






Host and Parasite Transcriptomic Changes upon Successive *Plasmodium falciparum* Infections in Early Childhood

 Katie R. Bradwell,^a Drissa Coulibaly,^b Abdoulaye K. Koné,^b  Matthew B. Laurens,^c Ahmadou Dembélé,^b Youssouf Tolo,^b Karim Traoré,^b Amadou Niangaly,^b Andrea A. Berry,^c Bourema Kouriba,^b Christopher V. Plowe,^d Ogobara K. Doumbo,^{b†} Kirsten E. Lyke,^c Shannon Takala-Harrison,^c Mahamadou A. Thera,^b  Mark A. Travassos,^c David Serre^a

^aInstitute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland, USA

^bMalaria Research and Training Center, University of Science, Techniques and Technologies, Bamako, Mali

^cMalaria Research Program, Center for Vaccine Development and Global Health, University of Maryland School of Medicine, Baltimore, Maryland, USA

^dDuke Global Health Institute, Duke University, Durham, North Carolina, USA

ABSTRACT Children are highly susceptible to clinical malaria, and in regions where malaria is endemic, their immune systems must face successive encounters with *Plasmodium falciparum* parasites before they develop immunity, first against severe disease and later against uncomplicated malaria. Understanding cellular and molecular interactions between host and parasites during an infection could provide insights into the processes underlying this gradual acquisition of immunity, as well as to how parasites adapt to infect hosts that are successively more malaria experienced. Here, we describe methods to analyze the host and parasite gene expression profiles generated simultaneously from blood samples collected from five consecutive symptomatic *P. falciparum* infections in three Malian children. We show that the data generated enable statistical assessment of the proportions of (i) each white blood cell subset and (ii) the parasite developmental stages, as well as investigations of host-parasite gene coexpression. We also use the sequences generated to analyze allelic variations in transcribed regions and determine the complexity of each infection. While limited by the modest sample size, our analyses suggest that host gene expression profiles primarily clustered by individual, while the parasite gene expression profiles seemed to differentiate early from late infections. Overall, this study provides a solid framework to examine the mechanisms underlying acquisition of immunity to malaria infections using whole-blood transcriptome sequencing (RNA-seq).

IMPORTANCE We show that dual RNA-seq from patient blood samples allows characterization of host/parasite interactions during malaria infections and can provide a solid framework to study the acquisition of antimalarial immunity, as well as the adaptations of *P. falciparum* to malaria-experienced hosts.

KEYWORDS malaria, transcriptomics

Despite tremendous progress in the last decades, malaria still has devastating consequences throughout Africa, where *Plasmodium falciparum* causes more than 200 million malaria cases and close to half a million deaths every year, the majority of them children (1). In Mali, malaria remains a leading cause of death in children under 5 years of age (2). In Bandiagara, a town of approximately 14,000 inhabitants in central Mali, malaria is highly seasonal with a transmission that peaks in September (3) and each child typically experiences one or two clinical episodes of malaria every year (4).

With repeated exposures to malaria parasites, children living in high-transmission settings gradually acquire immunity against the disease, first against severe malaria manifestations, then against milder symptoms that characterize uncomplicated malaria,

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

Editor Paola Flórez de Sessions, Oxford Nanopore Technologies

Copyright © 2020 Bradwell et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Mark A. Travassos, mtravass@som.umaryland.edu, or David Serre, dserre@som.umaryland.edu.

† Deceased.

Received 13 February 2020

Accepted 13 June 2020

Published 7 July 2020

until they eventually develop asymptomatic infections (5) with increased parasite clearance (6, 7). A better understanding of the processes accompanying the acquisition of immunity could shed light on the mechanisms underlying disease resistance and parasite tolerance and could guide more effective antimalarial treatments. Antidisease immunity is thought to be partially mediated by recognition of multiple antigens, and longitudinal studies following children from infancy to adulthood have demonstrated recognition of an increasing number of antigens and antigen variants over time (8). Innate immunity may also play a role in protecting against malaria and in modulating the adaptive immune response afforded by this repertoire (9). However, adaptive immunity appears to be relatively short-lived, as evidenced by loss of immunity upon discontinued exposure (10). Aside from “strain-specific” acquired immunity, the concept of a “strain-transcending” immunity that may be mediated by host age has been proposed (11), although it is important to note that severe malaria does occur in older children in low-transmission areas (12). Parasite tolerance, possibly acquired via immunoregulatory mechanisms (13), remains a poorly characterized yet intriguing avenue of research. Study of immune cell repertoires have shown that complicated malaria cases present different CD4⁺ T-cell phenotypes than those seen in uncomplicated or asymptotically infected individuals (14). Platelets, which act as first responders to infection, secrete platelet factor 4 (which lyses the parasitic vacuole), and present antigens in the context of major histocompatibility complex (MHC) class I, have also been shown to exhibit clear differences in counts between patients with complicated malaria (greater thrombocytopenia) and those with uncomplicated malaria (7).

