Integrated pathogen load and dual transcriptome analysis of systemic host-pathogen interactions in severe malaria

Hyun Jae Lee[1], Michael Walther[2], Athina Georgiadou[3], Davis Nwakanma[2], Lindsay B. Stewart[4], Michael Levin[3], Thomas D. Otto[5#], David J. Conway[4#], Lachlan J. Coin[1#], Aubrey J. Cunnington[3]*

[1]Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia

[2]Medical Research Council Gambia Unit, Fajara, The Gambia

[3] Section of Paediatrics, Imperial College, London, UK

[4] Department of Pathogen Molecular Biology, London School of Hygiene and Tropical Medicine, UK

[5] Wellcome Trust Sanger Centre, Hinxton, United Kingdom

*Corresponding author: a.cunnington@imperial.ac.uk (AJC), [#]Denotes equal contribution

Abstract

The pathogenesis of severe *Plasmodium falciparum* malaria is incompletely understood. Since the pathogenic stage of the parasite is restricted to blood, dual RNA-sequencing of host and parasite transcripts in blood can reveal their interactions at a systemic scale. Here we identify human and parasite gene expression associated with severe disease features in Gambian children. Differences in parasite load explained up to 99% of differential expression of human genes but only a third of the differential expression of parasite genes. Co-expression analyses showed a remarkable co-regulation of host and parasite genes controlling translation, and host granulopoiesis genes uniquely co-regulated and differentially expressed in severe malaria. Our results indicate that high parasite load is the proximal stimulus for severe *P. falciparum* malaria, that there is an unappreciated role for many parasite genes in determining virulence, and hint at a molecular arms-race between host and parasite to synthesise protein products.

Introduction

Plasmodium falciparum malaria is one of the most important infectious diseases affecting humankind[1]. Progress has been made in malaria treatment and control in the last decade but this is threatened by the spread of antimalarial and insecticide resistance[2-4]. Understanding of pathogenic mechanisms associated with severe malaria (SM), which puts individuals at risk of death, has also progressed[1, 5, 6]. Immunopathology, vascular endothelial dysfunction and parasite sequestration (obstruction of the microvasculature by cytoadherent parasites) all have putative roles in SM[5], and high parasite load is also strongly associated with greater risk of severe disease[5, 7-10]. Rodent models have contributed to mechanistic dissection of the pathogenic processes, but these cannot yet reproduce all of the features of naturally-occurring *P. falciparum* malaria[6, 11]. An integrated understanding of the respective roles and interactions of host and parasite in human SM is notably lacking, and whether SM involves excessive, proportionate or insufficient host responses to the parasite is largely unknown. Here we combine estimates of parasite load with host and parasite whole blood gene expression to investigate their associations with severity and different pathological features of SM, aiming to provide a global view of systemic host-parasite interaction. This approach allows us to harness the natural variation which occurs between humans infected with *P. falciparum* to better understand the different pathogenic processes which underlie SM.

Results

We performed dual-RNA sequencing on whole blood of 46 Gambian children with uncomplicated (UM, n=21) and severe (SM, n=25) *P. falciparum* malaria (Supplementary Table 1). After exclusion of *var*, *rifin*, *stevor[12]* and highly polymorphic regions in the parasite genome (see Methods), we obtained median 26.6 million (26.6 million SM, 26.7 million UM, P=0.913) human and 9.61 million (10.3 million SM, 5.03 million UM, P=0.346) parasite uniquely mapped reads from each subject (Fig 1a), with considerably greater parasite read depth than a previous study conducted in adults with UM[13]. Systemic infection provokes changes in blood leukocyte subpopulations which could

dominate changes in gene expression[14] so we performed "gene signature" based deconvolution[15] to identify and adjust for heterogeneity in the major leukocyte subpopulations in each sample (Fig 1b, Supplementary Fig 1). Parasite gene expression *in vivo* is also influenced by the mixture of parasite developmental stages at the time of sampling because there is phasic variation in gene expression[16] and increasing RNA content during the intraerythrocytic developmental cycle[17]. Therefore we used the same deconvolution approach with "gene signatures" derived from highly synchronous parasite cultures[16, 18] to identify the contribution of parasites at different developmental stages (Supplementary Fig 2 and Fig 1c). There was a trend towards greater proportions of late stage asexual parasites and gametocytes in children with SM (Fig 1d).

Examination of principal component plots before and after adjustment for heterogeneity in the mixture of leukocytes and parasite developmental stages revealed that segregation of SM and UM cases was improved after adjustment (Fig 1e,f). Therefore we used these adjusted gene expression values for all subsequent analyses, essentially allowing us to compare gene expression as if all subjects had the same leukocyte and parasite population compositions.

Whole blood genome-wide gene expression can be used to characterise host cellular responses and infer upstream regulators[19], whilst variations in *P. falciparum* gene expression are believed to reflect adaptation to the host environment and contribute to virulence[20, 21]. We identified significantly differentially expressed genes from human and parasite in SM vs UM (Fig 2a,b) and also for different subtypes of SM (hyperlactatemia (HL) and cerebral malaria (CM), alone or in combination) vs UM (Supplementary Figure 3). There were 770 human and 236 parasite significantly differentially expressed genes (DEGs, with false discovery rate (FDR)-adjusted $P<0.05$) between SM and UM (Supplementary Table 2 and 3). Some human genes had both conspicuously high expression in SM relative to UM (high log-fold change) and were also highly significant (Fig 2a): the four most upregulated (*MMP8*, matrix metallopeptidase 8; *OLFM4*, olfactomedin 4; *DEFA3*, defensin A3;

ELANE, neutrophil elastase) notably all encode neutrophil granule proteins[22]. Interestingly the number of human and parasite DEGs was substantially higher when comparing the subgroup of subjects with cerebral malaria plus hyperlactatemia (CH, the most severe phenotype, n=12) vs UM, despite the smaller number of subjects (Supplementary Table 2, Supplementary Figure 3).

Previous studies have shown a correlation between host gene expression and circulating parasitemia[23, 24], suggesting that this might explain some of the differences in gene expression between SM and UM. However peripheral blood parasite measurements underestimate the total number of parasites in the body because parasitized red blood cells can also become sequestered, accumulating in small blood vessels rather than remaining in circulation[12, 25]. The parasite protein, *P. falciparum* histidine rich protein 2 (PfHRP2), can be used as a plasma biomarker of total parasite load and is more strongly associated with severity[7, 8, 10] (Supplementary Table 1) and death[8, 10]. Therefore we examined the association of host and parasite gene expression with both circulating parasite density and PfHRP2 (restricting comparisons to subjects with data for both). We found 1886 human genes significantly (FDR $P<0.05$) correlated with log parasite density and 616 significantly correlated with log PfHRP2 (102 common to both), whilst only 2 and 10 parasite genes were significant in the corresponding analyses (none common to both) (Supplementary Tables 4 and 5). We then asked to what extent the differences between SM and UM phenotypes were dependent on parasite load. The number of human SM vs UM DEGs remained almost unchanged before and after adjustment for parasite density but was reduced by 98.6% after adjustment for PfHRP2, whilst parasite DEGs changed much less after the same adjustments (Figure 2c, Supplementary Tables 2 and 3). Findings were similar when adjusting for parasite load in comparisons of each of the SM subtypes vs UM (Supplementary Tables 2 and 3).

Genes associated with severity after adjustment for parasite load may include determinants of susceptibility to severe disease. Of particular interest amongst these, MMP8 (also known as

collagenase 1) is a metallopeptidase which causes endothelial barrier damage in several infection models[26, 27]; *AZI2* (also known as NF-Kappa-B-Activating Kinase-Associated Protein 1, NAP1) encodes a regulator of the type 1 interferon response[28], a pathway which is known to control severity of disease in rodent malaria models[29]; whilst CX3CR1 is the receptor for fractalkine (a biomarker of CM in humans[30]) and a marker for a subset of monocytes which are particularly efficient at killing malaria parasites[31].

Next we performed pathway analyses to better understand the biological functions of the significant genes in the preceding analyses. Human genes correlated with log parasite density were particularly enriched in pathways related to translation (especially exported proteins), oxidative phosphorylation and ubiquitination (Fig 2d, Supplementary Table 6), with predicted upstream regulation by RICTOR (RPTOR independent companion of MTOR complex 2), HNF4A (hepatocyte nuclear factor 4 alpha) and XBP1 (X-box binding protein 1; Supplementary Table 7). Genes correlated with log PfHRP2 were particularly enriched in inflammatory and immune response functions (specifically innate response and type 1 interferon, Fig 2d, Supplementary Table 6), with predicted regulation by IFN-γ, TGM2 (transglutaminase 2) and IFN-α2 (Supplementary Table 7). Some of these immune response functions were also correlated with parasite density but were not significantly enriched because of the larger denominator in this analysis (Fig 2d, Supplementary Tables 6). We compared human DEGs from SM vs UM comparisons without adjustment and with adjustment for parasite density or PfHRP2 (Figure 2e, Supplementary Table 6)). In unadjusted analyses human genes were particularly enriched in processes controlling protein synthesis and targeting to the endoplasmic reticulum, cell stress and immune response and the most significant predicted upstream regulators were CSF3 (colony stimulating factor 3, also known as granulocyte colony stimulating factor, GCSF), FAS (Fas cell surface death receptor) and PTGER2 (Prostaglandin E receptor 2, Supplementary Table 7). After adjustment for parasite density we observed little change in pathway enrichment associated with severe malaria, whilst adjustment for PfHRP2 reduced all pathway enrichments. In contrast parasite

pathways were less influenced by adjustment for either parasite density or PfHRP2 (Fig 2f, Supplementary Table 8), the most consistently significant being RNA processing, protein transport, and hemoglobin catabolism. Findings were broadly similar in corresponding analyses for all subtypes of SM (Supplementary Figure 4, Supplementary Table 6), with the exception that adjustment for parasite density produced a 5-fold increase in the number of DEGs in the hyperlactatemia (HL) vs UM comparison. This intriguing finding suggests that the response to circulating parasites in these groups with similar parasite density (Supplementary Table 1) may partially mask differences in gene expression associated with hyperlactatemia.

To further explore pathophysiology we examined the correlation of human and parasite gene expression with lactate and hemoglobin concentrations and platelet count[32]. 1012 human genes were significantly correlated with lactate concentration, reducing by half after adjustment for parasite density and by 95% after adjustment for PfHRP2 (Fig 2c,g, Supplementary Table 4). Immune response pathways were prominent in unadjusted analysis (the negative association with type 1 interferon being particularly notable), and the most significant predicted upstream regulators were interferon-γ, interferon-α, and TNF. Adjustment for parasite density retained most enrichment terms whilst adjustment for PfHRP2 removed almost all significant enrichment (Fig 2g, Supplementary Table 6) but remaining genes included *PKM* (encoding the glycolytic enzyme pyruvate kinase M) and *GYS1* (encoding the glycogenic enzyme glycogen synthase 1) (Supplementary Table 4). These findings suggest hyperlactatemia is driven by the parasite load-dependent inflammatory response, but also influenced by some parasite load-independent variation in control of host metabolism. 100 parasite genes were significantly correlated with lactate, with much less dependency on parasite load (Fig 2c,h, Supplementary Table 5). Unexpectedly these included two glycolysis genes (hexokinase and acetylCoA synthetase), negatively correlated with lactate (Fig 2h, Supplementary Table 9), suggesting that rather than parasite-derived lactate driving hyperlactatemia, host-derived lactate may negatively regulate parasite glycolysis. Compared to lactate, platelet count was associated with

157 fewer human genes, and these were less dependent on parasite load (Fig 2c,i, Supplementary Table
158 4). The most enriched pathways also differed considerably, with nucleosome assembly
159 (predominantly histone genes), coagulation, and response to wounding genes, all negatively
160 correlated with platelet count (Fig 2i), and the most significant predicted upstream regulators being
161 IL13, RB1 (RB transcriptional corepressor 1), and IL1RN (Supplementary Table 7). Activation of
162 coagulation pathways is increasingly recognised in severe malaria[33, 34], but free histones can also
163 induce thrombocytopenia[35] and may be relevant in malaria. No human genes and few parasite
164 genes were associated hemoglobin concentration in any analyses and only one parasite gene was
165 associated with platelet count (Figure 2c, Supplementary Table 4,5).
166
167 Taken together the preceding findings indicate that total parasite load is the dominant driver of host
168 leukocyte gene expression in malaria, particularly inflammatory and immune response genes, and
169 differences in parasite load explain almost all of the human gene expression differences between SM
170 and UM. Despite this, specific parasite load-dependent pathways were differentially associated with
171 distinct aspects of systemic pathophysiology, and circulating parasites correlated with patterns of
172 host gene expression suggesting that parasite localization substantially alters the host-parasite
173 interaction. In contrast to host genes, parasite gene expression showed little association with
174 parasite load, implying that the non-polymorphic genes differentially expressed between SM and
175 UM do not directly contribute to high parasite load, but may contribute to other aspects of
176 pathogenesis or simply reflect parasite responses to the perturbed host environment.
177
178 Whilst these independent analyses of host and parasite gene expression associations with severity
179 are enlightening, dual-RNA sequencing can also be used to identify molecular interactions within and
180 between species and their associations with severity[36]. Expression of groups of genes with
181 common functional roles are often highly correlated and can be identified through co-expression
182 network analysis[37]. We applied this methodology to identify correlated modules of genes, which

could originate from either or both species and were named according to the "hub gene" which has the greatest connectivity within the module. First we analysed all subjects together and generated a network with 26 modules (Fig 3 and Supplementary Table 9): 10 containing exclusively human genes, 5 exclusively parasite genes, and 11 with both human and parasite genes (although most of these were highly skewed to a single species). All modules showed significant functional enrichments regardless of host or parasite origin. The composite expression of genes within a module can be described by a module eigengene value and, as expected, there were significant associations between module eigengene values and severity, parasite load, and laboratory parameters (Figure 3). Only the *HSPH1* (heat shock protein family H (Hsp110) member 1) module contained more than 10 genes from both human and parasite, strongly enriched in human heat shock response genes and parasite RNA metabolism genes, perhaps indicating that these parasite genes particularly promote human cell stress. Some host-dominated and parasite-dominated modules were also highly correlated with each other, most notably the *RPL24* (ribosomal protein L24) module (highly enriched in translation pathways) was strongly correlated with the remarkably homologous *PF3D7_0721600* (putative 40S ribosomal protein S5) parasite module. We excluded read mapping errors as an explanation for this, and suggest that this indicates co-regulation of conserved host and parasite translation machinery. Furthermore, most of these genes were also differentially expressed between SM and UM, perhaps indicating a "molecular arms race" between parasite and host to synthesise proteins which may, in excess, contribute to collateral tissue damage.

