmodium spp. Cell Calcium 57:186–193. https://doi.org/10.1016/j.ceca.2014.12.010.

61. Jimenez-Ruiz E, Morlon-Guyot J, Daher W, Meissner M. 2016. Vacuolar protein sorting mechanisms in apicomplexan parasites. Mol Biochem Parasitol 209:18–25. https://doi.org/10.1016/j.molbiopara.2016.01.007.

62. Vallejo AF, Read RC, Arevalo-Herrera M, Herrera S, Elliott T, Polak ME. 2018. Malaria systems immunology: Plasmodium vivax induces tolerance during primary infection through dysregulation of neutrophils and dendritic cells. J Infect 77:440–447. https://doi.org/10.1016/j.jinf.2018.09.005.

63. Aitken EH, Alemu A, Rogerson SJ. 2018. Neutrophils and malaria. Front Immunol 9:3005. https://doi.org/10.3389/fimmu.2018.03005.

64. Ortega-Pajares A, Rogerson SJ. 2018. The rough guide to monocytes in malaria infection. Front Immunol 9:2888. https://doi.org/10.3389/fimmu.2018.02888.

65. Tan QW, Mutwil M. 2020. Malaria.tools—comparative genomic and transcriptomic database for Plasmodium species. Nucleic Acids Res 48:D768–D775. https://doi.org/10.1093/nar/gkz662.

66. Li H, O'Donoghue AJ, van der Linden WA, Xie SC, Yoo E, Foe IT, Tilley L, Craik CS, da Fonseca PCA, Bogyo M. 2016. Structure- and function-based design of Plasmodium-selective proteasome inhibitors. Nature 530:233–236. https://doi.org/10.1038/nature16936.

67. Stokes BH, Yoo E, Murithi JM, Luth MR, Afanasyev P, da Fonseca PCA, Winzeler EA, Ng CL, Bogyo M, Fidock DA. 2019. Covalent Plasmodium falciparum-selective proteasome inhibitors exhibit a low propensity for generating resistance in vitro and synergize with multiple antimalarial agents. PLoS Pathog 15:e1007722. https://doi.org/10.1371/journal.ppat.1007722.

68. Breinich MS, Ferguson DJP, Foth BJ, van Dooren GG, Lebrun M, Quon DV, Striepen B, Bradley PJ, Frischknecht F, Carruthers VB, Meissner M. 2009. A dynamin is required for the biogenesis of secretory organelles in Toxoplasma gondii. Curr Biol 19:277–286. https://doi.org/10.1016/j.cub.2009.01.039.

69. Melatti C, Pieperhoff M, Lemgruber L, Pohl E, Sheiner L, Meissner M. 2019. A unique dynamin-related protein is essential for mitochondrial

fission in Toxoplasma gondii. PLoS Pathog 15:e1007512. https://doi.org/10.1371/journal.ppat.1007512.

70. Wang L, Delahunty C, Fritz-Wolf K, Rahlfs S, Helena Prieto J, Yates JR, Becker K. 2015. Characterization of the 26S proteasome network in Plasmodium falciparum. Sci Rep 5:17818. https://doi.org/10.1038/srep17818.

71. Aminake MN, Arndt H-D, Pradel G. 2012. The proteasome of malaria parasites: a multi-stage drug target for chemotherapeutic intervention? Int J Parasitol Drugs Drug Resist 2:1–10. https://doi.org/10.1016/j.ijpddr.2011.12.001.

72. Wu Y, Encheva V, Green JL, Lasonder E, Prommaban A, Kunzelmann S, Christodoulou E, Grainger M, Truongvan N, Bothe S, Sharma V, Song W, Pinzuti I, Uthaipibull C, Srichairatanakool S, Barault V, Langsley G, Schindelin H, Stieglitz B, Snijders AP, Holder AH. 2019. Protein ubiquitylation is essential for the schizont to merozoite transition in Plasmodium falciparum blood-stage development. bioRxiv https://doi.org/10.1101/755439.

73. Cai H, Lilburn TG, Hong C, Gu J, Kuang R, Wang Y. 2015. Predicting and exploring network components involved in pathogenesis in the malaria parasite via novel subnetwork alignments. BMC Syst Biol 9:S1. https://doi.org/10.1186/1752-0509-9-S4-S1.

74. Burgmann H, Looareesuwan S, Viravan C, Vanijanonta S, Zedwitz-Liebenstein K, Vorbach H, Graninger W. 1996. Serum laminin and basic fibroblast growth factor concentrations in patients with complicated Plasmodium falciparum malaria. J Clin Immunol 16:278–282. https://doi.org/10.1007/BF01541393.

75. Wenisch C, Graninger W, Viravan C, Looareesuwan S, Parschalk B, Wernsdorfer W. 1994. Serum laminin in malaria. J Clin Pathol 47:787–789. https://doi.org/10.1136/jcp.47.9.787.