Gene expression analyses have the potential to complement immunological studies and could reveal molecular processes underlying the acquisition of immunity to malaria. Transcriptome sequencing (RNA-seq) studies have thus reported gene expression differences between malaria-naive and -experienced individuals (15), as well as between severe and uncomplicated malaria cases (16). In addition, microarray studies have shown that recent and heavy exposure to malaria is associated with a loss of proinflammatory cytokine production (17), and higher levels of the anti-inflammatory cytokine interleukin 10 (IL-10) (18).

To date, most malaria gene expression studies focused either on the host response to infection or on parasite gene expression and its association with disease phenotypes, with the exception of a few studies that characterized general interactions between infecting malaria parasites and their hosts (16, 19). Simultaneous characterization of host and parasite gene expression profiles, sometimes referred to as dual RNA-seq, could provide novel perspectives on the interactions between host and pathogen during an infection (20) and address an important but understudied aspect of immunity. In particular, study of the dynamic changes occurring over successive infections could identify molecular pathways involved in host immunity acquisition and mechanisms used by parasites to overcome the immunity of more experienced hosts. Here, we describe the parasite and host gene expression profiles generated from blood samples collected in the context of a malaria incidence study (4) from three young Malian children during five successive clinical malaria episodes. We show that RNA-seq provides robust characterization of both organisms’ transcriptomes without requiring sample processing or culture. We present statistical analyses of the temporal changes in gene expression and show that the gene expression profiles of both organisms cluster differently over time. We also demonstrate how RNA-seq data can support (i) gene expression deconvolution analyses to estimate the proportions of the different parasite stages and white blood cell (WBC) subsets in each sample, (ii) analysis of host-parasite gene coexpression, and (iii) robust genotyping to examine the complexity of each infection.

RESULTS

Changes in host and parasite gene expression over successive infections. We simultaneously analyzed the parasite and host gene expression profiles from three Malian children (aged 1 to 2 years old) enrolled in a longitudinal study of malaria

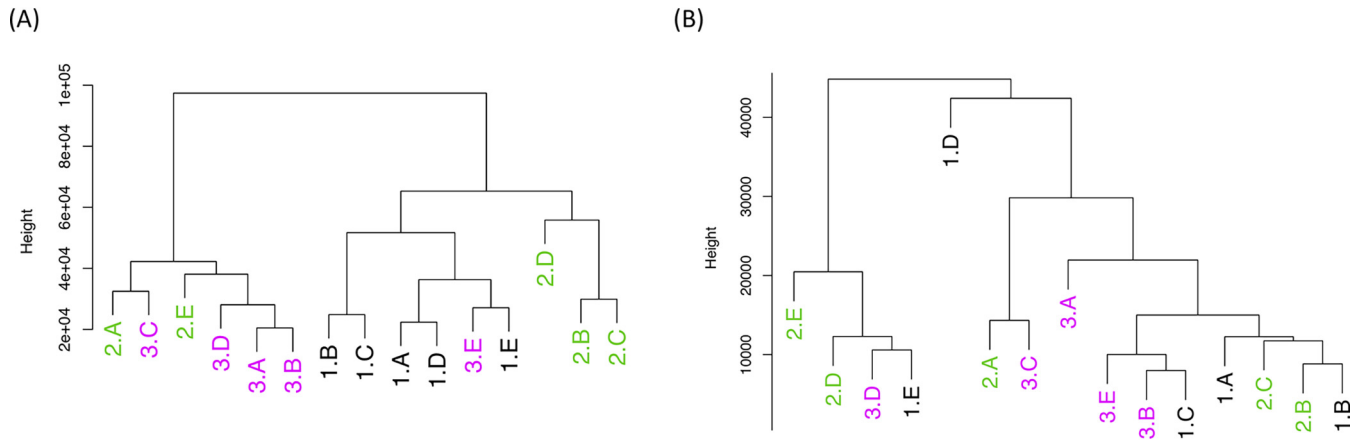


FIG 1 Unsupervised clustering of the 15 samples according to the host (A) and parasite (B) gene expression profiles. The colors and numbers (1 to 3) indicate which patient the sample is derived from. The letters distinguish the five symptomatic infections from each patient, with A representing the earliest infection and E the latest. Tree height refers to dissimilarities in terms of squared Euclidean distance between cluster means.

incidence in Bandiagara, Mali, between 2009 and 2014 (4). For each child, we extracted RNA from five blood samples collected during successive symptomatic *P. falciparum* infections for a total of 15 samples (for details, see Table S1 in the supplemental material). After ribosomal and globin RNA depletion, we prepared a stranded RNA-seq library and generated 34 to 67 million read pairs from each sample (Table S2).