Co-expression network modules can be used as units of analysis, affording considerable dimension reduction for whole-genome expression data. We used module eigengene values[37, 38] and parasite load (with which many modules were correlated, Fig 3) in linear regression models to determine the best within-sample predictors of severity, starting with all significant univariate associations and proceeding by backward selection (Supplementary Table 10). The best multivariate model combined *MMP8*, *OAS1* (2'-5'-oligoadenylate synthetase 1) and *LYSMD3* (LysM, putative

peptidoglycan-binding, domain containing 3) module eigengenes, but not parasite load. Interestingly these modules represent distinct aspects of the immune response: the *MMP8* module, highly enriched in defence response genes with predicted upstream regulators CEBPA (CCAAT/enhancer binding protein alpha, a myeloid transcription factor) and CSF3, likely reflects granulopoiesis[22]; the *OAS1* module is highly enriched for type 1 interferon response genes; the small *LYSMD3* module, with limited GO enrichment, contains a functional network around interferon-γ (Supplementary Figure 5). The direction of association of the *OAS1* module with severity changed from negative in univariate analysis to positive in the multivariate analysis, suggesting that inadequate downregulation of the type-1 interferon response in conjunction with upregulation of granulopoiesis and interferon-γ signalling may contribute to pathogenesis.

Considering all subjects together for generation of co-expression networks maximises power to detect consistently co-regulated genes but may not identify sets of genes where co-regulation is altered by severity. For this reason we also created separate co-expression networks for UM and SM and compared the modules to identify differential co-expression (Fig 4, Supplementary Table 11). Eight modules showed significant preservation between networks, seven were partially preserved, and two were unique to SM (Figure 4a, Supplementary Table 11). Partial preservation was common amongst modules comprised predominantly from human or parasite genes (Figure 4a,b), and module preservation was not dependent on the proportion of module genes differentially expressed between SM and UM (Figure 4a,c). Again, a *MMP8* module was identified (exclusively human genes, many encoding neutrophil granule and phagosome components), unique to SM with 38% of genes significantly differentially expressed between SM and UM, enriched in host defence pathways (Supplementary Table 11) and predicted to be regulated by CEBPA, CSF3 and TNF. These findings strongly suggest this module represents emergency granulopoiesis[22] and mark this as a specific feature of SM. The *TIPRL* (TOR Signaling Pathway Regulator) module (99.2% human genes) was also unique to SM but contained very few (1.3%) DEGs, had limited GO enrichment (Supplementary Table

11), and the most significant predicted upstream regulator was the transcription factor HNF4A. Both TIPRL and HNF4A have regulatory roles in metabolic, inflammatory and apoptosis signal pathways, so the minimal change in expression of this module may represent an aberrant response in SM. Amongst the partially preserved modules we found evidence that host and parasite translation pathways were more tightly co-regulated in SM than UM, genes being distributed across fewer modules in SM (Fig 4a, Supplementary Table 11).

Discussion

We have shown that dual-RNA sequencing can be used to identify systemic host-pathogen interactions and potential pathogenic mechanisms associated with severe infection in humans. The differences in human and parasite gene expression between SM and UM were much clearer after adjusting for heterogeneity of leukocyte population and parasite developmental stage. Although the importance of accounting for such variation is well recognised[14], it has rarely been done in malaria or other infectious disease transcriptomic studies.

Our most striking finding came from integrating parasite load with global gene expression, revealing the overriding effect of parasite load on the differences in human gene expression between SM and UM. Previous studies have examined the association between human gene expression and circulating parasitemia[13, 23, 24], but we found that estimation of total body parasite load was necessary to appreciate the full effect on host response. Our findings imply that SM is not the consequence of an excessive host response, but that there is an appropriate host response to an excessive pathogen load. This has important implications for other infectious disease, immunology, and pathogenesis research in humans. Total body pathogen load is much harder to measure in other infections in humans[39], yet failure to account for it may lead to misinterpretation of associations between host factors and severity or protection.

Despite the dominant effect of parasite load, we found that specific sets of genes induced by infection were associated with different pathophysiological consequences of malaria. Distinct sets of

genes were correlated lactate concentration and platelet count, and associated with different clinical presentations of SM. Alternative analytical approaches repeatedly identified the association of genes expressed during neutrophil granulopoiesis (such as *MMP8*) and translation pathways with severe outcomes. There is plentiful evidence that neutrophil granule proteins are released in severe malaria[40, 41], can impair vascular endothelial functions such as barrier integrity[26, 27], and may therefore have a direct role in the pathogenesis of SM. Unfortunately neutrophil related signatures are not differentially expressed in the whole blood transcriptome of the widely used rodent experimental cerebral malaria model[42], which means that experimental testing of the role of neutrophils may be challenging.

We observed an intriguing relationship between type 1 interferon responses and severity, which may help to tie together data from previous observations in humans and animal models. A previous small study found higher expression of type-1 interferon response genes in UM than SM and suggested that this may be protective against developing SM[43]. However we found that type-1 interferon response genes were negatively correlated with parasite load, indicating that downregulation with increasing parasite load (and severity) is a more likely explanation. When we performed multivariate analyses using gene expression modules to explain severity, our results suggested that insufficient downregulation of type 1 interferons was in fact associated with severity. This would be more consistent with results in several animal models where genetic or antibody-mediated ablation of type-1 interferon signalling improves outcome[44-47].

The role of translation pathways is more speculative, but co-regulation of these genes between host and parasite, which becomes tighter in more severe disease, implies that there may be an inter-species feedback loop. Increased translation is important for production of host defence effector proteins[48] and parasite proteins which enable survival[49]. Perhaps, as parasite load increases the host response increases, the parasite produces more proteins necessary to survive, and the cycle amplifies until parasite load and host response cause host organ damage and severe disease.

In addition to translation-related genes, we also identified hundreds of other parasite genes which were associated with severe disease. Many of these genes have as yet unknown function. However the enrichment of genes involved in protein transport, for example, suggests there may be layers of control which determine parasite protein export into the host cell and the molecular host-parasite interactions which predispose to SM. Some of these aspects of parasite gene regulation may only be appreciated *in vivo*, in the parasite's natural environment.

The data we have generated and comprehensive analyses we have performed provide a unique and valuable resource for the research community. These should be launch points for future studies using alternative approaches to assess whether the mechanisms we have implicated through gene expression do indeed play causal roles in SM and may be targets for much needed adjunctive therapies.

Methods

Subjects and samples

Gambian children (under 16 years old) with *P. falciparum* malaria were recruited from three peri-urban health centres, The MRC Gate Clinic, Brikama Health Centre, and The Jammeh Foundation for Peace Hospital, Serekunda, as part of a larger study of severe malaria[7, 50, 51]. Informed consent was obtained from the child's parent or legal guardian for collection and subsequent use of samples. The study was approved by the Gambian Government / MRC Laboratories Joint Ethics Committee. All children underwent full clinical examination and were managed in accordance with the Gambian government guidelines. Malaria was defined by the occurrence of fever in the last 48 hours before recruitment and >5000 asexual parasites/μL in the peripheral blood. Subjects were further categorized into different severe malaria phenotypes using modified World Health Organization criteria: cerebral malaria (CM) was defined as Blantyre Coma Score (BCS) of 1 or 2, or a BCS of 3 if the motor response was 1, AND no hypoglycaemia, no rapid improvement in response to fluid resuscitation, no suspicion of meningitis; hyperlactatemia (HL), blood lactate concentration > 5mmol/L; both CM and HL (CH)[7]. At the time of presentation to the clinic, prior to any antimalarial treatment or blood transfusion, capillary blood was used for measurement of lactate and glucose concentrations and thick and thin blood films, venous blood was collected into EDTA for sickle cell screen and full blood count, PAXgene blood RNA tube (BD), and sodium heparin (BD) for plasma separation[50]. Parasitemia was calculated using 50 high power fields on Giemsa-stained thin blood smears. Plasma *P. falciparum* histidine-rich protein II (PfHRP2) was measured by ELISA (Cellabs)[7].

For the present study we used 46 subjects selected from those with ≥1μg RNA available which showed no / minimal evidence of degradation on visual inspection of a Bioanalyser (Agilent) trace (RNA integrity number calculations are not valid for dual species RNA analysis). To reduce potential confounding we aimed to frequency match subjects between SM and UM groups as closely as possible by age and gender, and if there remained a choice of samples available we selected those

with the most complete additional clinical and laboratory data. For UM samples we aimed to include an equal number with parasitemia above and below 5% (to maximise the chance of obtaining parasite reads in some of the subjects). For SM samples we aimed to include subjects with each of the common SM phenotypes seen in this part of the Gambia in approximately equal numbers, although final numbers were determined by availability and quality of RNA. Detailed information about the study subjects is shown in Supplementary Table 1 and Supplementary Dataset 1. Characteristics were compared between subject groups using one-way ANOVA for continuous data and Fisher's exact test for categorical data.

RNA sequencing

Total RNA was extracted using the PAXgene Blood RNA kit (BD). Libraries were prepared from 1µg of total RNA using the ScriptSeq v2 RNA-seq library preparation kit (Illumina) with additional steps to remove ribosmal RNA (rRNA) and globin messenger RNA (mRNA) using the Globin-Zero Gold kit (Epicentre). Strand-specific libraries were sequenced using the 2x100 bp protocol with an Illumina HiSeq 2500 instrument. In order to eliminate batch effects, samples were randomized for the order of library preparation. For sequencing, 5-6 samples were run per lane, and each lane contained at least one sample from each disease type, randomly allocated in a block design. Library preparation and sequencing were carried out by Exeter University sequencing service.

Genomes and RNA annotations

Human reference genome (hg38) was obtained from UCSC genome browser (http://genome.ucsc.edu/) and *P. falciparum* reference genome (release 24) was obtained from PlasmoDB (http://plasmodb.org/). Human gene annotation was obtained from GENCODE (release 22) (http://gencodegenes.org/releases/) and *P. falciparum* gene annotation from PlasmoDB (release 24) (http://plasmodb.org).

Read Mapping and quantification

RNA-seq data was mapped to the combined genomic index containing both human and *P. falciparum* genomes using the splice-aware STAR aligner, allowing up to 8 mismatches for each paired-end read[52]. Reads were extracted from the output BAM file to separate parasite-mapped reads from human-mapped reads. Reads mapping to both genomes were counted for each sample and removed. BAM files were sorted, read groups replaced with a single new read group and all reads assigned to it, and indexed to run RNA-SeQC, a tool for computing quality control metrics for RNA-seq data[53]. HTSeq-count was used to count the reads mapped to exons with the parameter "-m union"[54]. Only uniquely mapping reads were counted.

Since our analysis of *P. falciparum* gene expression was reliant on a reference genome, families of highly polymorphic *var*, *stevor*, and *rifin* genes were removed from downstream analyses as these exhibit great sequence diversity between parasites and are likely to be incorrectly characterized[12]. Additional highly polymorphic regions within the *P. falciparum* genome which might also be incorrectly characterized were identified using schizont stage RNA-seq data from 9 clinical isolates (Duffy et al., manuscript submitted). In total, 139 genes were identified with highly polymorphic regions. A reference GTF file containing *P. falciparum* gene annotations was modified to remove these regions without removing the genes, and the resulting read count data generated using the modified GTF file was used for downstream analysis.

