Dual RNA-Seq meta-analysis in *Plasmodium* infection

Parnika Mukherjee[1] and Emanuel Heitlinger[1,2]

[1]Department of Molecular Parasitology, Humboldt University Berlin, Philippstr 13, Haus 14, 10115 Berlin, Germany

[2]Leibniz-Institut für Zoo- und Wildtierforschung (IZW), Alfred-Kowalke-Straße 17, 10315 Berlin, Germany

March 13, 2019

Abstract

Dual RNA-Seq is the simultaneous analysis of host and parasite transcriptomes. This approach can identify host-parasite interactions by correlated gene expression. Co-expression might highlight interlinked signaling, metabolic or gene regulatory pathways in addition to potentially physically interacting proteins. Numerous studies have used gene expression data to investigate *Plasmodium* infection causing malaria. Usually such studies focus on one organism – either the host or the parasite – and the other is considered "contaminant". Dual RNA-Seq, in contrast, follows the rationale that cross-species interactions determine not only virulence of the parasite but also tolerance, resistance or susceptibility of the host.

Here we propose a meta-analysis approach for dual RNA-Seq. We screened malaria transcriptome experiments for studies providing gene expression data from both *Plasmodium* and its host. Out of 105 malaria studies in *Homo sapiens, Macaca mulatta* and *Mus musculus*, we identified 56 studies with the potential to provide host and parasite data. While 15 studies (1935 total samples) of these 56 explicitly aimed to generate dual RNA-Seq data, 41 (1129 samples) had an original focus on either the host or the parasite. We show that a total of up to 2530 samples are suitable for dual RNA-Seq analysis providing an unexplored potential for meta analysis.

We argue that the multitude of variations in experimental conditions found in the selected studies should help narrow down a conserved core of cross-species interactions. Different hosts used as laboratory models for human malaria infection are infected by evolutionarily diverse species of genus *Plasmodium*. We propose that a conserved core of interacting pathways and co-regulated genes might be identified using overlying interaction networks of different host-parasite species pairs based on orthologous genes. Our approach might also provide the opportunity to gauge the applicability of model systems for different pathways in malaria studies.

Introduction

Transcriptomes are often analysed in a first attempt to understand cellular and organismic events, because a comprehensive profile of RNA expression can be obtained at reasonable cost and with high technical accuracy [1]. Microarrays dominated transcriptomics for over ten years since 1995 [2–4]. Microarrays quantify gene expression based on hybridisation of a target sequence to an immobilised probe of known sequence. Technical difficulties associated with microarrays lie in probe selection, cross-hybridization, and design cost of custom chips [5]. RNA sequencing (RNA-Seq) eliminates these difficulties and provides deep and accurate expression estimates for all RNAs in a sample. RNA-Seq has thus replaced microarrays as the predominant tool for transcriptomics [1, 6]. RNA-Seq assesses host and parasite transcriptomes simultaneously, if RNA of both organisms is contained in a sample. Virulence of infectious disease is often a result of interlinked processes of both host and pathogen ("host-pathogen interactions") and it has been proposed to analyse transcriptomes of both organisms involved in an infection to obtain a more complete understanding of disease [5–7]. This approach is called dual RNA-Seq.

In case of malaria, unlike in bacterial infections, both the pathogen and the host are eukaryotic organisms with similar transcriptomes. Host and parasite mRNA is selected simultaneously when poly-dT priming is used to amplify polyadenylated transcripts [5, 6]. This makes most malaria transcriptome datasets potentially suitable for dual RNA-Seq analysis. Malaria research, especially transcriptomics, is traditionally designed to target one organism, either the host or the parasite. Expression of mRNA, for example, can be compared between different time points in the life cycle of *Plasmodium* or between different drug treatment conditions. In the mammalian intermediate host, *Plasmodium* invades first liver and then red blood cells (RBCs) for development and asexual expansion. While the nuclear machinery of cells from both host and parasite produces mRNA in the liver, RBCs are enucleated and transcriptionally inactive in mammalian host. In blood infections leukocytes are thus the source of host mRNA. Researchers conducting a targeted experiment might regard transcripts from the non-target organism as "contamination". Nevertheless, expression of those transcripts potentially responds to stimuli during the investigation. Additionally, some recent studies on malaria make intentional use dual RNA-Seq. Malaria is the most thoroughly investigated disease caused by an eukaryotic organism and accumulation of these two kinds of studies, RNA-Seq with "contaminants"

and intentional dual RNA-Seq, provides a rich resource for meta-analysis.

Such a meta analysis can use co-regulated gene expression to infer host-parasite interactions. Correlation of mRNA expression can be indicative of different kinds of biological "interactions": On one hand, protein products could be directly involved in the formation of complexes and might therefore be produced at quantities varying similarly under altered conditions. On the other, involvement in the same biological pathways can result in co-regulated gene expression without physical interaction. This broad concept of interaction has long been exploited in single organisms (e.g. [**?**, 8–10]). We (and others before [11]) propose to extrapolate this to interactions between the host and pathogen. 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 co-regulated at different time points or under different conditions.

In this paper we explore first steps in a comparative meta-analysis of dual RNA-Seq transcriptomes. Existing raw read datasets collectively present an unexplored potential to answer questions that have not been investigated by individual studies. Meta-analysis increases the number of observations and statistical power and helps eliminate false positives and true negatives which may otherwise conceal important biological inferences [12–14]. Since mice- and macaque-malaria are often used as laboratory models for human malaria, we analyse the availability and suitability of mRNA sequencing data from three evolutionarily close hosts - *Homo sapiens, Macaca mulatta* and *Mus musculus* - and their associated *Plasmodium* parasites. We summarize available data, challenges and approaches to obtain host-parasite interactions and discuss orthology across different host-parasite systems as a means to enrich information.

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), EBI Sequence Read Archive (ERA) and DDBJ Sequence Read Archive (DRA). Comprehensive query tools to access these databases via web interfaces and programmatically via scriptable languages exist (for example, SRAdb,

ENAbrowseR). In these databases, all experiments submitted under a single accession are given a single "study accession number" and are collectively referred to as a "study" here onwards.

We used SRAdb [15], a Bioconductor/R package [16, 17], to query SRA [18, 19] 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" given as "RNA-Seq" and "Plasmodium" in study title, abstract or sample attributes using the "dbGetQuery" function. 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 SRA, PubMed and Google Scholar. During this search, we disregarded 91 studies, all of which provide data from vectors and non-target hosts (e.g. avian malaria). 49 more studies were excluded because their gene expression data was derived from *Plasmodium. spp* cultures in erythrocytes, blood or RPMI and thus can be expected to be devoid of host mRNA. We then used the SRAdb Bioconductor/R package, and the prefetch and fastq-dump functions from SRAtoolkit, to download all replicate samples (called "runs" in the databases) of the selected studies. The curation of studies and the download was performed on 21 January, 2019.

In total we found 56 potentially suitable studies in this database and literature review. The host organism for 22 studies was *Homo sapiens*, for 24, *Mus musculus* and for 10, *Macaca mulatta*. The corresponding infecting parasites were *P. falciparum, P. vivax* and *P. berghei* in human studies (including four artificial infections of human liver cell culture with *P. berghei*), *P. yoelii*, *P. chabaudi* and *P. berghei* in mouse studies and *P. cynomolgi* and *P. coatneyi* in macaque studies(table 1).

We note that 20 of the 56 studies depleted (or enriched, respectively) specific classes of cells from their samples. Some studies, for example, targeted the parasite using vaccines derived from sporozoites during liver infection [20–22]. Such infection is physiologically asymptomatic and a low number of parasites cells [23] makes it difficult to study *Plasmodium* transcriptomes in this stage. To reduce overwhelming host RNA levels, 3 out of 10 liver studies sorted infected hepatoma cells from uninfected cells. Similarly, 17 other studies have depleted or enriched host WBCs (leukocytes) to focus expression analysis on *Plasmodium* or the host immune system, respectively. In all these scenarios, we suspect depletion to be imperfect and thus the samples to potentially include mRNA of both or-

ganisms. We note, however, that host gene-expression for WBC-depleted samples might be problematic as incomplete depletion might affect different types of WBCs differentially and hence bias the detectable host mRNA expression in the direction of less-depleted cell types. For the similar reasons, parasite depletion might be challenging to control for.

For 15 out of the 56 studies the authors state that they intended to simultaneously study host and parasite transcriptomes ("dual RNA-Seq"). This includes 8 studies from MaHPIC (Malaria Host-Pathogen Interaction Center), based at Emory University, that made extensive omics measurements in macaque malaria. The original focus of the remaining 41 studies was on the parasite in 20 and on the host in 21 cases.

Plasmodium parasites sequester in bone marrow, adipose tissue, lung, spleen and brain (the latter causing cerebral malaria) [24, 25]. To study a comprehensive spectrum of host-parasite interactions it would be optimal to have data from these different tissues. Our collection of studies represent data derived from blood and liver for all three host organisms. In addition, we have seven spleen studies ([26–32]) and two studies of cerebral malaria ([33,34]) from mice. MaHPIC offers a collection of blood and bone marrow studies in macaques.

Experiments performed on mouse blood focus on the parasite instead of the host (11 vs. 0). Studies on human blood infection focus more often on the host immune response than on the parasite (9 vs. 5). 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 cell cultures (*in vitro*). We, here, argue that small clusters of genes co-expressed across several of such diverse conditions might help to point towards potentially novel core host-parasite interactions.

Dual RNA-Seq suitability analysis

A sample (experimental replicate or "run" in the jargon of sequencing databases) suitable for dual RNA-Seq analysis must provide "sufficient" gene expression from both host and parasite. To assess the proportion for host and parasite RNA sequencing reads in each study and sample we mapped sequencing reads onto concatenated host and parasite reference genomes using STAR [35,36]. Simultaneous mapping against both genomes should avoid non-specific 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 [37] and calculated the proportion of reads mapping to host and parasite genes.