76. Mahamar A, Attaher O, Swihart B, Barry A, Diarra BS, Kanoute MB, Cisse KB, Dembele AB, Keita S, Gamain B, Gaoussou S, Issiaka D, Dicko A, Duffy PE, Fried M. 2017. Host factors that modify Plasmodium falciparum adhesion to endothelial receptors. Sci Rep 7:13872. https://doi.org/10.1038/s41598-017-14351-7.

77. Gnädig NF, Stokes BH, Edwards RL, Kalantarov GF, Heimsch KC, Kuderjavy M, Crane A, Lee MCS, Straimer J, Becker K, Trakht IN, Odom John AR, Mok S, Fidock DA. 2020. Insights into the intracellular localization, protein associations and artemisinin resistance properties of Plasmodium falciparum K13. PLoS Pathog 16:e1008482. https://doi.org/10.1371/journal.ppat.1008482.

78. Birnbaum J, Scharf S, Schmidt S, Jonscher E, Hoeijmakers WAM, Flemming S, Toenhake CG, Schmitt M, Sabitzki R, Bergmann B, Fröhlke U, Mesén-Ramírez P, Blancke Soares A, Herrmann H, Bártfai R, Spielmann T. 2020. A Kelch13-defined endocytosis pathway mediates artemisinin resistance in malaria parasites. Science 367:51–59. https://doi.org/10.1126/science.aax4735.

79. Bridgford JL, Xie SC, Cobbold SA, Pasaje CFA, Herrmann S, Yang T, Gillett DL, Dick LR, Ralph SA, Dogovski C, Spillman NJ, Tilley L. 2018. Artemisinin kills malaria parasites by damaging proteins and inhibiting the proteasome. Nat Commun 9:3801. https://doi.org/10.1038/s41467-018-06221-1.

80. Tilley L, Straimer J, Gnädig NF, Ralph SA, Fidock DA. 2016. Artemisinin action and resistance in Plasmodium falciparum. Trends Parasitol 32:682–696. https://doi.org/10.1016/j.pt.2016.05.010.

81. Ng CL, Fidock DA, Bogyo M. 2017. Protein degradation systems as antimalarial therapeutic targets. Trends Parasitol 33:731–743. https://doi.org/10.1016/j.pt.2017.05.009.

82. Dogovski C, Xie SC, Burgio G, Bridgford J, Mok S, McCaw JM, Chotivanich K, Kenny S, Gnädig N, Straimer J, Bozdech Z, Fidock DA, Simpson JA, Dondorp AM, Foote S, Klonis N, Tilley L. 2015. Targeting the cell stress response of Plasmodium falciparum to overcome artemisinin resistance. PLoS Biol 13:e1002132. https://doi.org/10.1371/journal.pbio.1002132.

83. Liu J, Wetzel L, Zhang Y, Nagayasu E, Ems-McClung S, Florens L, Ke H. 2013. Novel thioredoxin-like proteins are components of a protein complex coating the cortical microtubules of Toxoplasma gondii. Eukaryot Cell 12:1588–1599. https://doi.org/10.1128/EC.00082-13.

84. Liu J, He Y, Benmerzouga I, Sullivan WJ, Jr, Morrissette NS, Murray JM, Hu K. 2016. An ensemble of specifically targeted proteins stabilizes cortical microtubules in the human parasite Toxoplasma gondii. Mol Biol Cell 27:549–571. https://doi.org/10.1091/mbc.E15-11-0754.

85. Chen C, Itakura E, Weber KP, Hegde RS, de Bono M. 2014. An ER complex of ODR-4 and ODR-8/Ufm1 specific protease 2 promotes GPCR maturation by a Ufm1-independent mechanism. PLoS Genet 10:e1004082. https://doi.org/10.1371/journal.pgen.1004082.

86. Dwyer ND, Troemel ER, Sengupta P, Bargmann CI. 1998. Odorant receptor localization to olfactory cilia is mediated by ODR-4, a novel membrane-associated protein. Cell 93:455–466. https://doi.org/10.1016/S0092-8674(00)81173-3.

87. Dellibovi-Ragheb TA, Jhun H, Goodman CD, Walters MS, Ragheb DRT, Matthews KA, Rajaram K, Mishra S, McFadden GI, Sinnis P, Prigge ST. 2018. Host biotin is required for liver stage development in malaria parasites. Proc Natl Acad Sci U S A 115:E2604–E2613. https://doi.org/10.1073/pnas.1800717115.

88. Sarafoglou K, Bentler K, Gaviglio A, Redlinger-Grosse K, Anderson C, McCann M, Bloom B, Babovic-Vuksanovic D, Gavrilov D, Berry SA. 2009. High incidence of profound biotinidase deficiency detected in newborn screening blood spots in the Somalian population in Minnesota. J Inherit Metab Dis 32:169–173. https://doi.org/10.1007/s10545-009-1135-7.