Outlier identification

With the R package edgeR, raw read counts of each data set were normalized using a trimmed mean of M-values (TMM), which takes into account the library size and the RNA composition of the input data[55]. A multi-dimensional scaling (MDS) plot was used to identify the distances between samples that correspond to leading biological coefficient of variation. Up to the 6th dimension of MDS was plotted to fully observe the variation between samples, with two dimensions visualized at a time in scatter plot format. Three parasite samples were consistently found to be positioned away from other samples in each pair of dimensions, indicating outliers. This was further supported by low

correlations observed between either of these outliers with other samples. These three samples were excluded from further parasite gene expression analysis: one sample (HL_478) had very low parasite reads making estimation of gene expression impossible and the other two samples (CH_285 and UM_589) were conspicuous outliers on MDS plots, possibly due to imperfect library preparation.

Deconvolution analysis

To account for inter-individual variation in the proportions of different types of blood leukocyte, and for variation in the distribution of circulating parasites through the intraerythrocytic developmental cycle, deconvolution analysis was performed on RNA-seq data using CellCODE[15]. This uses a multi-step statistical framework to compute the relative differences in cell proportion represented as surrogate proportion variables (SPVs). It requires a reference data set that contains gene expression profiles for each cell type of interest. Five major immune cell populations were selected from Immune Response In Silico (IRIS)[56] to constitute the human reference data set: neutrophil, monocyte, CD4+ T-cell, CD8+ T-cell, and B-cell. Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values were calculated from human RNA-seq data and log-transformed to simulate a microarray data set. For the parasite reference data set, RNA-seq data sets were obtained for four specific stages in the parasite asexual and sexual stage (0 hour, 24 hour, 48 hour, and gametocyte stage V)[16, 18], normalized by relative library sizes of samples (i.e. size factors) using edgeR. An identical normalization method was also applied for the input parasite RNA-seq data. A trial-and-error approach was taken to obtain the optimum SPV values for each cell-type. For human deconvolution, a cutoff value of 1.2 and a maximum number of marker genes of 50 appeared optimal. For parasite deconvolution, a cutoff value of 1.7 and a maximum number of marker genes of 50 appeared optimal.

Validation of CellCODE for *P. falciparum* developmental stage deconvolution (for which its use has not previously been reported) was performed by comparison with previously reported “stage-specific” marker genes[57] and by assessing performance in synthetic data sets constructed by

mixing together in varying proportions randomly selected reads from RNA-seq reference datasets[16, 18] of the different parasite developmental stages.

Differential gene expression and linear regression analysis

Prior to carrying out any downstream analyses genes with very low TMM-normalized read counts (< 5 counts-per-million (cpm) in < 3 samples and undetected in the remainder) were excluded. The generalized linear model tool in edgeR was employed to perform differential gene expression analysis (DGEA) between disease groups with adjustment for leukocyte and parasite SPVs, and in subsequent analyses additional adjustment for log parasite density and log PfHRP2.

Linear regression analysis was performed in edgeR to identify genes significantly associated with clinical variables of interest. Input gene expression values included adjustment for SPVs. The variables considered were: log PfHRP2, log parasite density, lactate concentration, platelet counts, and hemoglobin concentration. Additional analysis for lactate, platelets and hemoglobin were conducted including adjustment for log parasite density and log PfHRP2.

In both DGEA and linear regression analyses, false Discovery Rate (FDR) was computed for each individual analysis using the Benjamini-Hochberg procedure[58]. Genes with FDR below 0.05 were considered to be differentially expressed.

Gene ontology and KEGG pathway enrichment analysis

Gene ontology (GO) terms for genes were obtained from Bioconductor package “org.Hs.eg.db” for human and “org.Pf.plasmo.db” for parasite. Input gene lists were significantly differentially expressed genes or genes that were significantly associated with laboratory variables. Fisher’s exact test was used to identify significantly over-represented GO terms from these gene lists. The background sets for each species consisted of all expressed genes detected in the data set with the exclusion of those with very low expression as described above. Enrichment analysis for biological process terms was carried out using the "goana()" function in edgeR. The least redundant GO terms with greatest significance in each analysis were identified for reporting using the tool REVIGO[59].

Ingenuity Pathway Analysis (Qiagen) was used for prediction of upstream regulators of groups of differentially expressed genes, and to identify functional networks.

Construction of a coexpression network

The weighted gene coexpression network analysis (WGCNA) tool was used to construct a gene coexpression network[38]. The input data for WGCNA was read counts for each gene feature normalized using TMM method and then adjusted for SPVs using the command "removeBatchEffect()" from the R package edgeR. Both human and parasite expression data were analyzed together as a single set of genes for each subject. In order to comprehensively study the relationships between genes, two sets of networks were created: one with all samples from SM and UM groups, and the other with two separate sub-networks, generated from samples from SM and UM groups respectively. Network creation was conducted following the WGCNA tool guidelines:

1) Hierarchical clustering was performed at a sample level to detect outliers based on the WGCNA tool threshold, which were removed from the subsequent network generation (HL_171 and UM_492).

2) An appropriate soft-thresholding power (b) was chosen by applying the scale-free topology criterion. This was such that the power value enables the resulting gene network to satisfy the scale-free topology of approximately ($R^2 > 0.80$).

3) Adjacency, which represents the connection strength of two genes in a network, was calculated. Coexpression similarity was calculated by taking the absolute value of the correlation coefficient, multiplying by 0.5 and adding 0.5 to create a signed network, where the presence of strongly negatively correlated gene pairs is downsized.

4) The adjacency matrix was transformed into a topological overlap matrix (TOM) in order to minimize the effects of spurious associations and noise in the network.

5) Hierarchical clustering on TOM dissimilarity was done to create hierarchical clustering tree of genes.

6) The dynamic tree cut method was used to group the genes that are highly correlated with one another into gene modules where minimum module size and the tree height at which genes below the height is grouped together were specified.

7) Module eigengene values for each module were calculated, which represents the overall gene expression profile of a module. Correlation analysis between modules was performed using eigengene values to identify modules with high similarity, which were then merged together.

The resulting network consisted of genes (represented as nodes in the network) and correlations between genes (represented as edges in the network), and highly correlated genes grouped together into modules. To characterize the gene network, several analysis steps were carried out. The most connected genes in each module were identified as the hub genes. Based on the module eigengene value, the connections between modules were determined. Pearson correlation analysis between module eigengene values and clinical variables was performed to identify gene clusters that are highly associated with clinical traits. Gene set enrichment analysis was performed on each module to identify significantly enriched GO terms. This data was summarised using OmicCircos[60].

From two separate sub-networks generated from SM and UM groups respectively, the preservation of gene connections across SM and UM groups was determined by assessing an overlap of genes for each module pair (from SM and UM sub-networks respectively), the significance of overlap was measured using the hypergeometric test. For each significantly preserved module pair, the hub genes and the significantly enriched GO terms were compared.

The gene network was exported to Cytoscape (http://www.cytoscape.org/) for visualisation. Only gene pairs with adjacency value of 0.03 or higher were exported to remove genes with low connections from the network visualization. SM and UM sub-networks were exported separately and subsequently combined into a single network using the Cytoscape embedded tool "Merge". By doing so, duplicate genes representing overlap between SM and UM sub-networks were removed, and connections between genes remained intact such that genes that can only be found on SM sub-network and also connected to the genes that can be found on both networks were not connected to the genes that can be found on UM sub-network and also connected to the same overlapping genes.

Logistic regression for association of module eigengenes with severity

Logistic regression was performed using the glm package in R to identify module eigengene values with univariate association with severity. All modules with significant univariate associations

(P<0.01) in addition to log PfHRP2 concentration were used in backward selection to identify the best multivariate model in which all terms were significant.

Data availability

Sequence data that support the findings of this study will be deposited in ArrayExpress with the accession codes made available at the time of publication. Source data for Supplementary Table 1 are provided with the paper as Supplementary Dataset 1.

Acknowledgements

This work was funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement and is also part of the EDCTP2 programme supported by the European Union (MR/L006529/1), MRC core funding of the MRC Gambia Unit (MRCG), and Wellcome Trust (098051). We are grateful to the study subjects, staff at MRCG, Jammeh Foundation for Peace Hospital, and Brikama Health Centre; Konrad Paszkiewicz and staff at Exeter Sequencing Service at the University of Exeter (supported by Medical Research Council Clinical Infrastructure award (MR/M008924/1), Wellcome Trust Institutional Strategic Support Fund (WT097835MF), Wellcome Trust Multi User Equipment Award (WT101650MA) and BBSRC LOLA award (BB/K003240/1)).

Author contributions

AJC, ML and DJC conceived the study; HJL, TDO and AG performed formal analysis; MW and AJC performed investigations on clinical samples; MW, DJC, DN and LBS provided samples; DJC, TDO and LBS provided methodology; AJC and HJL wrote the original draft; all authors contributed to review and editing of the manuscript; LJC, DJC, ML, TDO and AJC provided supervision; AJC obtained funding.

Competing financial interests

504 The authors declare no competing financial interests.

505

506 **Materials & Correspondence**

507 Correspondence should be addressed to Dr Aubrey Cunnington, Clinical Senior Lecturer,

508 Section of Paediatrics, Department of Medicine, Imperial College London (St Mary's Campus)

509 231, Medical School Building, Norfolk Place, London W2 1PG. a.cunnington@imperial.ac.uk

510 **References**

511 1. Phillips MA, Burrows JN, Manyando C, van Huijsduijnen RH, Van Voorhis WC, Wells TNC.
512 Malaria. Nat Rev Dis Primers. 2017;3:17050. doi: 10.1038/nrdp.2017.50. PubMed PMID: 28770814.
513 2. Blasco B, Leroy D, Fidock DA. Antimalarial drug resistance: linking Plasmodium falciparum
514 parasite biology to the clinic. Nat Med. 2017;23(8):917-28. doi: 10.1038/nm.4381. PubMed PMID:
515 28777791.
516 3. Ranson H, Lissenden N. Insecticide Resistance in African Anopheles Mosquitoes: A
517 Worsening Situation that Needs Urgent Action to Maintain Malaria Control. Trends Parasitol.
518 2016;32(3):187-96. doi: 10.1016/j.pt.2015.11.010. PubMed PMID: 26826784.
519 4. Churcher TS, Lissenden N, Griffin JT, Worrall E, Ranson H. The impact of pyrethroid
520 resistance on the efficacy and effectiveness of bednets for malaria control in Africa. eLife. 2016;5.
521 doi: 10.7554/eLife.16090. PubMed PMID: 27547988; PubMed Central PMCID: PMC5025277.
522 5. Cunnington AJ, Walther M, Riley EM. Piecing together the puzzle of severe malaria. Sci Transl
523 Med. 2013;5(211):211ps18. doi: 10.1126/scitranslmed.3007432. PubMed PMID: 24225942.
524 6. Wassmer SC, Grau GE. Severe malaria: what's new on the pathogenesis front? Int J Parasitol.
525 2017;47(2-3):145-52. doi: 10.1016/j.ijpara.2016.08.002. PubMed PMID: 27670365; PubMed Central
526 PMCID: PMC5285481.
527 7. Cunnington AJ, Bretscher MT, Nogaro SI, Riley EM, Walther M. Comparison of parasite
528 sequestration in uncomplicated and severe childhood Plasmodium falciparum malaria. J Infect.
529 2013;67(3):220-30. doi: 10.1016/j.jinf.2013.04.013. PubMed PMID: 23623771; PubMed Central
530 PMCID: PMC3744804.
531 8. Hendriksen IC, Mwanga-Amumpaire J, von Seidlein L, Mtove G, White LJ, Olaosebikan R, et
532 al. Diagnosing severe falciparum malaria in parasitaemic African children: a prospective evaluation of
533 plasma PfHRP2 measurement. PLoS Med. 2012;9(8):e1001297. Epub 2012/08/29. doi:
534 10.1371/journal.pmed.1001297. PubMed PMID: 22927801; PubMed Central PMCID: PMC3424256.

9. Hendriksen IC, White LJ, Veenemans J, Mtove G, Woodrow C, Amos B, et al. Defining falciparum-malaria-attributable severe febrile illness in moderate-to-high transmission settings on the basis of plasma PfHRP2 concentration. J Infect Dis. 2013;207(2):351-61. doi: 10.1093/infdis/jis675. PubMed PMID: 23136222; PubMed Central PMCID: PMC3532834.