The proportions of host and parasite reads for each run does not always reflect the original focus of a study (fig. 1a). Studies using no depletion or enrichment give us an idea how skewed overall RNA expression is towards one organism under native conditions: in studies on blood stage infections, the original focus is mostly on immune gene expression from leukocytes. In the respective samples the number of host reads is often overwhelming unless parasitemia is very high, like in studies originally designed to use a dual RNA-Seq approach on blood stages. Samples with lower parasitemia are mostly not suitable for dual RNA-Seq analysis (table 1).

Many studies using depletion or enrichment prior to RNA sequencing ("enriched/depleted" in fig. 1a) show considerable expression of the non-target organism. Studies on liver infection, for example, [38] and [39], comprise several runs with balanced proportions of host and parasite reads. This is a result of infected liver cells being sorted from uninfected cells in culture. While the parasite has been the original target organism in most studies they provide data suitable for dual RNA-Seq. Studies depleting whole blood from leukocytes to focus on parasite transcriptomes still show considerable host gene expression and provide principally suitable runs for the analysis of blood infection at lower intensities. The latter comes with the caveat that host expression might be biased by unequal depletion of particular cell types.

To establish suitability thresholds for inclusion of individual samples (runs) in further analysis we plotted the number of host and parasite reads against the number of host and parasite genes expressed (fig. 1b and fig. 1c). For runs with high sequencing depth the total number of expressed genes of the host and parasite approaches the number of annotated genes: around 30000 for the mammalian host and around 4500 for *Plasmodium*. When sequencing depth is lower, the number of genes detected as expressed is lower and a decrease in sensitivity can be expected to prevent analysis of lowly expressed genes. We propose four parameters for suitability thresholds in dual RNA-Seq analysis: the number reads mapping to host (1) and (2) parasite genes and the number of genes these reads map to (expressed genes) in host and parasite (3, 4). In table 1, we give the number of runs considered suitable for three different combinations of thresholding. Without claiming a particular thresholds to be ideal we propose to use thresholds to avoid uninformative runs in further processing to reduce the computational burden of co-expression analysis.

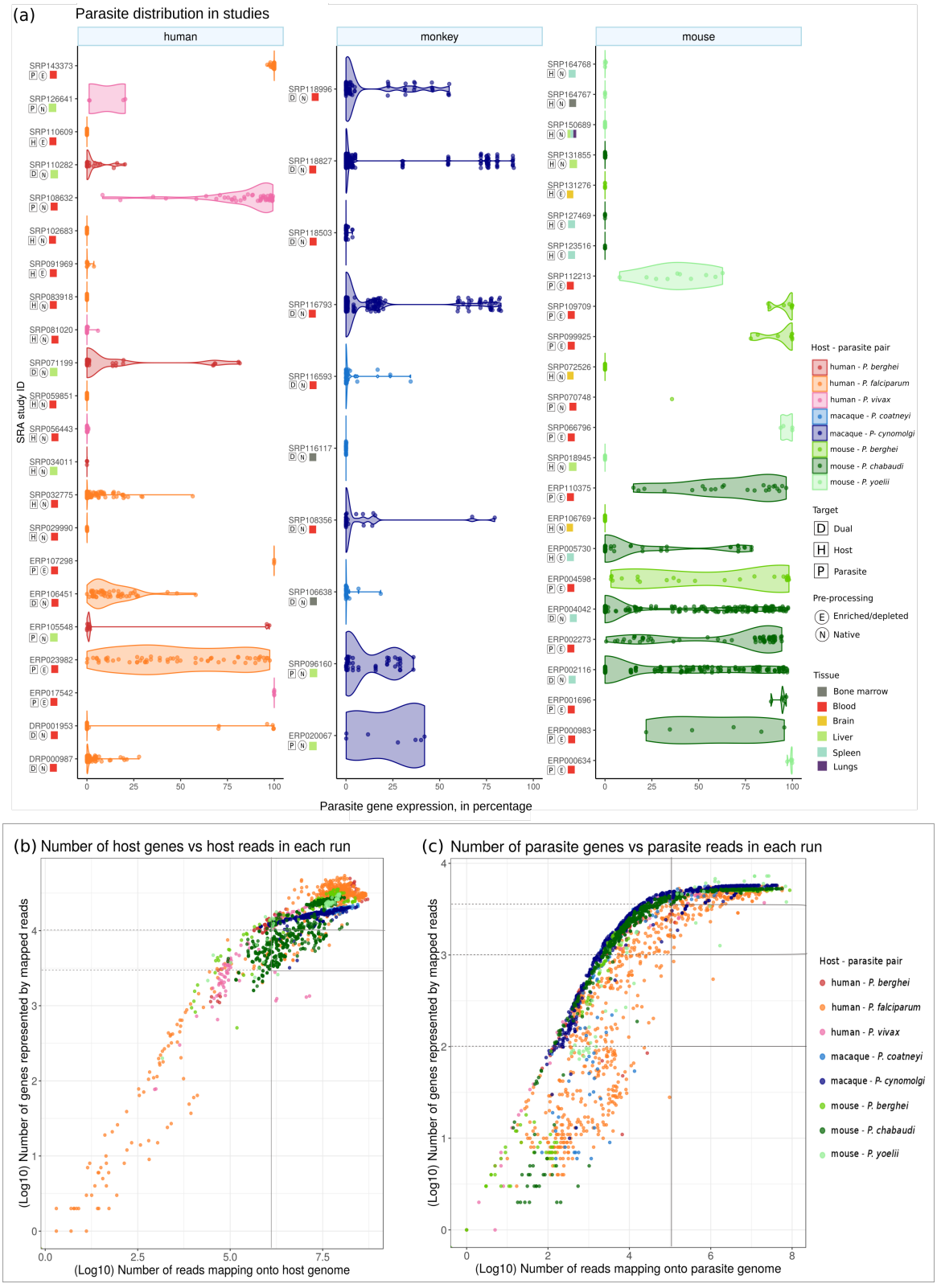

Figure 1: Proportion and number of sequencing reads and expressed genes from parasite and host in selected malaria RNA-Seq studies. We mapped sequencing reads from studies selected for their potential to provide both host and parasite gene expression data (studies N=56, total runs n=3064) mapped against appropriate host and parasite genomes. (a) The percentage of parasite reads (x-axis) is plotted for runs in each study. The studies are categorised according to the host organisms and labeled "enriched/depleted" 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 (a) and parasite (b). The number of expressed genes increases with sequencing depth towards the maximum of all annotated genes for the respective organism. The vertical lines indicate a threshold of 1.000.000 and 100.000 reads for host and parasite, respectively. The horizontal lines correspond to thresholds on the number of expressed genes at 10.000 and 3.000 for host and at 1.000 and 100 for the parasite. At such exemplary thresholds data could be considered sufficient for dual RNA-Seq analysis on both organisms.

Table 1: Number of studies for each host-parasite pair and suitability analysis of their runs.

						Run suitability analysis from all studies					
Host	Parasite	# total studies	# total runs	# depletion studies	# depletion runs	# host genes	# parasite genes	no threshold on read count		prc[a] >= 10^5 hrc[b] >= 10^6	
								# studies	# runs	# studies	# runs
Human	*P. falciparum*	13	840	5	216	3000	100	11	523	6	270
						3000	1000	11	362	6	268
						10000	3000	7	219	5	207
Human	*P. berghei*	4	77	4	77	3000	100	3	60	2	16
						3000	1000	3	42	2	16
						10000	3000	3	25	2	16
Human	*P. vivax*	5	135	1	5	3000	100	3	39	1	1
						3000	1000	3	36	1	1
						10000	3000	2	4	1	1
Mouse	*P. berghei*	7	128	4	50	3000	100	6	74	4	23
						3000	1000	6	54	4	23
						10000	3000	4	24	4	22
Mouse	*P. chabaudi*	10	926	7	174	3000	100	8	869	7	341
						3000	1000	7	801	7	341
						10000	3000	7	252	6	109
Mouse	*P. yoelii*	7	65	3	19	3000	100	7	34	2	8
						3000	1000	6	21	2	8
						10000	3000	2	7	2	6
Macaque	*P. coatneyi*	3	70	0	0	3000	100	3	74	3	26
						3000	1000	3	53	3	26
						10000	3000	3	33	3	26
Macaque	*P. cynomolgi*	7	849	0	0	3000	100	7	857	7	375
						3000	1000	7	706	7	357
						10000	3000	7	514	7	365

[a]prc: parasite read count
[b]hrc: host read count

Suitable runs at the thresholds chosen here are identified from human-*P. falciparum*, monkey-*P. cynomolgi*, human-*P. berghei* and mouse-*P. berghei* systems. Unfortunately, with current thresholds and currently available data, we highly under-represent human-*P. vivax* and human-*P. berghei* systems, the two liver *in vitro* models. This outcome is understandable owing to the low parasitemia in liver cultures [40]. We note that the thresholds could further be made lenient enough to include more runs for these systems at the cost of analysing only the most highly expressed parasite genes. An alternative approach relies on depleted/enriched samples for these systems. For further analysis, however, we it could prove challenging to include depleted/enriched samples as discussed before. Analysis approaches such as multilayer networks (see below) might help to gauge problems with such runs for the inference of co-expression in further steps of analysis.