89. Mustière R, Vanelle P, Primas N. 2020. Plasmodial kinase inhibitors targeting malaria: recent developments. Molecules 25:5949. https://doi.org/10.3390/molecules25245949.

90. Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database Collaboration. 2011. The Sequence Read Archive. Nucleic Acids Res 39:D19–D21. https://doi.org/10.1093/nar/gkq1019.

91. Leinonen R, Akhtar R, Birney E, Bower L, Cerdeno-Tárraga A, Cheng Y, Cleland I, Faruque N, Goodgame N, Gibson R, Hoad G, Jang M, Pakseresht N, Plaister S, Radhakrishnan R, Reddy K, Sobhany S, Ten Hoopen P, Vaughan R, Zalunin V, Cochrane G. 2011. The European Nucleotide Archive. Nucleic Acids Res 39:D28–D31. https://doi.org/10.1093/nar/gkq967.

92. Ogasawara O, Kodama Y, Mashima J, Kosuge T, Fujisawa T. 2020. DDBJ Database updates and computational infrastructure enhancement. Nucleic Acids Res 48:45–50. https://doi.org/10.1093/nar/gkz982.

93. Zhu Y, Stephens RM, Meltzer PS, Davis SR. 2013. SRAdb: query and use public next-generation sequencing data from within R. BMC Bioinformatics 14:19. https://doi.org/10.1186/1471-2105-14-19.

94. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Oleś AK, Pagès H, Reyes A, Shannon P, Smyth GK, Tenenbaum D, Waldron L, Morgan M. 2015. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods 12:115–121. https://doi.org/10.1038/nmeth.3252.

95. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. https://doi.org/10.1093/bioinformatics/bts635.

96. Dobin A, Gingeras TR. 2015. Mapping RNA-seq reads with STAR. Curr Protoc Bioinformatics 51:11.14.1–11.14.19. https://doi.org/10.1002/0471250953.bi1114s51.

97. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. 2013. Software for computing and annotating genomic ranges. PLoS Comput Biol 9:e1003118. https://doi.org/10.1371/journal.pcbi.1003118.

98. Knijnenburg TA, Wessels LFA, Reinders MJT, Shmulevich I. 2009. Fewer permutations, more accurate P-values. Bioinformatics 25:i161–i168. https://doi.org/10.1093/bioinformatics/btp211.

99. Phipson B, Smyth GK. 2010. Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. Stat Appl Genet Mol Biol 9:Article39. https://doi.org/10.2202/1544-6115.1585.

100. Jaccard P. 1912. The distribution of the flora in the alpine zone. New Phytol 11:37–50. https://doi.org/10.1111/j.1469-8137.1912.tb05611.x.

101. Koonin EV. 2001. An apology for orthologs - or brave new memes. Genome Biol 2:COMMENT1005. https://doi.org/10.1186/gb-2001-2-4-comment1005.

102. Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20:238. https://doi.org/10.1186/s13059-019-1832-y.

103. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

104. van Dongen S, Abreu-Goodger C. 2012. Using MCL to extract clusters from networks. Methods Mol Biol 804:281–295. https://doi.org/10.1007/978-1-61779-361-5_15.

105. Li J, Wang S, Chen Z, Wang Y. 2020. A bipartite network module-based project to predict pathogen-host association. Front Genet 10:1357. https://doi.org/10.3389/fgene.2019.01357.

106. Csardi G, Nepusz T. 2006. The Igraph software package for complex network research. InterJournal 2006, Complex Systems, 1695.

107. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. https://doi.org/10.1101/gr.1239303.

108. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. 2014. UpSet: visualization of intersecting sets. IEEE Trans Vis Comput Graph 20:1983–1992. https://doi.org/10.1109/TVCG.2014.2346248.

109. Gaudet P, Škunca N, Hu JC, Dessimoz C. 2017. Primer on the Gene Ontology. Methods Mol Biol 1446:25–37. https://doi.org/10.1007/978-1-4939-3743-1_3.

110. Alexa A, Rahnenfuhrer J. 2007. Gene set enrichment analysis with topGO.

111. Bahl A, Brunk B, Crabtree J, Fraunholz MJ, Gajria B, Grant GR, Ginsburg H, Gupta D, Kissinger JC, Labo P, Li L, Mailman MD, Milgram AJ, Pearson DS, Roos DS, Schug J, Stoeckert CJ, Jr, Whetzel P. 2003. PlasmoDB: the Plasmodium Genome Resource. A database integrating experimental and computational data. Nucleic Acids Res 31:212–215. https://doi.org/10.1093/nar/gkg081.

112. R Core Team. 2020. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.