10. Dondorp AM, Desakorn V, Pongtavornpinyo W, Sahassananda D, Silamut K, Chotivanich K, et al. Estimation of the total parasite biomass in acute falciparum malaria from plasma PfHRP2. PLoS Med. 2005;2(8):e204. Epub 2005/08/18. doi: 04-PLME-RA-0297R2 [pii] 10.1371/journal.pmed.0020204. PubMed PMID: 16104831; PubMed Central PMCID: PMC1188247.

11. Craig AG, Grau GE, Janse C, Kazura JW, Milner D, Barnwell JW, et al. The role of animal models for research on severe malaria. PLoS Pathog. 2012;8(2):e1002401. doi: 10.1371/journal.ppat.1002401. PubMed PMID: 22319438; PubMed Central PMCID: PMC3271056.

12. Wahlgren M, Goel S, Akhouri RR. Variant surface antigens of Plasmodium falciparum and their roles in severe malaria. Nat Rev Microbiol. 2017;15(8):479-91. doi: 10.1038/nrmicro.2017.47. PubMed PMID: 28603279.

13. Yamagishi J, Natori A, Tolba ME, Mongan AE, Sugimoto C, Katayama T, et al. Interactive transcriptome analysis of malaria patients and infecting Plasmodium falciparum. Genome Res. 2014;24(9):1433-44. doi: 10.1101/gr.158980.113. PubMed PMID: 25091627; PubMed Central PMCID: PMC4158759.

14. Shen-Orr SS, Gaujoux R. Computational deconvolution: extracting cell type-specific information from heterogeneous samples. Curr Opin Immunol. 2013;25(5):571-8. doi: 10.1016/j.coi.2013.09.015. PubMed PMID: 24148234; PubMed Central PMCID: PMC3874291.

15. Chikina M, Zaslavsky E, Sealfon SC. CellCODE: a robust latent variable approach to differential expression analysis for heterogeneous cell populations. Bioinformatics. 2015;31(10):1584-91. doi: 10.1093/bioinformatics/btv015. PubMed PMID: 25583121; PubMed Central PMCID: PMCPMC4426841.

560 16. Otto TD, Wilinski D, Assefa S, Keane TM, Sarry LR, Bohme U, et al. New insights into the
561 blood-stage transcriptome of Plasmodium falciparum using RNA-Seq. Mol Microbiol. 2010;76(1):12-
562 24. doi: 10.1111/j.1365-2958.2009.07026.x. PubMed PMID: 20141604; PubMed Central PMCID:
563 PMC2859250.
564 17. Hoeijmakers WA, Bartfai R, Stunnenberg HG. Transcriptome analysis using RNA-Seq.
565 Methods Mol Biol. 2013;923:221-39. doi: 10.1007/978-1-62703-026-7_15. PubMed PMID:
566 22990781.
567 18. Lopez-Barragan MJ, Lemieux J, Quinones M, Williamson KC, Molina-Cruz A, Cui K, et al.
568 Directional gene expression and antisense transcripts in sexual and asexual stages of Plasmodium
569 falciparum. BMC Genomics. 2011;12:587. doi: 10.1186/1471-2164-12-587. PubMed PMID:
570 22129310; PubMed Central PMCID: PMC3266614.
571 19. Blankley S, Berry MP, Graham CM, Bloom CI, Lipman M, O'Garra A. The application of
572 transcriptional blood signatures to enhance our understanding of the host response to infection: the
573 example of tuberculosis. Phil Trans R Soc Lond B, Biol Sci. 2014;369(1645):20130427. doi:
574 10.1098/rstb.2013.0427. PubMed PMID: 24821914; PubMed Central PMCID: PMC4024221.
575 20. Daily JP, Scanfeld D, Pochet N, Le Roch K, Plouffe D, Kamal M, et al. Distinct physiological
576 states of Plasmodium falciparum in malaria-infected patients. Nature. 2007;450(7172):1091-5. doi:
577 10.1038/nature06311. PubMed PMID: 18046333.
578 21. Coleman BI, Duraisingh MT. Transcriptional control and gene silencing in Plasmodium
579 falciparum. Cell Microbiol. 2008;10(10):1935-46. doi: 10.1111/j.1462-5822.2008.01203.x. PubMed
580 PMID: 18637022.
581 22. Cowland JB, Borregaard N. Granulopoiesis and granules of human neutrophils. Immunol Rev.
582 2016;273(1):11-28. doi: 10.1111/imr.12440. PubMed PMID: 27558325.
583 23. Griffiths MJ, Shafi MJ, Popper SJ, Hemingway CA, Kortok MM, Wathen A, et al. Genomewide
584 analysis of the host response to malaria in Kenyan children. J Infect Dis. 2005;191(10):1599-611.
585 Epub 2005/04/20. doi: JID33481 [pii]

586 10.1086/429297. PubMed PMID: 15838786.

587 24. Idaghdour Y, Quinlan J, Goulet JP, Berghout J, Gbeha E, Bruat V, et al. Evidence for additive
588 and interaction effects of host genotype and infection in malaria. Proc Natl Acad Sci U S A.
589 2012;109(42):16786-93. doi: 10.1073/pnas.1204945109. PubMed PMID: 22949651; PubMed Central
590 PMCID: PMC3479498.

591 25. Cunnington AJ, Riley EM, Walther M. Stuck in a rut? Reconsidering the role of parasite
592 sequestration in severe malaria syndromes. Trends Parasitol. 2013;29(12):585-92. doi:
593 10.1016/j.pt.2013.10.004. PubMed PMID: 24210256; PubMed Central PMCID: PMC3880783.

594 26. Vandenbroucke RE, Dejonckheere E, Van Lint P, Demeestere D, Van Wonterghem E,
595 Vanlaere I, et al. Matrix metalloprotease 8-dependent extracellular matrix cleavage at the blood-CSF
596 barrier contributes to lethality during systemic inflammatory diseases. J Neurosci. 2012;32(29):9805-
597 16. doi: 10.1523/JNEUROSCI.0967-12.2012. PubMed PMID: 22815495.

598 27. Schubert-Unkmeir A, Konrad C, Slanina H, Czapek F, Hebling S, Frosch M. Neisseria
599 meningitidis induces brain microvascular endothelial cell detachment from the matrix and cleavage
600 of occludin: a role for MMP-8. PLoS Pathog. 2010;6(4):e1000874. doi:
601 10.1371/journal.ppat.1000874. PubMed PMID: 20442866; PubMed Central PMCID: PMC2861698.

602 28. Sasai M, Oshiumi H, Matsumoto M, Inoue N, Fujita F, Nakanishi M, et al. Cutting Edge: NF-
603 kappaB-activating kinase-associated protein 1 participates in TLR3/Toll-IL-1 homology domain-
604 containing adapter molecule-1-mediated IFN regulatory factor 3 activation. J Immunol.
605 2005;174(1):27-30. PubMed PMID: 15611223.

606 29. Mooney JP, Wassmer SC, Hafalla JC. Type I Interferon in Malaria: A Balancing Act. Trends
607 Parasitol. 2017;33(4):257-60. doi: 10.1016/j.pt.2016.12.010. PubMed PMID: 28094200.

608 30. Tahar R, Albergaria C, Zeghidour N, Ngane VF, Basco LK, Roussilhon C. Plasma levels of eight
609 different mediators and their potential as biomarkers of various clinical malaria conditions in African
610 children. Malar J. 2016;15:337. doi: 10.1186/s12936-016-1378-3. PubMed PMID: 27357958; PubMed
611 Central PMCID: PMC4928328.

612 31. Chimma P, Roussilhon C, Sratongno P, Ruangveerayuth R, Pattanapanyasat K, Perignon JL, et
613 al. A distinct peripheral blood monocyte phenotype is associated with parasite inhibitory activity in
614 acute uncomplicated Plasmodium falciparum malaria. PLoS Pathog. 2009;5(10):e1000631. doi:
615 10.1371/journal.ppat.1000631. PubMed PMID: 19851453; PubMed Central PMCID: PMC2759288.
616 32. Cserti-Gazdewich CM, Dhabangi A, Musoke C, Ssewanyana I, Ddungu H, Nakiboneka-
617 Ssenabulya D, et al. Inter-relationships of cardinal features and outcomes of symptomatic pediatric
618 Plasmodium falciparum malaria in 1,933 children in Kampala, Uganda. Am J Trop Med Hyg.
619 2013;88(4):747-56. doi: 10.4269/ajtmh.12-0668. PubMed PMID: 23358640; PubMed Central PMCID:
620 PMC3617864.
621 33. Moxon CA, Chisala NV, Mzikamanda R, MacCormick I, Harding S, Downey C, et al. Laboratory
622 evidence of disseminated intravascular coagulation is associated with a fatal outcome in children
623 with cerebral malaria despite an absence of clinically evident thrombosis or bleeding. J Thromb
624 Haemost. 2015;13(9):1653-64. doi: 10.1111/jth.13060. PubMed PMID: 26186686; PubMed Central
625 PMCID: PMC4605993.
626 34. O'Sullivan JM, Preston RJ, O'Regan N, O'Donnell JS. Emerging roles for hemostatic
627 dysfunction in malaria pathogenesis. Blood. 2016;127(19):2281-8. doi: 10.1182/blood-2015-11-
628 636464. PubMed PMID: 26851291.
629 35. Fuchs TA, Bhandari AA, Wagner DD. Histones induce rapid and profound thrombocytopenia
630 in mice. Blood. 2011;118(13):3708-14. doi: 10.1182/blood-2011-01-332676. PubMed PMID:
631 21700775; PubMed Central PMCID: PMC3186342.
632 36. Westermann AJ, Barquist L, Vogel J. Resolving host-pathogen interactions by dual RNA-seq.
633 PLoS Pathog. 2017;13(2):e1006033. doi: 10.1371/journal.ppat.1006033. PubMed PMID: 28207848;
634 PubMed Central PMCID: PMC5313147.
635 37. Fuller TF, Ghazalpour A, Aten JE, Drake TA, Lusis AJ, Horvath S. Weighted gene coexpression
636 network analysis strategies applied to mouse weight. Mamm. Genome. 2007;18(6-7):463-72. doi:
637 10.1007/s00335-007-9043-3. PubMed PMID: 17668265; PubMed Central PMCID: PMC1998880.

638 38. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis.
639 BMC Bioinformatics. 2008;9:559. doi: 10.1186/1471-2105-9-559. PubMed PMID: 19114008; PubMed
640 Central PMCID: PMC2631488.
641 39. Cunnington AJ. The importance of pathogen load. PLoS Pathog. 2015;11(1):e1004563. doi:
642 10.1371/journal.ppat.1004563. PubMed PMID: 25569282; PubMed Central PMCID: PMC4287534.
643 40. Dietmann A, Helbok R, Lackner P, Issifou S, Lell B, Matsiegui PB, et al. Matrix
644 metalloproteinases and their tissue inhibitors (TIMPs) in Plasmodium falciparum malaria: serum
645 levels of TIMP-1 are associated with disease severity. J Infect Dis. 2008;197(11):1614-20. doi:
646 10.1086/587943. PubMed PMID: 18700258.
647 41. Feintuch CM, Saidi A, Seydel K, Chen G, Goldman-Yassen A, Mita-Mendoza NK, et al.
648 Activated Neutrophils Are Associated with Pediatric Cerebral Malaria Vasculopathy in Malawian
649 Children. mBio. 2016;7(1):e01300-15. doi: 10.1128/mBio.01300-15. PubMed PMID: 26884431;
650 PubMed Central PMCID: PMC4791846.
651 42. Oakley MS, Anantharaman V, Venancio TM, Zheng H, Mahajan B, Majam V, et al. Molecular
652 correlates of experimental cerebral malaria detectable in whole blood. Infect Immun.
653 2011;79(3):1244-53. doi: 10.1128/IAI.00964-10. PubMed PMID: 21149594; PubMed Central PMCID:
654 PMC3067492.
655 43. Krupka M, Seydel K, Feintuch CM, Yee K, Kim R, Lin CY, et al. Mild Plasmodium falciparum
656 malaria following an episode of severe malaria is associated with induction of the interferon
657 pathway in Malawian children. Infect Immun. 2012;80(3):1150-5. doi: 10.1128/IAI.06008-11.
658 PubMed PMID: 22232187; PubMed Central PMCID: PMC3294667.
659 44. Haque A, Best SE, Ammerdorffer A, Desbarrieres L, de Oca MM, Amante FH, et al. Type I
660 interferons suppress CD4(+) T-cell-dependent parasite control during blood-stage Plasmodium
661 infection. Eur J Immunol. 2011;41(9):2688-98. doi: 10.1002/eji.201141539. PubMed PMID:
662 21674481.