Identification of co-expressed genes via correlation techniques

Some genes are likely to show almost uniform expression under different experimental conditions ("housekeeping genes"). Naive assessments of correlation could, however, identify pairs of such genes as highly correlated. An analysis of co-expression can deal with this challenge in two different ways:

Firstly, the most variable genes within and across studies can be selected and other genes discarded. While requiring little computational time and resources, exclusion of genes with too little variance in expression from downstream analysis should be performed with caution, as seemingly small variations might result in a suitable signal over a large set of runs. To select only variable genes, one option is to compute their variance across all samples (in one or multiple studies). Genes with variance below a threshold may then be excluded from further analysis. As variance increases with the mean for gene expression data, the Biological Coefficient of Variation (BCV) [41, 42] may provide a more robust threshold. Secondly, one can compute empirical correlation indices, similar to p-values, for any gene-pair. Empirical p-values are 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) [43, 44]. These methods construct a null distribution using permutations of the given data instead of assuming a null distribution

in advance. Since host and parasite genomes total nearly 30,000 genes, the number of permutations has to be around 1.6×10^9 to be suitable for corrections for multiple-testing. Alternatively, as computational costs for these permutations can be expected to be too high for datasets with thousands of samples, non-corrected "p-values" may be considered a ranking for host-parasite gene correlation, following the suggestion of Reid and Berriman [11]. Nevertheless, reliance on empirical computation of p-values without prior variance/BCV filtering might become impracticable for very large datasets in the proposed meta- analysis.

We consider partial correlation as an additional approach that could be combined with the above methods. Partial correlation can control pairwise correlations for the influence of other genes [45]. In transcriptomic applications full-conditioned partial correlation is computationally very expensive. Some studies therefore resort to second-order partial correlation (relationship between two genes independent of two other genes) [46–48]. A suitable pipeline might first use (zero-order partial, that is "regular") correlation with empirical p-values to remove constitutively expressed gene-pairs. For all correlations with an empirical "p-value" below a certain threshold, one could compute e.g. first-order partial correlations reducing the number of computations. Iterations of such an approach with higher-order partial correlations are then possible.

Across different studies; across different host-parasite systems

Gene × *gene* matrices obtained from correlation analysis can be visualised and analysed as interaction networks. We have identified different but interlinked workflows to reconstruct a consensus network of expression correlation (fig. 2). A first approach (fig. 2(a)) integrates data from different studies of one host-parasite system by simply appending expression profiles of their runs.

Knowledge of 1:1 orthologs [49] between different host and different parasite species can be used in the next steps to integrate across different host-parasite systems. Humans and macaques share 18179 1:1 orthologous genes, humans and mice share 17089 orthologous genes and 14776 genes are 1:1:1 orthologs among all three species. Similarly, 7760 groups of orthologous genes exist among the *Plasmodium* species. A simple approach

to combine data across host-parasite systems could again append those orthologs in the original datasets before correlations of gene expression.

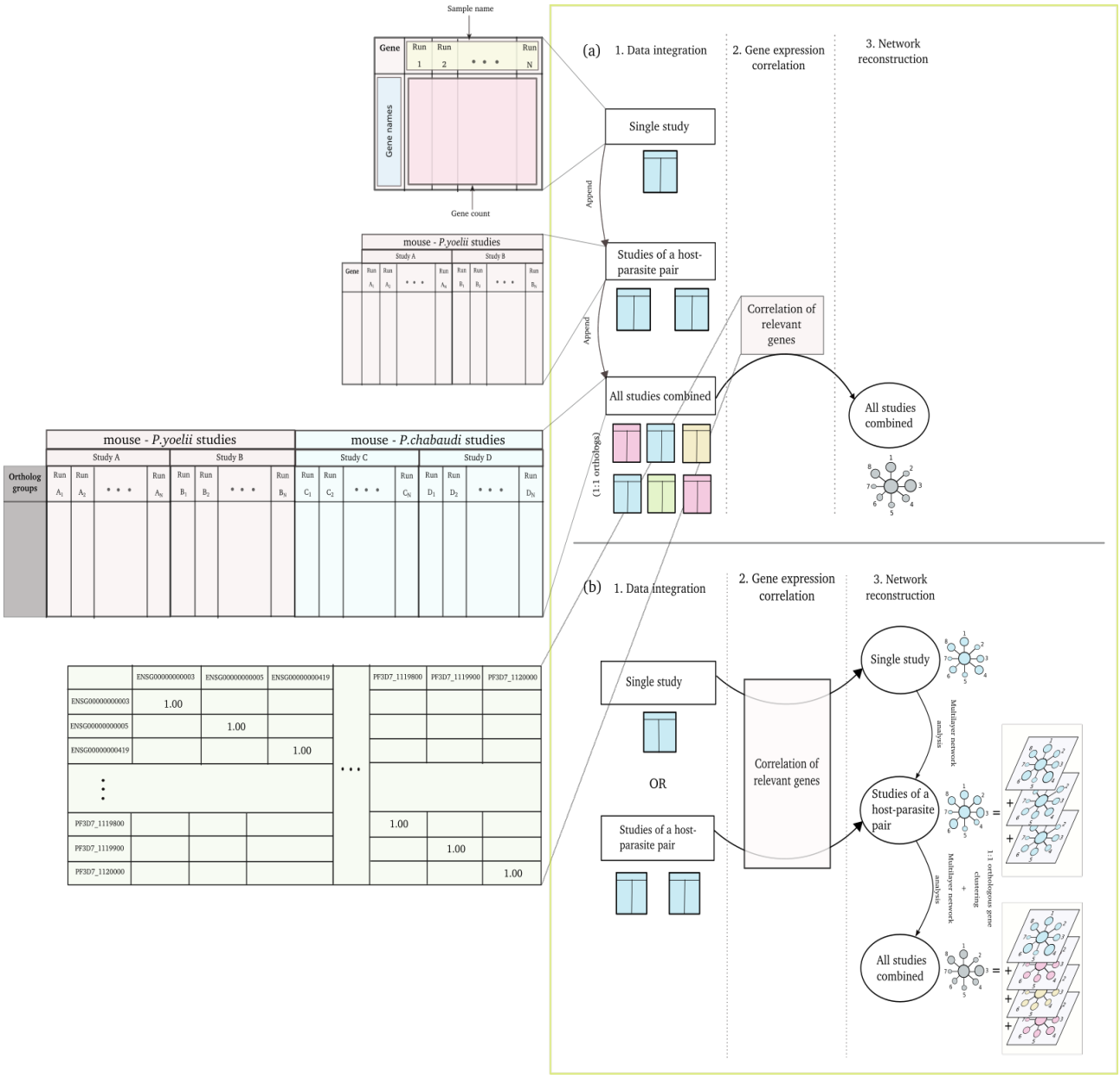

Figure 2: Two strategies identified to reconstruct host-parasite interaction networks from SRA. We identified two approaches to obtain a consensus network involving multiple hosts and multiple parasites. We selected appropriate studies from SRA for this analysis. The aim is to find a set of important interactions in malaria using co-regulated gene expression and visualising this information as a biological network. Using the first approach (figure (a)), we form single networks from single RNA-seq datasets or single networks from all studies of a host-parasite system appended one after the other, using cross-species gene correlation. To obtain a consensus network for all hosts and all parasites, we use 1:1 orthologous genes names for all hosts and all parasites, rename these genes to show their equivalency and append them to form one big dataset. Next we perform pairwise correlation of genes and finally, a network that will represent the direct interactions among orthologous genes. In (b), the second approach, we implement multi-layered network analysis to obtain a consensus network from several layers of individual networks. In this approach, we make single networks for individual RNA-seq datasets. To obtain a network for a host-parasite system, we either append all datasets of the host-parasite system with each other and form a network, or we apply multi-layered network analysis on single networks to get the consensus. To reconstruct a network involving multiple host-parasite systems, we rename orthologous genes in each layer and then look for overlapping communities.

Alternatively, to construct a consensus network involving all hosts and parasites, a multi-layer network analysis could align networks by orthologous genes. This approach

can offer more control when looking for similar correlation in different layers representing different host-parasite systems. Similarly, more insight could be possible when correlations from different types of tissues are combined as multilayer networks. This would only require the construction of networks for a single host-parasite system and multi-layered network analysis on networks from single studies of the same host-parasite system.

We hope correlation between host and parasite transcript expression to highlight host-parasite interactions worth scrutiny of further focussed research. As a second goal, meta-analysis involving different host-parasite systems could give insights into how easily other insights obtained in malaria models can be translated to human malaria. If e.g. certain groups of pathways show lower evolutionary conservation in host-parasite co-expression networks, one could expect results on those to be harder to translate between systems. Finally, one can ask whether expression correlation between host and parasite species is more or less evolutionarily conserved than within host species [50–52].

References

[1] S. Martin, C. Dehler, and E. Krol, "Transcriptomic responses in the fish intestine," *Developmental Comparative Immunology*, vol. 64, 03 2016.

[2] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown, "Quantitative monitoring of gene expression patterns with a complementary dna microarray," *Science*, vol. 270, no. 5235, pp. 467–470, 1995.

[3] R. E. Hayward, J. L. DeRisi, S. Alfadhli, D. C. Kaslow, P. O. Brown, and P. K. Rathod, "Shotgun dna microarrays and stage-specific gene expression in plasmodium falciparum malaria," *Molecular Microbiology*, vol. 35, no. 1, pp. 6–14, 2000.

[4] A. Craig and A. Scherf, "Transcriptional analysis in plasmodium falciparum," *Trends in Microbiology*, vol. 8, no. 8, pp. 350 – 351, 2000.

[5] A. J. Westermann, L. Barquist, and J. Vogel, "Resolving host–pathogen interactions by dual rna-seq," *PLOS Pathogens*, vol. 13, pp. 1–19, 02 2017.

[6] A. Westermann, S. A Gorski, and J. Vogel Prof. Dr, "Dual rna-seq of pathogen and host," *Nature reviews. Microbiology*, vol. 10, pp. 618–30, 08 2012.

[7] H. J. Lee, A. Georgiadou, M. Walther, D. Nwakanma, L. B. Stewart, M. Levin, T. D. Otto, D. J. Conway, L. J. Coin, and A. J. Cunnington, "Integrated pathogen load and dual transcriptome analysis of systemic host-pathogen interactions in severe malaria," *Science Translational Medicine*, vol. 10, no. 447, 2018.

[8] J. M. Stuart, E. Segal, D. Koller, and S. K. Kim, "A gene-coexpression network for global discovery of conserved genetic modules," *Science*, vol. 302, no. 5643, pp. 249–255, 2003.