663 45. Haque A, Best SE, Montes de Oca M, James KR, Ammerdorffer A, Edwards CL, et al. Type I
664 IFN signaling in CD8- DCs impairs Th1-dependent malaria immunity. J Clin Invest. 2014;124(6):2483-
665 96. doi: 10.1172/JCI70698. PubMed PMID: 24789914; PubMed Central PMCID: PMC4038565.
666 46. Sharma S, DeOliveira RB, Kalantari P, Parroche P, Goutagny N, Jiang Z, et al. Innate immune
667 recognition of an AT-rich stem-loop DNA motif in the Plasmodium falciparum genome. Immunity.
668 2011;35(2):194-207. Epub 2011/08/09. doi: S1074-7613(11)00301-3 [pii]
669 10.1016/j.immuni.2011.05.016. PubMed PMID: 21820332; PubMed Central PMCID: PMC3162998.
670 47. Zander RA, Guthmiller JJ, Graham AC, Pope RL, Burke BE, Carr DJ, et al. Type I Interferons
671 Induce T Regulatory 1 Responses and Restrict Humoral Immunity during Experimental Malaria. PLoS
672 Pathog. 2016;12(10):e1005945. doi: 10.1371/journal.ppat.1005945. PubMed PMID: 27732671;
673 PubMed Central PMCID: PMC5061386.
674 48. Arguello RJ, Rodriguez Rodrigues C, Gatti E, Pierre P. Protein synthesis regulation, a pillar of
675 strength for innate immunity? Curr Opin Immunol. 2015;32:28-35. doi: 10.1016/j.coi.2014.12.001.
676 PubMed PMID: 25553394.
677 49. Vembar SS, Droll D, Scherf A. Translational regulation in blood stages of the malaria parasite
678 Plasmodium spp.: systems-wide studies pave the way. Wiley Interdisciplin Rev RNA. 2016;7(6):772-
679 92. doi: 10.1002/wrna.1365. PubMed PMID: 27230797; PubMed Central PMCID: PMC5111744.
680 50. Walther M, Jeffries D, Finney OC, Njie M, Ebonyi A, Deininger S, et al. Distinct roles for
681 FOXP3 and FOXP3 CD4 T cells in regulating cellular immunity to uncomplicated and severe
682 Plasmodium falciparum malaria. PLoS Pathog. 2009;5(4):e1000364. Epub 2009/04/04. doi:
683 10.1371/journal.ppat.1000364. PubMed PMID: 19343213; PubMed Central PMCID: PMC2658808.
684 51. Walther M, De Caul A, Aka P, Njie M, Amambua-Ngwa A, Walther B, et al. HMOX1 gene
685 promoter alleles and high HO-1 levels are associated with severe malaria in Gambian children. PLoS
686 Pathog. 2012;8(3):e1002579.

687 52. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal
688 RNA-seq aligner. Bioinformatics. 2013;29(1):15-21. doi: 10.1093/bioinformatics/bts635. PubMed
689 PMID: 23104886; PubMed Central PMCID: PMC3530905.
690 53. DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C, et al. RNA-SeQC: RNA-
691 seq metrics for quality control and process optimization. Bioinformatics. 2012;28(11):1530-2. doi:
692 10.1093/bioinformatics/bts196. PubMed PMID: 22539670; PubMed Central PMCID: PMC3356847.
693 54. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput
694 sequencing data. Bioinformatics. 2015;31(2):166-9. doi: 10.1093/bioinformatics/btu638. PubMed
695 PMID: 25260700; PubMed Central PMCID: PMC4287950.
696 55. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential
697 expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40. doi:
698 10.1093/bioinformatics/btp616. PubMed PMID: 19910308; PubMed Central PMCID: PMC2796818.
699 56. Abbas AR, Baldwin D, Ma Y, Ouyang W, Gurney A, Martin F, et al. Immune response in silico
700 (IRIS): immune-specific genes identified from a compendium of microarray expression data. Genes
701 Immun. 2005;6(4):319-31. doi: 10.1038/sj.gene.6364173. PubMed PMID: 15789058.
702 57. Joice R, Narasimhan V, Montgomery J, Sidhu AB, Oh K, Meyer E, et al. Inferring
703 developmental stage composition from gene expression in human malaria. PLoS Comp Biol.
704 2013;9(12):e1003392. doi: 10.1371/journal.pcbi.1003392. PubMed PMID: 24348235; PubMed
705 Central PMCID: PMC3861035.
706 58. Benjamini Y, Hochberg Y. Controlling False Discovery Rate: A Practical and Powerful
707 Approach to Multiple Testing. J R Stat Soc B. 1995;57(1):289-300.
708 59. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene
709 ontology terms. PLoS One. 2011;6(7):e21800. doi: 10.1371/journal.pone.0021800. PubMed PMID:
710 21789182; PubMed Central PMCID: PMC3138752.
711 60. Hu Y, Yan C, Hsu CH, Chen QR, Niu K, Komatsoulis GA, et al. OmicCircos: A Simple-to-Use R
712 Package for the Circular Visualization of Multidimensional Omics Data. Cancer Informatics.

713 2014;13:13-20. doi: 10.4137/CIN.S13495. PubMed PMID: 24526832; PubMed Central PMCID:

714 PMC3921174.

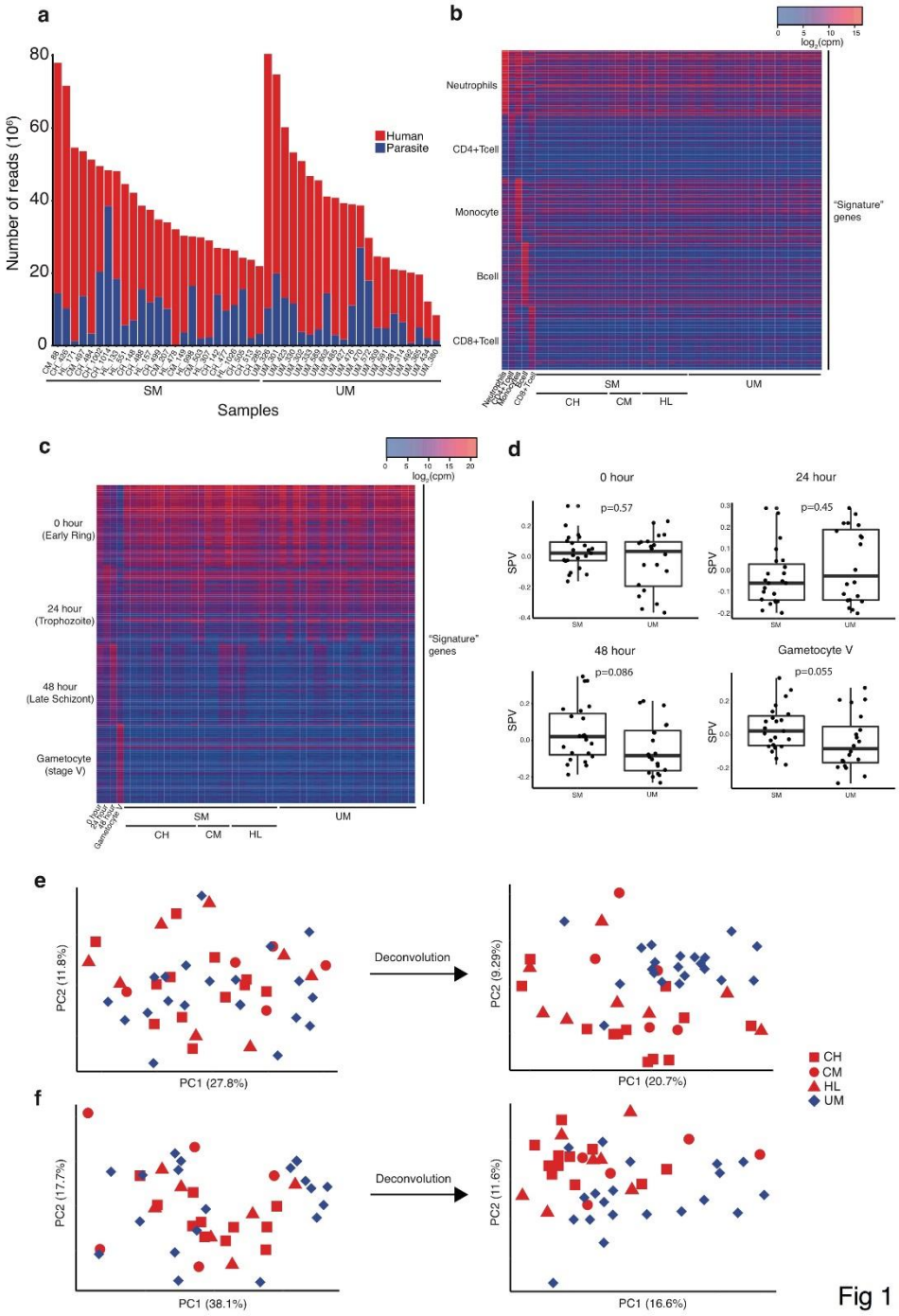
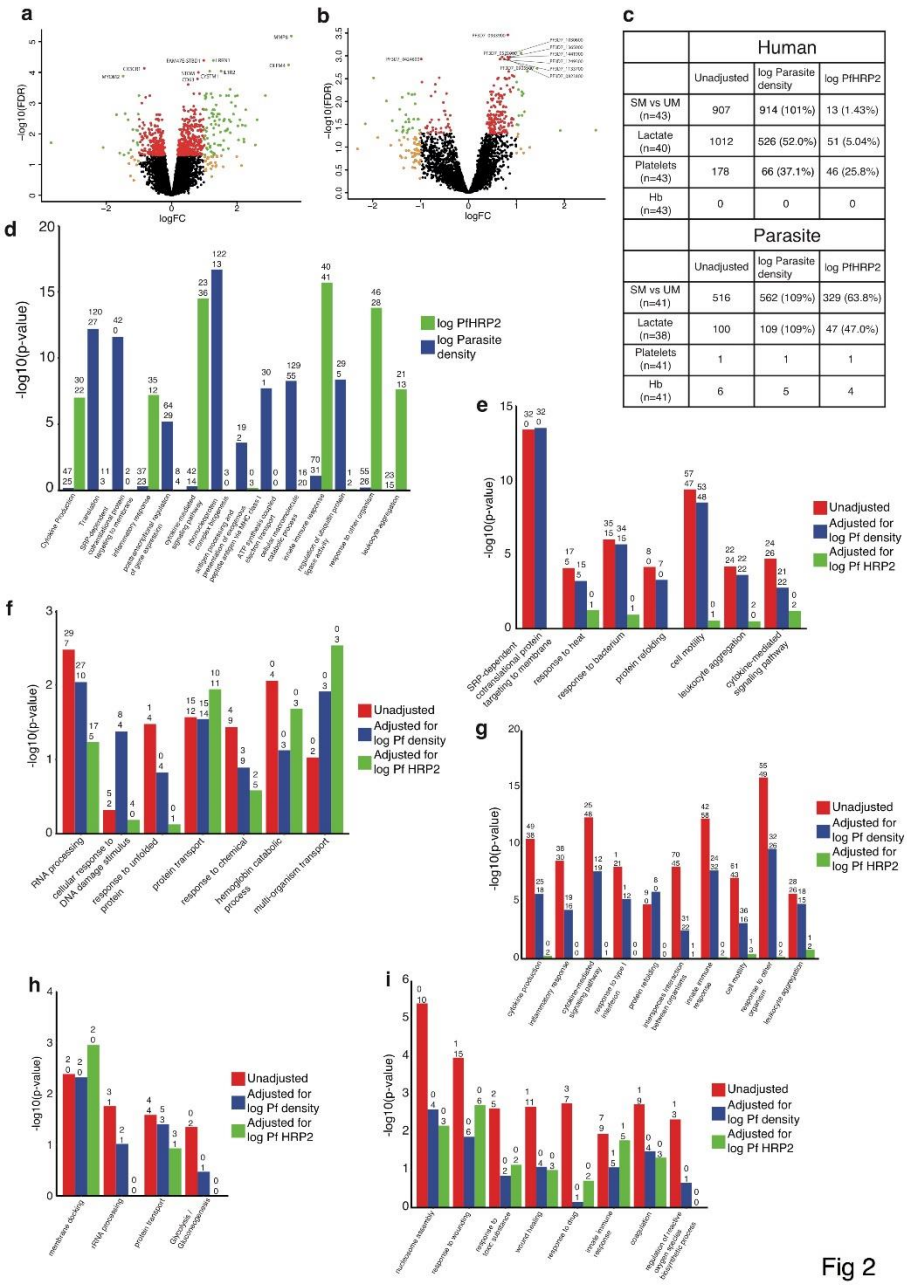


Fig 1

Figure 1. Whole blood dual RNA-sequencing and deconvolution. (**a**) Uniquely mapped reads from human (red) and *P. falciparum* (blue) from subjects with severe (SM, n=25) and uncomplicated malaria (UM, n=21). (**b,c**) Heatmaps showing signature gene expression for different leukocyte (**b**) and parasite developmental stage (**c**) populations (rows) and their relative intensity in individual subjects with SM, including different SM phenotypes (CH, cerebral malaria plus hyperlactatemia; CM, cerebral malaria; HL, hyperlactatemia), and UM (columns). (**d**) Surrogate proportion variables for parasite developmental stages compared between SM and UM using the Mann-Whitney test (bold line, box and whiskers indicate median, interquartile range and 1.5-times interquartile range respectively). (**e,f**) Principal component plots showing the effect of deconvolution on the segregation of subjects with UM and SM, adjusting human (**e**) and parasite (**f**) gene expression for differences in proportions of leukocytes or parasite developmental stages respectively. Analyses of human gene expression (**b**,**e**): SM, n=25; UM, n=21. Analyses of parasite gene expression (**c**,**d**, **f**): SM, n=23; UM, n=20.



	Human		
	Unadjusted	log Parasite density	log PfHRP2
SM vs UM (n=43)	907	914 (101%)	13 (1.43%)
Lactate (n=40)	1012	526 (52.0%)	51 (5.04%)
Platelets (n=43)	178	66 (37.1%)	46 (25.8%)
Hb (n=43)	0	0	0
	Parasite		
	Unadjusted	log Parasite density	log PfHRP2
SM vs UM (n=41)	516	562 (109%)	329 (63.8%)
Lactate (n=38)	100	109 (109%)	47 (47.0%)
Platelets (n=41)	1	1	1
Hb (n=41)	6	5	4

Fig 2

Figure 2. Association of gene expression with severity features and dependency on parasite load. (**a, b**) Volcano plots showing extent and significance of up- or down- regulation of human (**a**) or *P. falciparum* (**b**) gene expression in SM compared with UM (red and green, P <0.05 after Benjamini-Hochberg adjustment for false discovery rate (FDR); orange and green, absolute log_2-fold change (FC) in expression > 1; the 10 most significant genes are annotated; human comparison SM n=25, UM=21; parasite comparison SM n=23, UM=20). (**c**) Number of human and parasite genes associated with severity category and laboratory markers of severity before and after adjustment for parasite load expressed as either log circulating parasite density or log PfHRP2 concentration. Only subjects with complete data for every parameter are included. (**d-i**) Most significantly enriched, non-redundant, gene ontology terms for genes significantly associated with log parasite density (**d**) and log PfHRP2 (**e**), and the effect of adjustment for these measures on genes significantly associated with severity (**f-i**) (numbers above bars indicate the number of up-regulated/positively-associated and down-regulated/negatively-associated genes within each category). (**e,f**) Human (**e**) and *P. falciparum* (**f**) genes significantly differentially expressed in SM vs UM. (**g,h**) Human (**g**) and *P. falciparum* (**h**) genes significantly correlated with blood lactate concentration. (**i**) Human genes significantly correlated with platelet count.

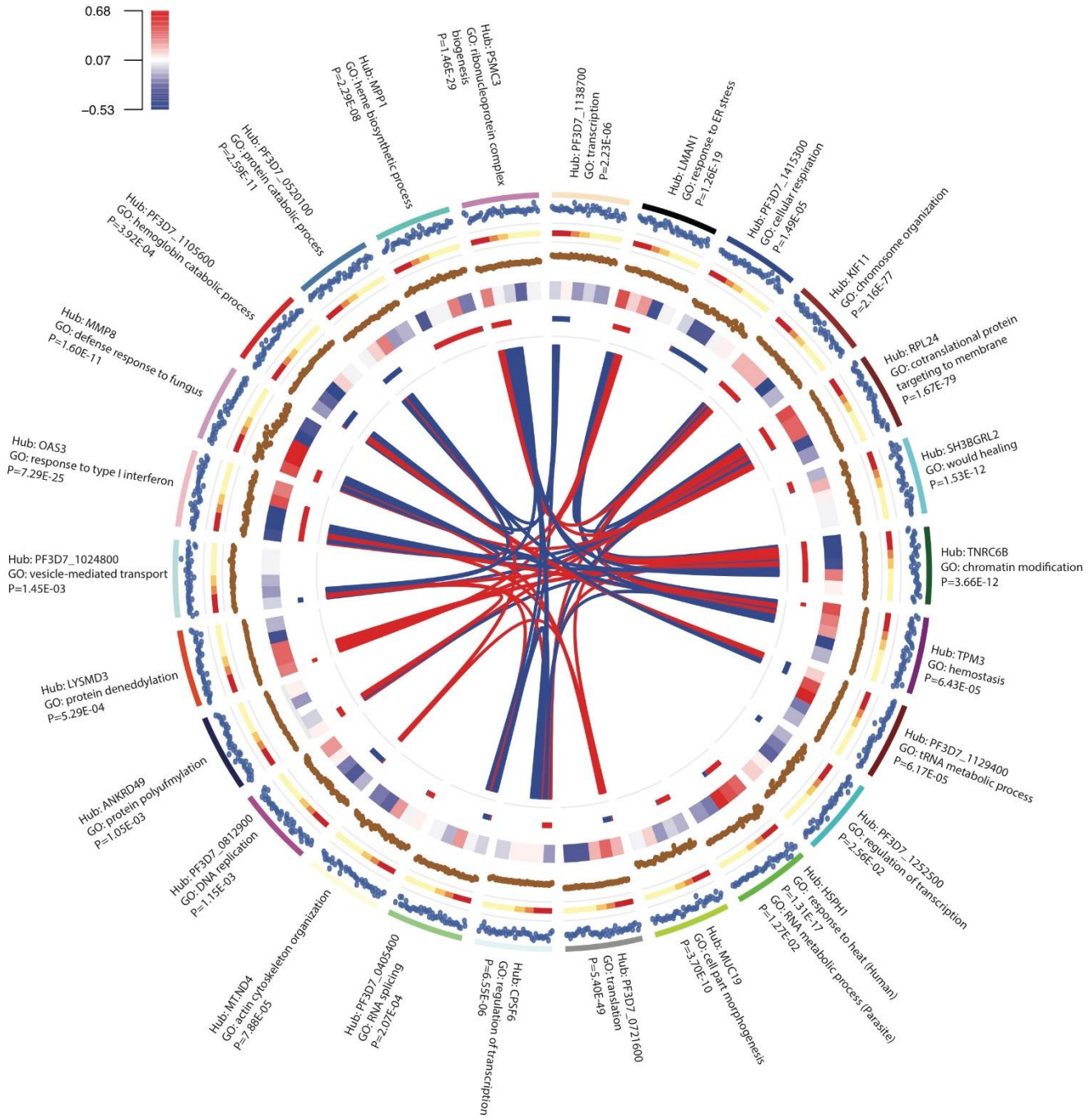
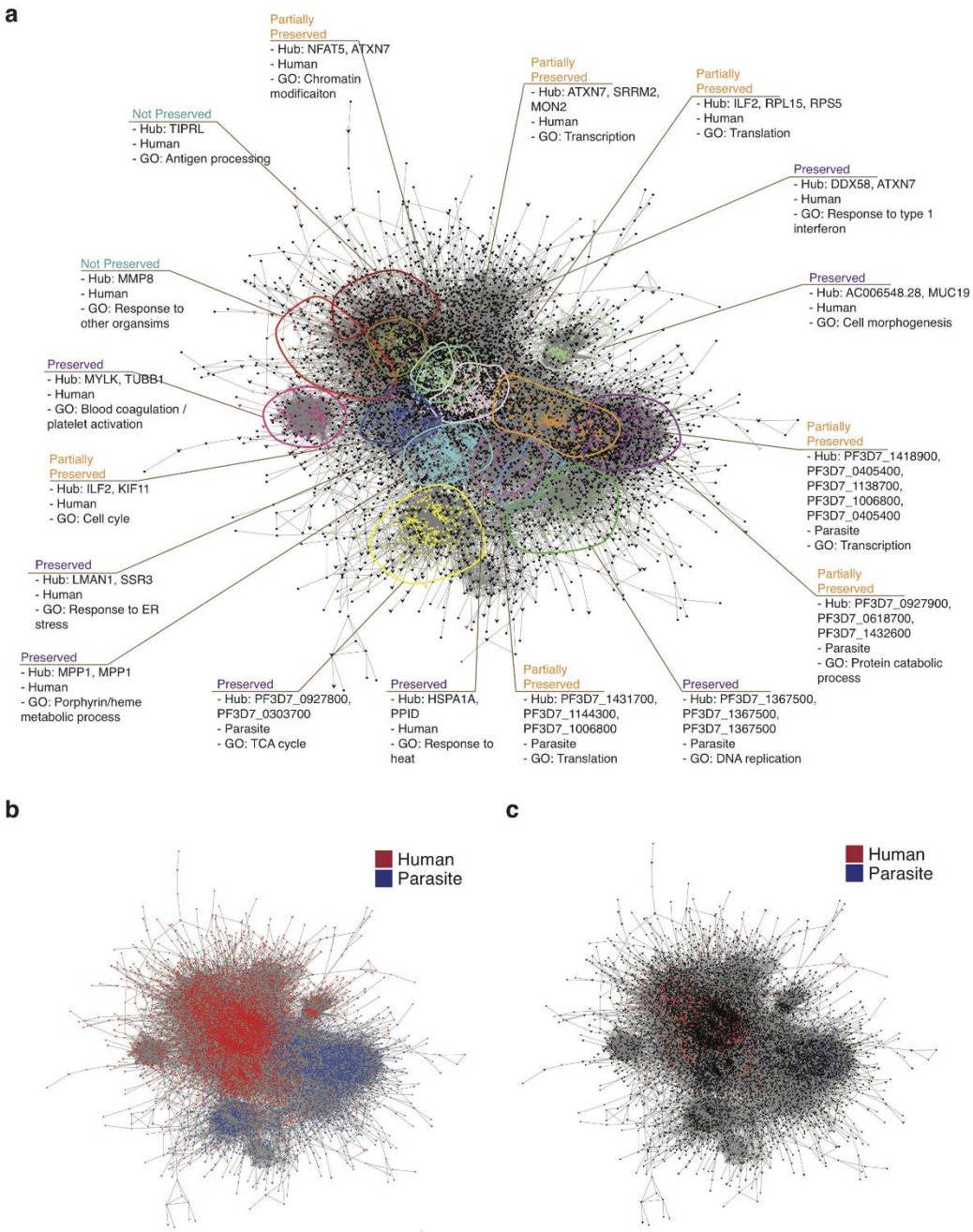


Figure 3. Interspecies gene expression modules and their associations with severity. Circos plot showing gene expression modules obtained from whole genome correlation network analysis using expression of all human and parasite genes from each subject (SM, n=22; UM, n=19) as the input. From outside to inside: labels, hub gene and most enriched GO term (with enrichment P-value) for each module; track 1, module eigengene value for each subject; track 2, clinical phenotype (Red=CH, Orange=CM, Green=HL, Yellow=UM); track 3, hub gene expression (log CPM) for each subject; track 4, heatmap for correlation with laboratory measurements (clockwise: log parasite density, log PfHRP2, lactate, platelets, haemoglobin; colour intensity represents Pearson correlation coefficient as shown in legend); track 5, module size and composition (length proportional to number of genes in module; red, human genes; blue, parasite genes); polygons connect modules with significant (FDR P<0.01) Pearson correlation between eigengene values (width proportional to $-\log_{10}$ FDR P-value; red=positive correlation, blue=negative correlation)



745 **Figure 4. Severity-associated differential co-expression within the interspecies gene expression network.** (**a-c**) Cytoscape visualisation of
746 merged co-expression networks derived separately from SM (n=22) and UM (n=19). Networks were merged such that genes found in both
747 sub-networks (represented as arrow-shaped, larger-sized nodes) are connected to genes found in only one sub-network (represented as
748 circular-shaped and smaller-sized nodes). (**a**) Genes and gene clusters are coloured and annotated by module, species, most enriched gene
749 ontology terms, and conservation between sub-networks (preserved, module pairs from SM and UM sub-networks display highly
750 significant overlap with each other and much less significant overlap with other modules; partially preserved, module clusters display
751 significant overlaps with two or more modules in the other subnetwork; unique, gene clustering only found in one sub-network); genes in
752 black do not belong to any characterized module. (**b**) Identical network layout with genes coloured by species (red, human; blue, *P.*
753 *falciparum*). (**c**) Identical network layout with genes coloured by whether they are significantly differentially expressed in SM vs UM (red,
754 human; blue, *P. falciparum*; black, not differentially expressed).