[9] H. Bono and Y. Okazaki, "Functional transcriptomes: comparative analysis of biological pathways and processes in eukaryotes to infer genetic networks among transcripts," *Current Opinion in Structural Biology*, vol. 12, no. 3, pp. 355 – 361, 2002.

[10] H. K. Lee, A. K. Hsu, J. Sajdak, J. Qin, and P. Pavlidis, "Coexpression analysis of human genes across many microarray data sets," *Genome Research*, vol. 14, year = 2004, doi = https://doi.org/10.1101/gr.1910904.

[11] A. J. Reid and M. Berriman, "Genes involved in host–parasite interactions can be revealed by their correlated expression," *Nucleic Acids Research*, vol. 41, no. 3, pp. 1508–1518, 2013.

[12] Fayyad, P.-S. Usama, S. Gregory, and Padhraic, "From data mining to knowledge discovery in databases," *AI Magazine*, vol. 17, 1996.

[13] A. Haidich, "Meta-analysis in medical research," *Hippokratia*, vol. 14, Dec 2011.

[14] R. H. Fagard, J. A. Staessen, and L. Thijs, "Advantages and disadvantages of the meta-analysis approach," *Journal of hypertension*, vol. 14, Oct 1996.

[15] Y. Zhu, R. M. Stephens, P. S. Meltzer, and S. R. Davis, "Sradb: query and use public next-generation sequencing data from within r," *BMC bioinformatics*, vol. 14, no. 1, p. 19, 2013.

[16] R. Ihaka and R. Gentleman, "R: A language for data analysis and graphics," *Journal of Computational and Graphical Statistics*, vol. 5, no. 3, pp. 299–314, 1996.

[17] W. Huber, V. J. Carey, R. Gentleman, S. Anders, M. Carlson, B. S. Carvalho, H. C. Bravo, S. Davis, L. Gatto, T. Girke, R. Gottardo, F. Hahne, K. D. Hansen, R. A. Irizarry, M. Lawrence, M. I. Love, J. MacDonald, V. Obenchain, A. K. Ole's, H. Pag'es, A. Reyes, P. Shannon, G. K. Smyth, D. Tenenbaum, L. Waldron, and M. Morgan, "Orchestrating high-throughput genomic analysis with Bioconductor," *Nature Methods*, vol. 12, no. 2, pp. 115–121, 2015.

[18] R. Leinonen, H. Sugawara, M. Shumway, and on behalf of the International Nucleotide Sequence Database Collaboration, "The sequence read archive," *Nucleic Acids Research*, vol. 39, no. supplement1, pp. D19 – D21, 2011.

[19]

[20] M. R. Hollingdale and M. Sedegah, "Development of whole sporozoite malaria vaccines," *Expert Review of Vaccines*, vol. 16, no. 1, pp. 45–54, 2017. PMID: 27327875.

[21] R. A. Seder, L.-J. Chang, M. E. Enama, K. L. Zephir, U. N. Sarwar, I. J. Gordon, L. A. Holman, E. R. James, P. F. Billingsley, A. Gunasekera, A. Richman, S. Chakravarty, A. Manoj, S. Velmurugan, M. Li, A. J. Ruben, T. Li, A. G. Eappen, R. E. Stafford, S. H. Plummer, C. S. Hendel, L. Novik, P. J. M. Costner, F. H. Mendoza, J. G. Saunders, M. C. Nason, J. H. Richardson, J. Murphy, S. A. Davidson, T. L. Richie, M. Sedegah, A. Sutamihardja, G. A. Fahle, K. E. Lyke, M. B. Laurens, M. Roederer, K. Tewari, J. E. Epstein, B. K. L. Sim, J. E. Ledgerwood, B. S. Graham, S. L. Hoffman, and , "Protection against malaria by intravenous

immunization with a nonreplicating sporozoite vaccine," *Science*, vol. 341, no. 6152, pp. 1359–1365, 2013.

[22] E. M. Bijker, S. Borrmann, S. H. Kappe, B. Mordmüller, B. K. Sack, and S. M. Khan, "Novel approaches to whole sporozoite vaccination against malaria," *Vaccine*, vol. 33, no. 52, pp. 7462 – 7468, 2015. Malaria Vaccines 2015.

[23] G. LaMonte, P. Orjuela-Sanchez, L. Wang, S. Li, J. Swann, A. Cowell, B. Y. Zou, A. Abdel-Haleem Mohamed, Z. Villa-Galarce, M. Moreno, C. Tong-Rios, J. Vinetz, N. Lewis, and E. A. Winzeler, "Dual rnaseq shows the human mucosal immunity protein, muc13, is a hallmark of plasmodium exoerythrocytic infection," *bioRxiv*, 2017.

[24] B. Franke-Fayard, J. Fonager, A. Braks, S. M. Khan, and C. J. Janse, "Sequestration and tissue accumulation of human malaria parasites: Can we learn anything from rodent models of malaria?," *PLOS Pathogens*, vol. 6, pp. 1–10, 09 2010.

[25] P. H. e. a. David, "Parasite sequestration in plasmodium falciparum malaria: spleen and antibody modulation of cytoadherence of infected erythrocytes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 80, 1983.

[26] S. M. Lai, J. Sheng, P. Gupta, L. Renia, K. Duan, F. Zolezzi, K. Karjalainen, E. W. Newell, and C. Ruedl, "Organ-specific fate, recruitment, and refilling dynamics of tissue-resident macrophages during blood-stage malaria," *Cell Reports*, vol. 25, no. 11, pp. 3099 – 3109.e3, 2018.

[27] "Erp002116." https://www.ncbi.nlm.nih.gov/sra/?term=ERP002116. Accessed: 21-01-2019.

[28] "Erp004042." https://www.ncbi.nlm.nih.gov/sra/?term=ERP004042. Accessed: 21-01-2019.

[29] "Srp164768." https://www.ncbi.nlm.nih.gov/sra/?term=SRP164768. Accessed: 21-01-2019.

[30] L. Gabryšová, M. Alvarez-Martinez, R. Luisier, L. S. Cox, J. Sodenkamp, C. Hosking, D. Pérez-Mazliah, C. Whicher, Y. Kannan, K. Potempa, X. Wu, L. Bhaw, H. Wende, M. H. Sieweke, G. Elgar, M. Wilson, J. Briscoe, V. Metzis, J. Langhorne, N. M. Luscombe, and A. O'Garra, "c-maf controls immune responses by regulating disease-specific gene networks and repressing il-2 in cd4+ t cells," vol. 19, 2018.

[31] M. R. Mamedov, A. Scholzen, R. V. Nair, K. Cumnock, J. A. Kenkel, J. H. M. Oliveira, D. L. Trujillo, N. Saligrama, Y. Zhang, F. Rubelt, D. S. Schneider, Y. hsiu Chien, R. W. Sauerwein, and M. M. Davis, "A macrophage colony-stimulating-factor-producing t cell subset prevents malarial parasitemic recurrence," *Immunity*, vol. 48, no. 2, pp. 350 – 363.e7, 2018.

[32] "Erp005730." https://www.ncbi.nlm.nih.gov/sra/?term=ERP005730. Accessed: 21-01-2019.

[33] S. Torre, M. J. Polyak, D. Langlais, N. Fodil, J. M. Kennedy, I. Radovanovic, J. Berghout, G. A. Leiva-Torres, C. M. Krawczyk, S. Ilangumaran, K. Mossman, C. Liang, K.-P. Knobeloch, L. M. Healy, J. Antel, N. Arbour, A. Prat, J. Majewski, M. Lathrop, S. M. Vidal, and P. Gros, "Usp15 regulates type i interferon response and is required for pathogenesis of neuroinflammation," *Nature Immunology*, vol. 18, no. 54, 2016.

[34] "Erp106769." https://www.ncbi.nlm.nih.gov/sra/?term=ERP106769. Accessed: 21-01-2019.

[35] A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras, "Star: ultrafast universal rna-seq aligner," *Bioinformatics*, vol. 29, no. 1, pp. 15–21, 2013.

[36] A. Dobin and T. R. Gingeras, "Mapping rna-seq reads with star," *Current Protocols in Bioinformatics*, vol. 51, no. 1, pp. 11.14.1–11.14.19.

[37] M. Lawrence, W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. Morgan, and V. Carey, "Software for computing and annotating genomic ranges," *PLoS Computational Biology*, vol. 9, 2013.

[38] "Erp023982." https://www.ncbi.nlm.nih.gov/sra/?term=ERP023982. Accessed: 21-01-2019.

[39] D. S. Guttery, B. Poulin, A. Ramaprasad, R. J. Wall, D. J. Ferguson, D. Brady, E.-M. Patzewitz, S. Whipple, U. Straschil, M. H. Wright, A. M. Mohamed, A. Radhakrishnan, S. T. Arold, E. W. Tate, A. A. Holder, B. Wickstead, A. Pain, and R. Tewari, "Genome-wide functional analysis of plasmodium protein phosphatases reveals key regulators of parasite development and differentiation," *Cell Host Microbe*, vol. 16, no. 1, pp. 128 – 140, 2014.

[40] E. R. Derbyshire, M. Prudêncio, M. M. Mota, and J. Clardy, "Liver-stage malaria parasites vulnerable to diverse chemical scaffolds," *Proceedings of the National Academy of Sciences*, vol. 109, no. 22, pp. 8511–8516, 2012.

[41] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, "edger: a bioconductor package for differential expression analysis of digital gene expression data," *Bioinformatics*, vol. 26, no. 1, pp. 139–140, 2010.

[42] McCarthy, D. J., Chen, Yunshun, Smyth, and G. K., "Differential expression analysis of multifactor rna-seq experiments with respect to biological variation," *Nucleic Acids Research*, vol. 40, no. 10, pp. 4288–4297, 2012.

[43] Phipson, S. B, and G, "Permutation p-values should never be zero: Calculating exact p-values when permutations are randomly drawn," *Statistical Applications in Genetics and Molecular Biology*, vol. 9, no. 1, 2010.

[44] T. A. Knijnenburg, L. F. A. Wessels, M. J. T. Reinders, and I. Shmulevich, "Fewer permutations, more accurate p-values," *Bioinformatics*, vol. 25, no. 12, pp. i161–i168, 2009.

[45] R. Khanin and E. Wit, *Construction of Malaria Gene Expression Network Using Partial Correlations*, vol. V, pp. 75–88. 02 2007.

[46] Y. Zuo, G. Yu, M. G. Tadesse, and H. W. Ressom, "Reconstructing biological networks using low order partial correlation," in *2013 IEEE International Conference on Bioinformatics and Biomedicine*, pp. 171–175, Dec 2013.

[47] Y. Zuo, G. Yu, M. G. Tadesse, and H. W. Ressom, "Biological network inference using low order partial correlation," *Methods*, vol. 69, no. 3, pp. 266 – 273, 2014. Recent development in bioinformatics for utilizing omics data.

[48] Y. Zuo, G. Yu, M. G. Tadesse, and H. W. Ressom, "Biological network inference using low order partial correlation," *Methods*, vol. 69, no. 3, pp. 266 – 273, 2014. Recent development in bioinformatics for utilizing omics data.

[49] L. L, J. Christian J. Stoeckert, and D. S. Roos *Genome Research*, vol. 13, 2003.

[50] I. Yanai, D. Graur, and R. Ophir, "Incongruent expression profiles between human and mouse orthologous genes suggest widespread neutral evolution of transcription control," *OMICS: A Journal of Integrative Biology*, vol. 8, no. 1, pp. 15–24, 2004. PMID: 15107234.

[51] B.-Y. Liao and J. Zhang, "Evolutionary Conservation of Expression Profiles Between Human and Mouse Orthologous Genes," *Molecular Biology and Evolution*, vol. 23, pp. 530–540, 11 2005.

[52] P. Tsaparas, L. Mariño-Ramírez, O. Bodenreider, E. V. Koonin, and I. K. Jordan, "Global similarity and local divergence in human and mouse gene co-expression networks," *BMC Evolutionary Biology*, vol. 6, p. 70, Sep 2006.

[53] W. Guo, C. P. G. Calixto, N. Tzioutziou, P. Lin, R. Waugh, J. W. S. Brown, and R. Zhang, "Evaluation and improvement of the regulatory inference for large co-expression networks with limited sample size," *BMC Systems Biology*, vol. 11, p. 62, Jun 2017.

[54] "Erp105548." https://www.ncbi.nlm.nih.gov/sra/?term=ERP105548. Accessed: 21-01-2019.

[55] "Srp071199." https://www.ncbi.nlm.nih.gov/sra/?term=SRP071199. Accessed: 21-01-2019.

[56] "Erp107298." https://www.ncbi.nlm.nih.gov/sra/?term=ERP107298. Accessed: 21-01-2019.

[57] "Srp059851." https://www.ncbi.nlm.nih.gov/sra/?term=SRP059851. Accessed: 21-01-2019.

[58] "Srp083918." https://www.ncbi.nlm.nih.gov/sra/?term=SRP083918. Accessed: 21-01-2019.

[59] "Srp143373." https://www.ncbi.nlm.nih.gov/sra/?term=SRP143373. Accessed: 21-01-2019.

[60] "Erp017542." https://www.ncbi.nlm.nih.gov/sra/?term=ERP017542. Accessed: 21-01-2019.

[61] "Srp116117." https://www.ncbi.nlm.nih.gov/sra/?term=SRP116117. Accessed: 21-01-2019.

[62] "Srp116593." https://www.ncbi.nlm.nih.gov/sra/?term=SRP116593. Accessed: 21-01-2019.

[63] "Srp116793." https://www.ncbi.nlm.nih.gov/sra/?term=SRP116793. Accessed: 21-01-2019.

[64] "Srp118503." https://www.ncbi.nlm.nih.gov/sra/?term=SRP118503. Accessed: 21-01-2019.

[65] "Srp118827." https://www.ncbi.nlm.nih.gov/sra/?term=SRP118827. Accessed: 21-01-2019.

[66] "Srp118996." https://www.ncbi.nlm.nih.gov/sra/?term=SRP118996. Accessed: 21-01-2019.

[67] "Srp070748." https://www.ncbi.nlm.nih.gov/sra/?term=SRP070748. Accessed: 21-01-2019.

[68] "Srp109709." https://www.ncbi.nlm.nih.gov/sra/?term=SRP109709. Accessed: 21-01-2019.

[69] "Erp002273." https://www.ncbi.nlm.nih.gov/sra/?term=ERP002273. Accessed: 21-01-2019.

[70] "Srp112213." https://www.ncbi.nlm.nih.gov/sra/?term=SRP112213. Accessed: 21-01-2019.

[71] "Srp164767." https://www.ncbi.nlm.nih.gov/sra/?term=SRP164767. Accessed: 21-01-2019.

[72] J. Sch'afer and K. Strimmer, "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics," *Statistical Applications in Genetics and Molecular Biology*, vol. 4.

[73] J. Schäfer and K. Strimmer, "An empirical Bayes approach to inferring large-scale gene association networks," *Bioinformatics*, vol. 21, pp. 754–764, 10 2004.

[74] D. Husmeier, A. V. Werhli, and M. Grzegorczyk, "Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical gaussian models and bayesian networks," *Bioinformatics*, vol. 22, pp. 2523–2531, 07 2006.

[75] M. Magnani, D. Vega, and M. Dubik, "Analysis and mining of multilayer social networks," 2019.

[76] Z. Hammoud and F. Kramer, "mully: An r package to create, modify and visualize multilayered graphs," *Genes*, vol. 9, no. 11, 2018.

[77] E. Belilovsky, G. Varoquaux, and M. Blaschko, "Testing for differences in gaussian graphical models: Applications to brain connectivity," in *Proceedings of the*

30th International Conference on Neural Information Processing Systems, NIPS'16, (USA), pp. 595–603, Curran Associates Inc., 2016.

[78] B. Wang, R. Singh, and Y. Qi, "A constrained L1 minimization approach for estimating multiple sparse gaussian or nonparanormal graphical models," *CoRR*, vol. abs/1605.03468, 2016.

[79] W. Liu, "Structural similarity and difference testing on multiple sparse gaussian graphical models," *Ann. Statist.*, vol. 45, pp. 2680–2707, 12 2017.

[80] K. Mohan, P. London, M. Fazel, D. M. Witten, and S.-I. Lee, "Node-based learning of multiple gaussian graphical models," *Journal of machine learning research : JMLR*, vol. 15 1, pp. 445–488, 2014.

[81] S. Zhang, H. Zhao, and M. K. Ng, "Functional module analysis for gene coexpression networks with network integration," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 12, pp. 1146–1160, Sep. 2015.

[82] M. M. Bourdakou, E. I. Athanasiadis, and G. M. Spyrou, "Discovering gene reranking efficiency and conserved gene-gene relationships derived from gene coexpression network analysis on breast cancer data," *Scientific Reports*, vol. 6, 2016.

[83] Y. R. Wang and H. Huang, "Review on statistical methods for gene network reconstruction using expression data," *Journal of Theoretical Biology*, vol. 362, pp. 53 – 61, 2014. Network-based biomarkers for complex diseases.

[84] R. Opgen-Rhein and K. Strimmer, "From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data," *BMC Systems Biology*, vol. 1, p. 37, Aug 2007.

[85] C. Frech and N. Chen, "Genome comparison of human and non-human malaria parasites reveals species subset-specific genes potentially linked to human disease," *PLOS Computational Biology*, vol. 7, pp. 1–18, 12 2011.

[86] K. S., "ppcor: An r package for a fast calculation to semi-partial correlation coefficients," *Commun. Stat. Appl.*, vol. 22, 2015.

[87] T. Brugat, A. J. Reid, J.-w. Lin, D. Cunningham, I. Tumwine, G. Kushinga, S. McLaughlin, P. Spence, U. Böhme, M. Sanders, S. Conteh, E. Bushell, T. Metcalf, O. Billker, P. E. Duffy, C. Newbold, M. Berriman, and J. Langhorne, "Antibody-

independent mechanisms regulate the establishment of chronic plasmodium infection," *Nature Microbiology*, vol. 2, 2017.

[88] A. Ramaprasad, A. K. Subudhi, R. Culleton, and A. Pain, "A fast and cost-effective microsampling protocol incorporating reduced animal usage for time-series transcriptomics in rodent malaria parasites," *bioRxiv*, 2018.

[89] J. Lawton, T. Brugat, Y. X. Yan, A. J. Reid, U. Böhme, T. D. Otto, A. Pain, A. Jackson, M. Berriman, D. Cunningham, P. Preiser, and J. Langhorne, "Characterization and gene expression analysis of the cir multi-gene family of plasmodium chabaudi chabaudi (as)," *BMC Genomics*, vol. 13, p. 125, Mar 2012.

[90] C. Joyner, A. Moreno, E. V. S. Meyer, M. Cabrera-Mora, The MaHPIC Consortium, A. Kong, C. Ibegbu, D. Arafat, D. Jones, D. P. Nace, E. O. Voit, J. Pohl, J. Humphrey, J. DeBarry, J. Sullivan, J. B. Gutierrez, K. J. Kevin, L. L. Fonseca, M. Styczynski, M. Nural, S. T. Pruett, S. A. Lapp, S. Pakala, T. J. Lamb, T. Williams, K. Uppal, V. Tran, W. Yin, Y. Yan, J. C. Kissinger, J. W. Barnwell, and M. R. Galinski, "Plasmodium cynomolgi infections in rhesus macaques display clinical and parasitological features pertinent to modelling vivax malaria pathology and relapse infections," *Malaria Journal*, vol. 15, p. 451, Sep 2016.