Supplementary Material

Supplementary Table 1. Characteristics of study subjects (n=46)

	CM (n=5)	CH (n=12)	HL (n=8)	UM (n=21)	*P (F value)*
Age (years)	4.3 (4.2-4.8)	4.9 (3.6-5.7)	5.0 (3.8-8.3)	6.0 (4.0-9.0)	0.24 (1.47)
Male (%)	3 (60%)	5 (42%)	7 (88%)	13 (62%)	0.24
Parasitemia (%)	8.3 (5.3-9.0)[4]	12.6 (9.4-19.0)	9.6 (1.8-12.2)	5.1 (3.8-7.0)	0.02 (3.67)
Parasites (x10^5 /uL)	2.3 (1.7-3.1) [3]	3.5 (2.7-8.4) [11]	2.8 (0.7-5.0)	2.3 (1.6-3.2)	0.02 (3.55)
Clones	2 (1.5-2.5) [4]	2 (1-2) [9]	1 (1-2) [5]	2 (1-2) [15]	0.68 (0.51)
***Pf*HRP2 (ng/mL)**	202 (93-528) [4]	763 (374-1750)	470 (164-2214)	163 (128-227)	0.001 (6.20)
Duration of illness (days)	2.0 (1.7-3.0)	2.0 (2.0-2.5)	2.0 (2.0-3.5)	2.7 (2.0-3.0)	0.57 (0.68)
Hb (g/dL)	9.7 (7.4-10.4)	9.3 (7.8-11.5) [11]	9.1 (7.4-11.0)	10.8 (9.9-12.1)	0.10 (2.24)
WBC (x10^9/L)	9.8 (8.2-12.9) [4]	8.8 (6.4-9.4) [11]	15.3 (7.9-16.8) [7]	9.5 (7.7-11.8)	0.30 (1.27)
Platelets (x10^9/L)	41 (40-82) [4]	36 (23-65) [11]	59 (33-132)	122 (96-132)	0.04 (3.15)
Lymphocyte (%)	29.8 (20.6-37.3) [4]	37.8 (29.9-49.9)[11]	22.3 (14.7-37.3)	23.9 (16.0-33.5)[20]	0.05 (2.78)
Neutrophil (%)	55.1 (49.0-69.6) [4]	48.3 (39.6-56.2) [10]	61.5 (55.6-74.9) [7]	68.0 (59.9-79.6)[20]	0.01 (4.25)
Monocyte (%)	7.1 (6.0-7.7) [4]	7.8 (6.8-8.6) [10]	6.6 (5.1-7.8) [7]	6.7 (4.8-7.3)[20]	0.38 (1.05)

CM, cerebral malaria; CH, cerebral malaria plus hyperlactatemia; HL, hyperlactatemia (CM, CH, and HL are all subgroups of severe malaria, SM); UM, uncomplicated malaria. Data are median (IQR), superscripts indicate the number of subjects with data for each variable if less than the total; *P (F)* for ANOVA comparing all groups (degrees of freedom =3) except for sex where *P* is for Fisher's exact test.

760 **Supplementary Table 2. Human genes differentially expressed between severe malaria**
761 **phenotypes and uncomplicated malaria in unadjusted and parasite load-adjusted analyses.**
762
763 **Supplementary Table 3. *P. falciparum* genes differentially expressed between severe malaria**
764 **phenotypes and uncomplicated malaria.**
765
766 **Supplementary Table 4. Human genes significantly correlated with parasite load measurements**
767 **and laboratory parameters.**
768
769 **Supplementary Table 5. *P.falciparum* genes significantly correlated with parasite load and**
770 **laboratory parameters.**
771
772 **Supplementary Table 6. Gene ontology terms associated with human differentially expressed or**
773 **significantly correlated genes in unadjusted and parasite load-adjusted analyses.**
774
775 **Supplementary Table 7. Predicted upstream regulators associated with human differentially**
776 **expressed or significantly correlated genes in unadjusted and parasite load-adjusted analyses.**
777
778 **Supplementary Table 8. Gene ontology terms associated with parasite differentially expressed or**
779 **significantly correlated genes in unadjusted and parasite load-adjusted analyses.**
780
781 **Supplementary Table 9. Summary of modules obtained from combined whole genome correlation**
782 **network.**
783
784

785 **Supplementary Table 10. Univariate and multivariate associations of module eigengene values and**

786 **parasite load with severity.**

787

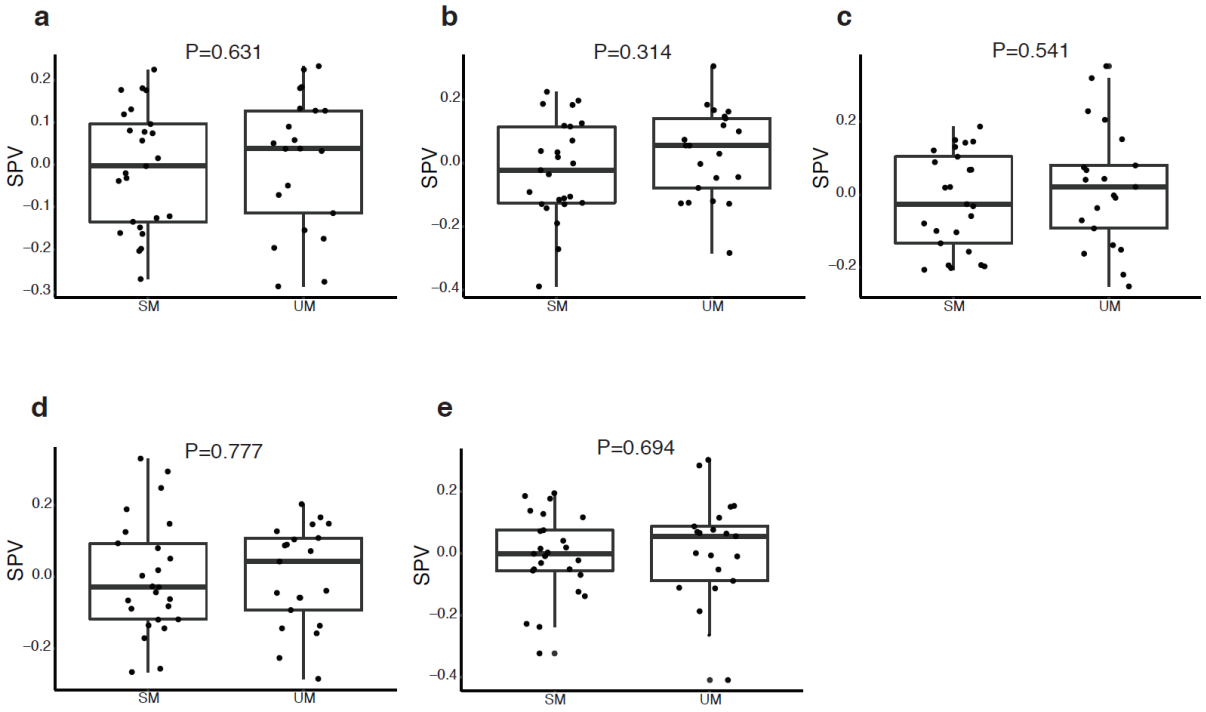
	Univariate log odds	Univariate P-value	Multivariate log odds	Multivariate P-value
MMP8	12.4	0.00130	**66.1**	**0.0152**
LYSMD3	9.49	0.00353	**28.6**	**0.0178**
PF3D7_1129400	10.0	0.00388		
RPL24	7.77	0.00463		
OAS3	-6.73	0.00943	**47.7**	**0.0261**
HSPH1	9.74	0.0117		
TNRC6B	-6.44	0.0136		
KIF11	6.15	0.0179		
PF3D7_1415300	-8.77	0.0184		
PF3D7_1252500	-5.50	0.0216		
PF3D7_1105600	-5.91	0.0218		
LMAN1	5.63	0.0255		
PF3D7_0721600	5.33	0.0283		
PF3D7_0520100	4.99	0.0307		
TPM3	4.44	0.0560		
ANKRD49	3.43	0.133		
MT.ND4	2.48	0.251		
PF3D7_0405400	2.18	0.289		
MUC19	-2.18	0.315		
PF3D7_0812900	-1.85	0.367		
SH3BGRL2	-1.41	0.491		
PSMC3	1.17	0.568		
PF3D7_1024800	-1.08	0.596		
PF3D7_1138700	0.938	0.643		
CPSF6	0.647402239	0.747		
MPP1	0.02183521	0.991		
Log parasite density	1.17	0.0339		
Log PfHRP2	1.4031	0.00440		

788 Log odds are per unit change in the variable, calculated using logistic regression. The multivariate
789 model was derived by backwards selection from all variables with univariate P<0.01.

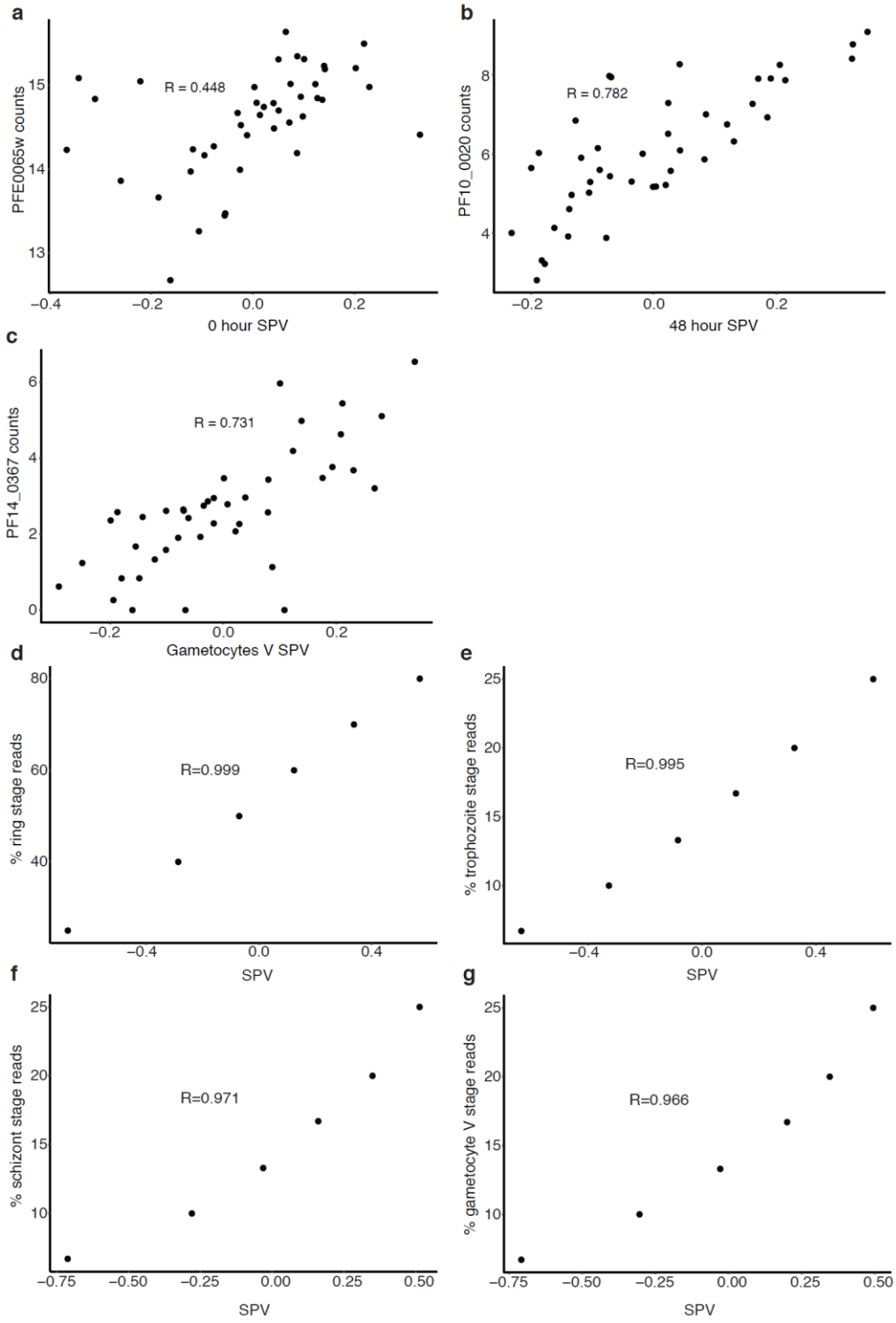
790 **Supplementary Table 11. Summary and overlap of whole genome correlation sub-networks for**