[91] J. E. Quin, I. Bujila, M. Chérif, G. S. Sanou, Y. Qu, M. Vafa Homann, A. Rolicka, S. B. Sirima, M. A. O'Connell, A. Lennartsson, M. Troye-Blomberg, I. Nebie, and A.-K. Östlund Farrants, "Major transcriptional changes observed in the fulani, an ethnic group less susceptible to malaria," *eLife*, vol. 6, p. e29156, sep 2017.

[92] K. S. Burrack, M. A. Huggins, E. Taras, P. Dougherty, C. M. Henzler, R. Yang, S. Alter, E. K. Jeng, H. C. Wong, M. Felices, F. Cichocki, J. S. Miller, G. T. Hart, A. J. Johnson, S. C. Jameson, and S. E. Hamilton, "Interleukin-15 complex treatment protects mice from cerebral malaria by inducing interleukin-10-producing natural killer cells," *Immunity*, vol. 48, no. 4, pp. 760 – 772.e4, 2018.

[93] S. A. Ogun, R. Tewari, T. D. Otto, S. A. Howell, E. Knuepfer, D. A. Cunningham, Z. Xu, A. Pain, and A. A. Holder, "Targeted disruption of py235ebp-1: Invasion of erythrocytes by plasmodium yoelii using an alternative py235 erythrocyte binding protein," *PLOS Pathogens*, vol. 7, pp. 1–14, 02 2011.

[94] A. Götz, M. S. Tang, M. C. Ty, C. Arama, A. Ongoiba, D. Doumtabe, B. Traore, P. D. Crompton, P. Loke, and A. Rodriguez, "Atypical activation of dendritic cells by plasmodium falciparum," *Proceedings of the National Academy of Sciences*, vol. 114, no. 49, pp. E10568–E10577, 2017.

[95] J. Lin, A. Reid, D. Cunningham, U. Böhme, I. Tumwine, S. Keller-Mclaughlin, M. Sanders, M. Berriman, and J. Langhorne, "Genomic and transcriptomic comparisons of closely related malaria parasites differing in virulence and sequestration pattern [version 1; referees: 2 approved]," *Wellcome Open Research*, vol. 3, no. 142, 2018.

[96] L. M. Yeoh, C. D. Goodman, V. Mollard, G. I. McFadden, and S. A. Ralph, "Comparative transcriptomics of female and male gametocytes in plasmodium berghei and the evolution of sex in alveolates," *BMC Genomics*, vol. 18, p. 734, Sep 2017.

[97] M. C. Rojas-Pena, D. C. Arafat, J. M. Velasquez, S. C. Garimalla, M. Arevalo-Herrera, S. Herrera, and G. Gibson, "Profiling gene expression of the host response to a plasmodium vivax irradiated sporozoite immunization and infectious challenge," *bioRxiv*, 2018.

[98] N. Gural, L. Mancio-Silva, A. B. Miller, A. Galstian, V. L. Butty, S. S. Levine, R. Patrapuvich, S. P. Desai, S. A. Mikolajczak, S. H. Kappe, H. E. Fleming, S. March, J. Sattabongkot, and S. N. Bhatia, "In vitro culture, drug sensitivity, and transcriptome of plasmodium vivax hypnozoites," *Cell Host Microbe*, vol. 23, no. 3, pp. 395 – 406.e4, 2018.

[99] I. C. Hirako, P. A. Assis, N. S. Hojo-Souza, G. Reed, H. Nakaya, D. T. Golenbock, R. S. Coimbra, and R. T. Gazzinelli, "Daily rhythms of tnf expression and food intake regulate synchrony of plasmodium stages with the host circadian cycle," *Cell Host Microbe*, vol. 23, no. 6, pp. 796 – 808.e6, 2018.

[100] A. V. van der Wel, G. Roma, D. K. Gupta, S. Schuierer, F. Nigsch, W. Carbone, A.-M. Zeeman, B. H. Lee, S. O. Hofman, B. W. Faber, J. Knehr, E. Pasini, B. Kinzel, P. Bifani, G. M. C. Bonamy, T. Bouwmeester, C. H. M. Kocken, and T. T. Diagana, "A comparative transcriptomic analysis of replicating and dormant liver stages of the relapsing malaria parasite plasmodium cynomolgi," *eLIFE*, vol. 6, 2017.

[101] A. Kim, J. Popovici, A. Vantaux, R. Samreth, S. Bin, S. Kim, C. Roesch, L. Liang, H. Davies, P. Felgner, S. Herrera, M. Arévalo-Herrera, D. Ménard, and D. Serre, "Characterization of p. vivax blood stage transcriptomes from field isolates reveals similarities among infections and complex gene isoforms," *Scientific Reports*, vol. 7, 2017.

[102] S. C. Galen, J. Borner, E. S. Martinsen, J. Schaer, C. C. Austin, C. J. West, and S. L. Perkins, "The polyphyly of *Plasmodium*: comprehensive phylogenetic analyses of the malaria parasites (order haemosporida) reveal widespread taxonomic conflict," *Royal Society Open Science*, vol. 5, no. 5, p. 171780, 2018.

[103] P. Legendre, "Comparison of permutation methods for the partial correlation and partial mantel tests," *Journal of Statistical Computation and Simulation*, vol. 67, no. 1, pp. 37–73, 2000.

[104] D. R. Rhodes, S. A. Tomlins, S. Varambally, V. Mahavisno, T. Barrette, S. Kalyana-Sundaram, D. Ghosh, A. Pandey, and A. M. Chinnaiyan, "Probabilistic model of the human protein-protein interaction network," *Nature Biotechnology*, vol. 23, 2005.

[105] S. Persson, H. Wei, J. Milne, G. P. Page, and C. R. Somerville, "Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets," *Proceedings of the National Academy of Sciences*, vol. 102, no. 24, pp. 8633–8638, 2005.

[106] S. Baskerville and D. P. Bartel, "Microarray profiling of micrornas reveals frequent coexpression with neighboring mirnas and host genes," *RNA*, vol. 11, 2005.

[107] S. Torre, M. J. Polyak, D. Langlais, N. Fodil, J. M. Kennedy, I. Radovanovic, J. Berghout, G. A. Leiva-Torres, C. M. Krawczyk, S. Ilangumaran, K. Mossman, C. Liang, K.-P. Knobeloch, L. M. Healy, J. Antel, N. Arbour, and Prat, "Usp15 regulates type i interferon response and is required for pathogenesis of neuroinflammation," *Nature Immunology*, vol. 18, 2016.

[108] S. Das, N. Hertrich, A. J. Perrin, C. Withers-Martinez, C. R. Collins, M. L. Jones, J. M. Watermeyer, E. T. Fobes, S. R. Martin, H. R. Saibil, G. J. Wright, M. Treeck, C. Epp, and M. J. Blackman, "Processing of plasmodium falciparum merozoite surface protein msp1 activates a spectrin-binding function enabling parasite egress from rbcs," *Cell Host Microbe*, vol. 18, no. 4, pp. 433 – 444, 2015.

[109] H. J. Lee, A. Georgiadou, M. Walther, D. Nwakanma, L. B. Stewart, M. Levin, T. D. Otto, D. J. Conway, L. J. Coin, and A. J. Cunnington, "Integrated pathogen load and dual transcriptome analysis of systemic host-pathogen interactions in severe malaria," *Science Translational Medicine*, vol. 10, no. 447, 2018.

[110] P. J. Spence, W. Jarra, P. Lévy, A. J. Reid, L. Chappell, T. Brugat, M. Sanders, M. Berriman, and J. Langhorne, "Vector transmission regulates immune control of plasmodium virulence," *Nature*, vol. 498.

[111] M. L. Rojas-Peña, A. Vallejo, S. Herrera, G. Gibson, and M. Arévalo-Herrera, "Transcription profiling of malaria-naïve and semi-immune colombian volunteers in a plasmodium vivax sporozoite challenge," *PLOS Neglected Tropical Diseases*, vol. 9, pp. 1–18, 08 2015.

[112] R. Cubi, S. S. Vembar, A. Biton, J.-F. Franetich, M. Bordessoulles, D. Sossau, G. Zanghi, H. Bosson-Vanga, M. Benard, A. Moreno, N. Dereuddre-Bosquet, R. Le Grand, A. Scherf, and D. Mazier, "Laser capture microdissection enables transcriptomic analysis of dividing and quiescent liver stages of plasmodium relapsing species," *Cellular Microbiology*, vol. 19, no. 8, p. e12735, 2017. e12735 CMI-17-0043.

[113] K. Modrzynska, C. Pfander, L. Chappell, L. Yu, C. Suarez, K. Dundas, A. R. Gomes, D. Goulding, J. C. Rayner, J. Choudhary, and O. Billker, "A knockout screen of apiap2 genes reveals networks of interacting transcriptional regulators controlling the plasmodium life cycle," *Cell Host Microbe*, vol. 21, no. 1, pp. 11 – 22, 2017.

[114] D. K. Jaijyan, H. Singh, and A. P. Singh, "A sporozoite- and liver stage-expressed tryptophan-rich protein plays an auxiliary role in plasmodium liver stage development and is a potential vaccine candidate," *Journal of Biological Chemistry*, vol. 290.

[115] T. M. Tran, M. B. Jones, A. Ongoiba, E. M. Bijker, R. Schats, P. Venepally, J. Skinner, S. Doumbo, E. Quinten, L. G. Visser, E. Whalen, S. Presnell, E. M. O'Connell, K. Kayentao, D. Chaussabel, H. Lorenzi, T. B. Nutman, T. H. M. Ottenhoff, M. C. Haks, B. Traore, E. F. Kirkness, R. W. Sauerwein, and P. D. Crompton, "Transcriptomic evidence for modulation of host inflammatory responses during febrile plasmodium falciparum malaria," *Scientific Reports*, vol. 6.