791 **severe and uncomplicated malaria.**

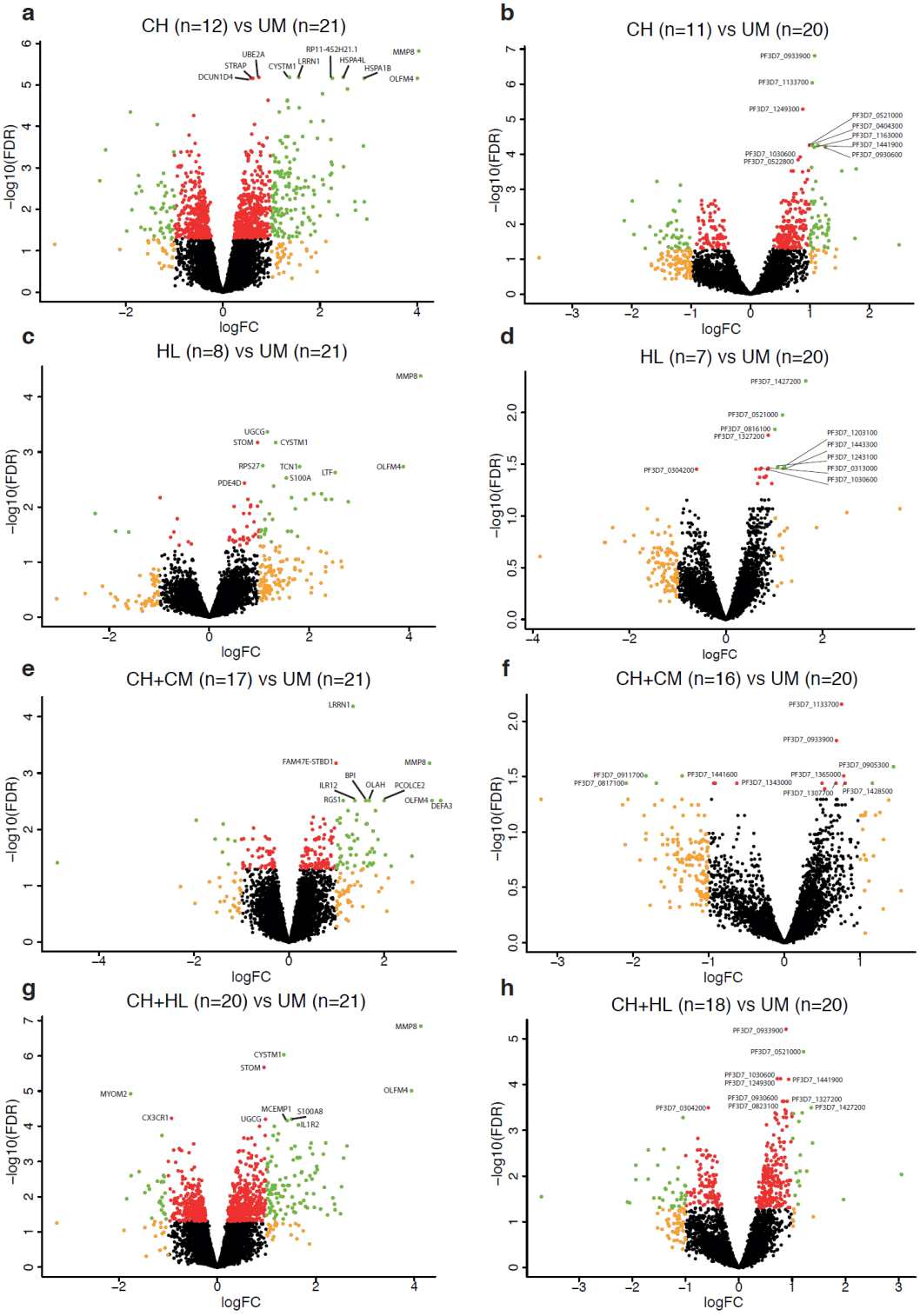
792



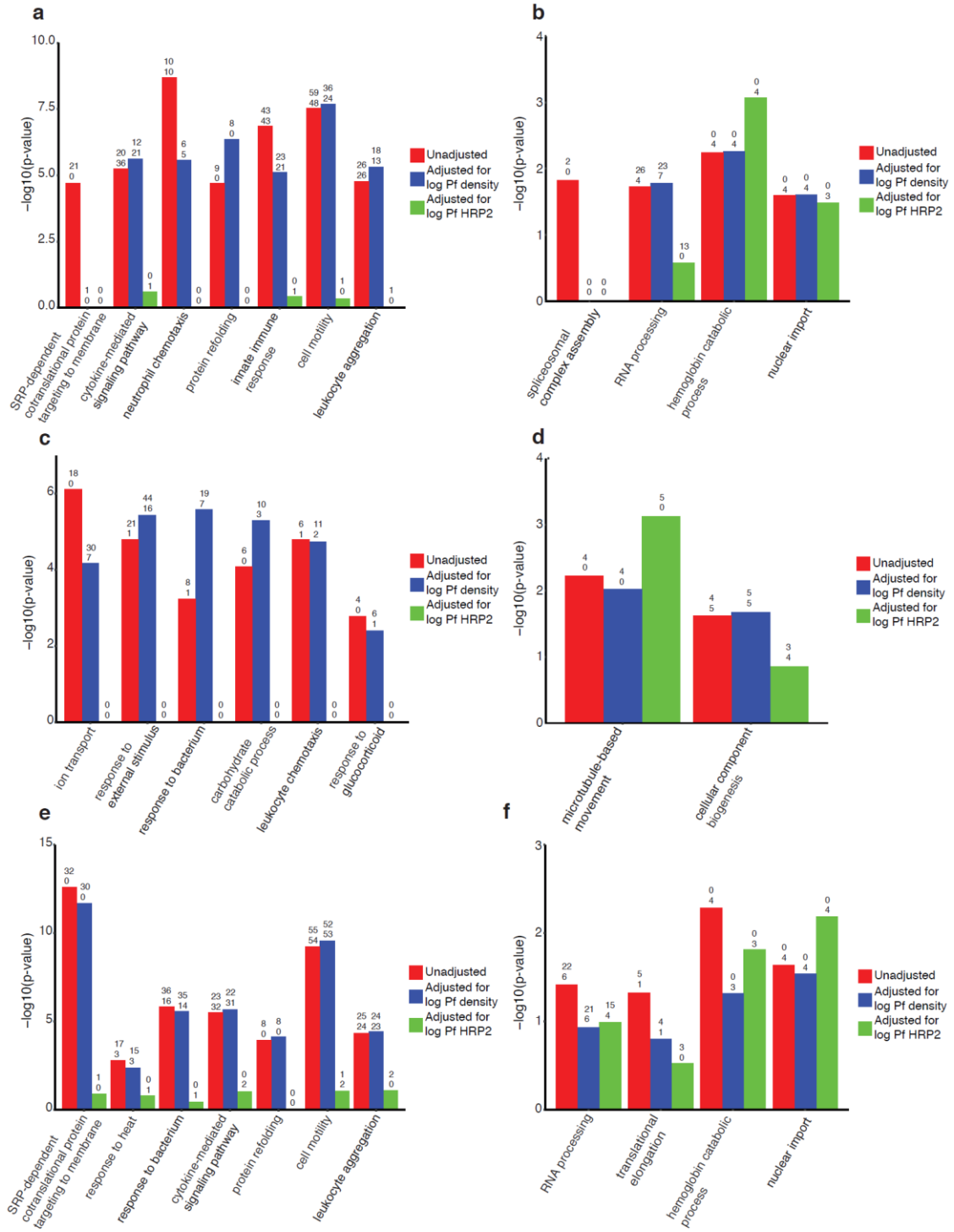
Supplementary Fig 1. Estimates of the relative proportions of leukocyte subpopulations in subjects with severe and uncomplicated malaria. (a-e) Surrogate proportion variables compared by severity category for neutrophils (**a**), monocytes (**b**), CD4+ T-lymphocytes (**c**), CD8+ T-lymphocytes (**d**), and B-lymphocytes (**e**) using the Mann-Whitney test (UM, n=21; SM, n=25; bold line, box and whiskers indicate median, interquartile range and 1.5-times interquartile range respectively).



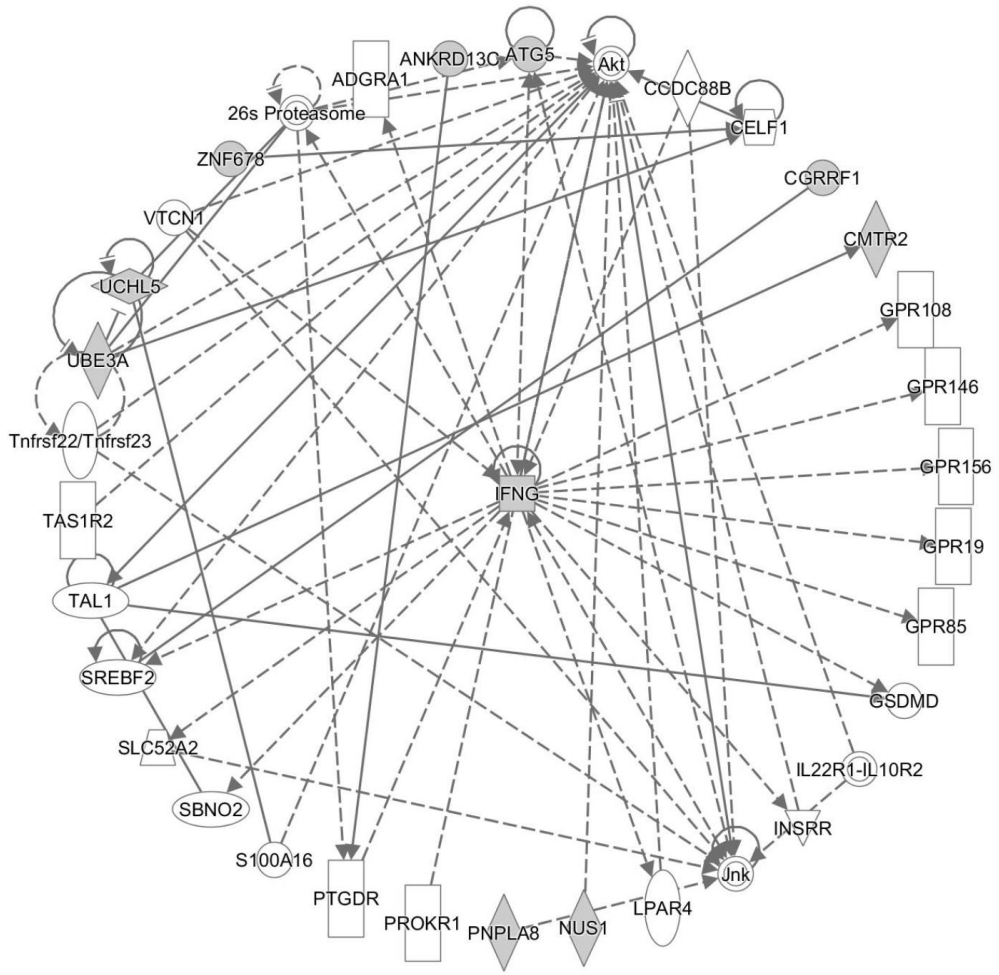
Supplementary Fig 2. Validation of the gene-signature approach to estimate parasite developmental stage proportions. (a-c) Correlation of surrogate proportion variables (SPV) with read counts for putative "stage-specific" marker genes (see Methods): (**a**) 0hr SPV vs. early asexual stage marker gene PFE0065w; (**b**) 48hr SPV vs. late asexual stage marker gene PF10_0020; (**c**) Gametocyte V SPV vs. developing gametocyte marker gene PF14_0367 (R for Pearson correlation, n=43). (**d-g**) Correlation of SPVs with actual proportion of reads derived from each parasite developmental stage in synthetic mixtures of varying proportions of stage-specific RNA-seq reads from early ring-stage (0 hour, **d**), trophozoite (24 hour, **e**), late schizont (48 hour, **f**) and mature gametocyte (Stage V, **g**). R for Pearson correlation, n=43.



Supplementary Fig 3. Differential gene expression between severe malaria phenotypes and uncomplicated malaria. Volcano plots showing extent and significance of up- or down- regulation of human (left hand column) or *P. falciparum* (right hand column) gene expression in comparisons between specific phenotypes of SM vs UM (red and green, P <0.05 after Benjamini-Hochberg adjustment for false discovery rate (FDR); orange and green, absolute $\log_2$-fold change (FC) in expression > 1; the 10 most significant genes are annotated).



Supplementary Fig 4. Gene ontology terms associated with genes differentially expressed between severe malaria phenotypes and uncomplicated malaria, unadjusted or adjusted for parasite load. Most significantly enriched, non-redundant, gene ontology terms for genes significantly differentially expressed (DEGs) between SM phenotypes and UM (numbers above each bar indicate the number of up-regulated and down-regulated genes within each category). Comparisons are only shown if they include multiple significantly enriched GO terms. (**a**) CH (n=11) vs UM (n=21) human DEGS, (**b**) CH (n=11) vs UM (n=20) *P. falciparum* DEGs, (**c**) HL (n=8) vs UM (n=21) human DEGS, (**d**) CH+CM (n=14) vs UM (n=20) *P. falciparum* DEGs, (**e**) CH+HL (n=19) vs UM (n=21) human DEGS, (**f**) CH+HL (n=18) vs UM (n=20) *P. falciparum* DEGs.



819

820 **Supplementary Fig 5. Top functional network for the small LYSMD3 module.** Functional networks were identified in Ingenuity Pathway

821 Analysis software and the top scoring network is portrayed in radial layout which places the most interconnected gene at the centre.

822 Genes within the module are shaded.

823

824 **Supplementary Dataset 1. Subject-level clinical and laboratory data**