[116] J. Rothen, C. Murie, J. Carnes, A. Anupama, S. Abdulla, M. Chemba, M. Mpina, M. Tanner, B. K. Lee Sim, S. L. Hoffman, R. Gottardo, C. Daubenberger, and

K. Stuart, "Whole blood transcriptome changes following controlled human malaria infection in malaria pre-exposed volunteers correlate with parasite prepatent period," *PLOS ONE*, vol. 13, pp. 1–18, 06 2018.

[117] J. L. Miller, B. K. Sack, M. Baldwin, A. M. Vaughan, and S. H. Kappe, "Interferon-mediated innate immune responses against malaria parasite liver stages," *Cell Reports*, vol. 7, no. 2, pp. 436 – 447, 2014.

[118] J. Li, B. Cai, Y. Qi, W. Zhao, J. Liu, R. Xu, Q. Pang, Z. Tao, L. Hong, S. Liu, M. Leerkes, M. Quiñones, and X.-z. Su, "Utr introns, antisense rna and differentially spliced transcripts between plasmodium yoelii subspecies," *Malaria Journal*, vol. 15, p. 30, Jan 2016.

[119] S. Aibar, C. Fontanillo, C. Droste, and J. De Las Rivas, "Functional gene networks: R/bioc package to generate and analyse gene networks derived from functional enrichment and clustering," *Bioinformatics*, vol. 31, no. 10, pp. 1686–1688, 2015.

[120] R. Zhang, Z. Ren, and W. Chen, "Silggm: An extensive r package for efficient statistical inference in large-scale gene networks," *PLOS Computational Biology*, vol. 14, pp. 1–14, 08 2018.

[121] X.-F. Zhang, L. Ou-Yang, S. Yang, X. Hu, and H. Yan, "Diffgraph: an r package for identifying gene network rewiring using differential graphical models," *Bioinformatics*, vol. 34, no. 9, pp. 1571–1573, 2018.

[122] P. S. T. Russo, G. R. Ferreira, L. E. Cardozo, M. C. Bürger, R. Arias-Carrasco, S. R. Maruyama, T. D. C. Hirata, D. S. Lima, F. M. Passos, K. F. Fukutani, M. Lever, J. S. Silva, V. Maracaja-Coutinho, and H. I. Nakaya, "Cemitool: a bioconductor package for performing comprehensive modular co-expression analyses," *BMC Bioinformatics*, vol. 19, p. 56, Feb 2018.

[123] T. L. Pedersen, "A tidy api for graph manipulation,"

[124] G. Csardi and T. Nepusz, "The igraph software package for complex network research," *InterJournal*, vol. Complex Systems, p. 1695, 2006.

[125] D. Li, S. He, Z. Pan, and G. Hu, "Active modules for multilayer weighted gene co-expression networks: a continuous optimization approach," *bioRxiv*, 2016.

[126] S. Kumari, J. Nie, H.-S. Chen, H. Ma, R. Stewart, X. Li, M.-Z. Lu, W. M. Taylor, and H. Wei, "Evaluation of gene association methods for coexpression network

construction and biological knowledge discovery," *PLOS ONE*, vol. 7, pp. 1–17, 11 2012.

[127] WHO, "World malaria report," tech. rep., 2018.

[128] Lee, H. Homin K, S. Amy K, Q. Jon, P. Jie, and Paul, "Coexpression analysis of human genes across many microarray data sets," *Genome research*, vol. 14, 2004.

[129] S. Gandon, M. J. Mackinnon, S. Nee, and A. F. Read, "Imperfect vaccines and the evolution of pathogen virulence," *Nature*, vol. 414, 2001.

[130] F. E. Cox, "History of the discovery of the malaria parasites and their vectors," *Parasites Vectors*, vol. 3, p. 5, Feb 2010.

[131] K. Haldar, S. Bhattacharjee, and I. Safeukui, "Drug resistance in plasmodium," *ature Reviews Microbiology*, vol. 16, 2018.

[132] T. Shay, V. Jojic, O. Zuk, K. Rothamel, D. Puyraimond-Zemmour, T. Feng, E. Wakamatsu, C. Benoist, D. Koller, and A. Regev, "Conservation and divergence in the transcriptional programs of the human and mouse immune systems," *Proceedings of the National Academy of Sciences*, vol. 110, no. 8, pp. 2946–2951, 2013.

[133] Carlton, S. Jane, H. Joana, and Neil, "The genome of model malaria parasites and comparative genomics," *Current Issues in Molecular Biology*, vol. 7, pp. 23–38, 2005.

[134] A. F. Cowman, D. Berry, and J. Baum, "The cellular and molecular basis for malaria parasite invasion of the human red blood cell," *The Journal of Cell Biology*, vol. 198, no. 6, pp. 961–971, 2012.

[135] M. Lawrence, W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, and V. J. Carey, "Software for computing and annotating genomic ranges," *PLOS Computational Biology*, vol. 9, pp. 1–10, 08 2013.

[136] E. L. Flannery, A. K. Chatterjee, and E. A. Winzeler, "Antimalarial drug discovery — approaches and progress towards new medicines," *Nature Reviews Microbiology*, vol. 11, 2013.

[137] A. Conesa, P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, M. W. Szcześniak, D. J. Gaffney, L. L. Elo, X. Zhang, and A. Mortazavi, "A survey of best practices for rna-seq data analysis," *Genome Biology*, vol. 17, p. 13, Jan 2016.

[138] N. Hall, M. Karras, J. D. Raine, J. M. Carlton, T. W. A. Kooij, M. Berriman, L. Florens, C. S. Janssen, A. Pain, G. K. Christophides, K. James, K. Rutherford, B. Harris, D. Harris, C. Churcher, M. A. Quail, D. Ormond, J. Doggett, H. E. Trueman, J. Mendoza, S. L. Bidwell, M.-A. Rajandream, D. J. Carucci, J. R. Yates, F. C. Kafatos, C. J. Janse, B. Barrell, C. M. R. Turner, A. P. Waters, and R. E. Sinden, "A comprehensive survey of the plasmodium life cycle by genomic, transcriptomic, and proteomic analyses," *Science*, vol. 307, no. 5706, pp. 82–86, 2005.

[139] M. Rug, S. W. Prescott, K. M. Fernandez, B. M. Cooke, and A. F. Cowman, "The role of kahrp domains in knob formation and cytoadherence of p falciparum-infected human erythrocytes," *Blood*, vol. 108, no. 1, pp. 370–378, 2006.

[140] E. Bushell, A. R. Gomes, T. Sanderson, B. Anar, G. Girling, C. Herd, T. Metcalf, K. Modrzynska, F. Schwach, R. E. Martin, M. W. Mather, G. I. McFadden, L. Parts, G. G. Rutledge, A. B. Vaidya, K. Wengelnik, J. C. Rayner, and O. Billker, "Functional profiling of a plasmodium genome reveals an abundance of essential genes," *Cell*, vol. 170, no. 2, pp. 260 – 272.e8, 2017.

[141] A. K. Dasanna, C. Lansche, M. Lanzer, and U. S. Schwarz, "Rolling adhesion of schizont stage malaria-infected red blood cells in shear flow," *Biophysical Journal*, vol. 112, no. 9, pp. 1908 – 1919, 2017.

[142] M. Recker, C. O. Buckee, A. Serazin, S. Kyes, R. Pinches, Z. Christodoulou, A. L. Springer, S. Gupta, and C. I. Newbold, "Antigenic variation in plasmodium falciparum malaria involves a highly structured switching pattern," *PLOS Pathogens*, vol. 7, pp. 1–11, 03 2011.

[143] K. W. Deitsch and R. Dzikowski, "Variant gene expression and antigenic variation by malaria parasites," *Annual Review of Microbiology*, vol. 71, no. 1, pp. 625–641, 2017. PMID: 28697665.

[144] A. S. Aly, A. M. Vaughan, and S. H. Kappe, "Malaria parasite development in the mosquito and infection of the mammalian host," *Annual Review of Microbiology*, vol. 63, no. 1, pp. 195–221, 2009. PMID: 19575563.

[145] E. Nourani, F. Khunjush, and S. Durmuş, "Computational approaches for prediction of pathogen-host protein-protein interactions," *Frontiers in Microbiology*, vol. 6, p. 94, 2015.

[146] P. F. L. Reyes, T. Michoel, A. Joshi, and G. Devailly, "Meta-analysis of liver and heart transcriptomic data for functional annotation transfer in mammalian orthologs," *bioRxiv*, 2017.

[147] S. Tornow and H. W. Mewes, "Functional modules by relating protein interaction networks and gene expression," *Nucleic Acids Research*, vol. 31, no. 21, pp. 6283–6289, 2003.

[148] Y.-C. Wang, C. Lin, M.-T. Chuang, W.-P. Hsieh, C.-Y. Lan, Y.-J. Chuang, and B.-S. Chen, "Interspecies protein-protein interaction network construction for characterization of host-pathogen interactions: a candida albicans-zebrafish interaction study," *BMC Systems Biology*, vol. 7, p. 79, Aug 2013.

[149] H. M. Abkallo, A. Martinelli, M. Inoue, A. Ramaprasad, P. Xangsayarath, J. Gitaka, J. Tang, K. Yahata, A. Zoungrana, H. Mitaka, A. Acharjee, P. P. Datta, P. Hunt, R. Carter, O. Kaneko, V. Mustonen, C. J. R. Illingworth, A. Pain, and R. Culleton, "Rapid identification of genes controlling virulence and immunity in malaria parasites," *PLOS Pathogens*, vol. 13, pp. 1–24, 07 2017.

[150] S. van Dam, U. Võsa, A. van der Graaf, L. Franke, and J. P. de Magalhães, "Gene co-expression analysis for functional classification and gene–disease predictions," *Briefings in Bioinformatics*, p. bbw139, 2017.

[151] S. Wuchty, J. H. Adams, and M. T. Ferdig, "A comprehensive plasmodium falciparum protein interaction map reveals a distinct architecture of a core interactome," *PROTEOMICS*, vol. 9, no. 7, pp. 1841–1849, 2009.

[152] M. Tomasini, "An introduction to multilayer networks," 2015.

[153] Langfelder, H. Peter, and Steve, "Fast r functions for robust correlations and hierarchical clustering," *Journal of Statistical Software*, vol. 46, no. 11, 2012.

[154] Pearson and Karl, "Mathematical contributions to the theory of evolution. on a form of spurious correlation which may arise when indices are used in the measurement of organs," *Proceedings of the Royal Society of London*, vol. 60, pp. 489–498, 1897.

[155] S. Boccaletti, G. Bianconi, R. Criado, C. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin, "The structure and dynamics of multilayer networks," *Physics Reports*, vol. 544, no. 1, pp. 1 – 122, 2014. The structure and dynamics of multilayer networks.

[156] F. Battiston, V. Nicosia, and V. Latora, "Structural measures for multiplex networks," *Phys. Rev. E*, vol. 89, p. 032804, Mar 2014.

[157] A. Aleta and Y. Moreno, "Multilayer networks in a nutshell," *Annual Review of Condensed Matter Physics*, vol. 10, no. 1, p. null, 2019.

[158] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for rna-seq data with deseq2," *Genome Biology*, vol. 15, p. 550, 2014.

[159] M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M. A. Porter, S. Gómez, and A. Arenas, "Mathematical formulation of multilayer networks," *Phys. Rev. X*, vol. 3, p. 041022, Dec 2013.

[160] W. Liu, T. Suzumura, H. Ji, and G. Hu, "Finding overlapping communities in multilayer networks," *PLOS ONE*, vol. 13, pp. 1–22, 04 2018.

[161] Y. Zhao, E. Levina, and J. Zhu, "Community extraction for social networks," *Proceedings of the National Academy of Sciences*, vol. 108, no. 18, pp. 7321–7326, 2011.

[162] B. Zhang and S. Horvath, "A general framework for weighted gene co-expression network analysis," *Statistical Applications in Genetics and Molecular Biology*, vol. 4, no. 1, 2005.

[163] Barabási, O. Albert-László, and Z. N., "Network biology: understanding the cell's functional organization," *Nature Reviews Genetics*, vol. 5, 2004.

[164] A. Reverter and E. K. F. Chan, "Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks," *Bioinformatics*, vol. 24, no. 21, pp. 2491–2497, 2008.

[165] O. D. Iancu, S. Kawane, D. Bottomly, R. Searles, R. Hitzemann, and S. McWeeney, "Utilizing rna-seq data for de novo coexpression network inference," *Bioinformatics*, vol. 28, no. 12, pp. 1592–1597, 2012.

[166] Segal, S. Eran, R. Michael, P. Aviv, B. Dana, K. David, F. Daphne, and Nir, "Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data," *Nature Genetics*, vol. 34, 2003.

[167] S. Wuchty, "Computational prediction of host-parasite protein interactions between p. falciparum and h. sapiens," *PLOS ONE*, vol. 6, pp. 1–8, 11 2011.

[168] J. Friedman and E. J. Alm, "Inferring correlation networks from genomic survey data," *PLOS Computational Biology*, vol. 8, pp. 1–11, 09 2012.

[169] Z. D. Kurtz, C. L. Müller, E. R. Miraldi, D. R. Littman, M. J. Blaser, and R. A. Bonneau, "Sparse and compositionally robust inference of microbial ecological networks," *PLOS Computational Biology*, vol. 11, pp. 1–25, 05 2015.

[170] Yamagishi, N. Junya, T. Anna, M. Mohammed E.M., S. Arthur E., K. Chihiro, K. Toshiaki, M. Shuichi, M. Wojciech, E. Ryuichiro, T. Yuki, Josef, Suzuki, and Yutaka, "Interactive transcriptome analysis of malaria patients and infecting plasmodium falciparum," *Genome Research*, vol. 24, pp. 1433–1444, 2014.

[171] B. M. Musungu, D. Bhatnagar, R. L. Brown, G. A. Payne, G. OBrian, A. M. Fakhoury, and M. Geisler, "A network approach of gene co-expression in the zea mays/aspergillus flavus pathosystem to map host/pathogen interaction pathways," *Frontiers in Genetics*, vol. 7, p. 206, 2016.

[172] C. W. Remmele, C. H. Luther, J. Balkenhol, T. Dandekar, T. Müller, and M. T. Dittrich, "Integrated inference and evaluation of host–fungi interaction networks," *Frontiers in Microbiology*, vol. 6, p. 764, 2015.

[173] S. Durmuş, T. Çakır, A. Özgür, and R. Guthke, "A review on computational systems biology of pathogen–host interactions," *Frontiers in Microbiology*, vol. 6, p. 235, 2015.

[174] L. Song, P. Langfelder, and S. Horvath, "Comparison of co-expression measures: mutual information, correlation, and model based indices," *BMC Bioinformatics*, vol. 13, p. 328, Dec 2012.

[175] A. K. Subudhi, P. A. Boopathi, I. Pandey, R. Kaur, S. Middha, J. Acharya, S. K. Kochar, D. K. Kochar, and A. Das, "Disease specific modules and hub genes for intervention strategies: A co-expression network based approach for plasmodium falciparum clinical isolates," *Infection, Genetics and Evolution*, vol. 35, pp. 96 – 108, 2015.

[176] M. L. Smith and M. P. Styczynski, "Systems biology-based investigation of host–plasmodium interactions," *Trends in Parasitology*, vol. 34, no. 7, pp. 617 – 632, 2018.

[177] A. Grönlund, R. P. Bhalerao, and J. Karlsson, "Modular gene expression in poplar: A multilayer network approach," *The New Phytologist*, vol. 181, no. 2, pp. 315–322, 2009.

[178] L. Cantini, E. Medico, S. Fortunato, and M. Caselle, "Detection of gene communities in multi-networks reveals cancer drivers," *Scientific Reports*, vol. 5, 2015.

[179] D. Li, Z. Zhu, Z. Pan, G. Hu, and S. He, "Extracting active modules from multilayer ppi network: a continuous optimization approach," *bioRxiv*, 2017.

[180] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, 2005.

[181] X. Li, M. Wu, C.-K. Kwoh, and S.-K. Ng, "Computational approaches for detecting protein complexes from protein interaction networks: a survey," *BMC Genomics*, vol. 11, p. S3, Feb 2010.

[182] M. Zitnik and J. Leskovec, "Predicting multicellular function through multi-layer tissue networks," *Bioinformatics*, vol. 33, no. 14, pp. i190–i198, 2017.

[183] D. Beisser, G. W. Klau, T. Dandekar, T. Müller, and M. T. Dittrich, "Bionet: an r-package for the functional analysis of biological networks," *Bioinformatics*, vol. 26, no. 8, pp. 1129–1130, 2010.

[184] P. Parsana, C. Ruberman, A. E. Jaffe, M. C. Schatz, A. Battle, and J. T. Leek, "Addressing confounding artifacts in reconstruction of gene co-expression networks," *bioRxiv*, 2018.

[185] S. Freytag, J. Gagnon-Bartsch, T. P. Speed, and M. Bahlo, "Systematic noise degrades gene co-expression signals but can be corrected," *BMC Bioinformatics*, vol. 16, p. 309, Sep 2015.

[186] J. Friedman and E. J. Alm, "Inferring correlation networks from genomic survey data," *PLOS Computational Biology*, vol. 8, pp. 1–11, 09 2012.

[187] Z. D. Kurtz, C. L. Müller, E. R. Miraldi, D. R. Littman, M. J. Blaser, and R. A. Bonneau, "Sparse and compositionally robust inference of microbial ecological networks," *PLOS Computational Biology*, vol. 11, pp. 1–25, 05 2015.

[188] Y. Deng, Y.-H. Jiang, Y. Yang, Z. He, F. Luo, and J. Zhou, "Molecular ecological network analyses," *BMC Bioinformatics*, vol. 13, p. 113, May 2012.

[189] N. A. Fonseca, J. Rung, A. Brazma, and J. C. Marioni, "Tools for mapping high-throughput sequencing data," *Bioinformatics*, vol. 28, no. 24, pp. 3169–3177, 2012.

[190] H. Li and R. Durbin, "Fast and accurate short read alignment with burrows–wheeler transform," *Bioinformatics*, vol. 25, no. 14, pp. 1754–1760, 2009.

[191] T. W. Lam, W. K. Sung, S. L. Tam, C. K. Wong, and S. M. Yiu, "Compressed indexing and local alignment of dna," *Bioinformatics*, vol. 24, no. 6, pp. 791–797, 2008.

[192] H. Yang and G. Churchill, "Estimating p-values in small microarray experiments," *Bioinformatics*, vol. 23, no. 1, pp. 38–43, 2007.

[193] J. Li, D. Witten, I. Johnstone, and R. Tibshirani, "Normalization, testing, and false discovery rate estimation for rna-sequencing data," *Biostatistics*, vol. 13, pp. 523–538, 7 2012.

[194] C. J. Wolfe, I. S. Kohane, and A. J. Butte, "Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks," *BMC Bioinformatics*, vol. 6, p. 227, Sep 2005.

[195] H. J. Lee, A. Georgiadou, T. D. Otto, M. Levin, L. J. Coin, D. J. Conway, and A. J. Cunnington, "Transcriptomic studies of malaria: a paradigm for investigation of systemic host-pathogen interactions," *Microbiology and Molecular Biology Reviews*, vol. 82, no. 2, 2018.

[196] O. Silvie, M. M. Mota, K. Matuschewski, and M. Prudêncio, "Interactions of the malaria parasite and its mammalian host," *Current Opinion in Microbiology*, vol. 11, no. 4, pp. 352 – 359, 2008. Host–microbe interactions: fungi/parasites/viruses.

[197] G. V. Glass, "Primary, secondary, and meta-analysis of research," *Educational Researcher*, vol. 5, no. 10, pp. 3–8, 1976.