To test if any of those infections contained more than one species of *Plasmodium* parasites (21), we mapped the reads to the genome sequences of different *Plasmodium* species infecting humans. In all samples, the reads mapping to *P. ovale*, *P. malariae*, and *P. vivax* represented less than 0.68% of all *Plasmodium* reads, suggesting that these blood samples were infected with only *P. falciparum*. Overall, 17 to 91% of the reads mapped to the human genome and 5 to 78% of the reads mapped to the *P. falciparum* genome (Table S2). After stringent quality filters, we obtained 5 to 24 and 1 to 12 million reads mapping to the human and *P. falciparum* genomes, respectively. The majority of reads mapped to annotated coding regions ($\sim 80\%$ for human and $\sim 97\%$ for *P. falciparum*) and provided sufficient information to analyze the expression levels of 8,896 host and 2,822 parasite genes (see Materials and Methods).

To assess how host and parasite gene expression profiles change over successive infections, we compared the transcriptomes of the 15 samples (3 children \times 5 successive infections). Unsupervised clustering revealed that host gene expression profiles tended to cluster each child's successive infections together (Fig. 1A), while the *P. falciparum* transcriptomes generated from the same successive infections tended to differentiate early from late infections, regardless of the individual (Fig. 1B). To further investigate this pattern, for each host and parasite gene, we tested whether the expression was influenced by the host and/or sequential infections (i.e., whether it was the first, second, third, fourth, or fifth infection) using a statistical framework that assessed whether the expression levels changed consistently over time. Consistent with the hierarchical clustering results, a greater number of host genes were differentially expressed according to the individual than the number of the infection (e.g., 4,581 versus 1,042 at a false discovery rate [FDR] of 0.2), while a greater number of parasite genes were differentially expressed according to the number of the infection rather than the individual (0 versus 68, FDR = 0.2) (Table 1 and Tables S3 and S4).

Several of the host genes whose expression changed the most (and consistently) over successive infections were involved in G-protein signaling, platelet aggregation, and immunoregulation (Fig. 2A and Table S3). To systematically examine whether some pathways were disproportionately represented among the genes differentially expressed according to the number of the infection ($n = 97$, FDR = 0.1), we performed enrichment analyses. PANTHER overrepresentation test (22) suggested that blood coagula-

TABLE 1 Number of host and parasite genes differentially expressed according to the patient and the number of the infection^a

Transcriptome	No. of genes tested	DE according to patient or infection no.	No. of DE genes at:	
			FDR = 0.2	FDR = 0.1
Host	8,896	Patient	4,581	2,876
		Infection no.	1,042	97
Parasite	2,822	Patient	0	0
		Infection no.	68	11

^aOnly genes expressed at more than 10 counts per million in more than six samples were tested (see Materials and Methods).

tion was also influenced by the number of the infection (P value = 1.63×10^{-5}). Gene set enrichment analysis (GSEA) for Reactome pathways (Fig. 3A) confirmed the roles of G-protein signaling and blood coagulation, as well as revealed enrichment in other pathways such as cytokine signaling. Table S5 shows the full results and information on the leading edge genes that drive the enrichment. Only a handful of parasite genes reached statistical significance (Fig. 2B and Table S4), including phospholipase A2, Alba 2, and glyceraldehyde-3-phosphate dehydrogenase, and no specific pathway was statistically enriched.

The host genes differentially expressed among the three children (Table S3) included CD36, also known as glycoprotein IV, a membrane protein present on the surface of many cell types that facilitates the binding and activation of platelets and monocytes (23) and is hypothesized to influence the host response to *P. falciparum* infection (7). Despite the large number of differentially expressed genes (2,876 genes at a FDR = 0.1), PANTHER analysis did not reveal any significant enrichment after multiple testing correction. GSEA results produced statistically significant results ($P < 0.05$) only for the child 1 versus child 2 comparison (Fig. 3B and Table S5), and leading edge analysis did not place CD36 in any significantly enriched pathways. Child 1, the oldest

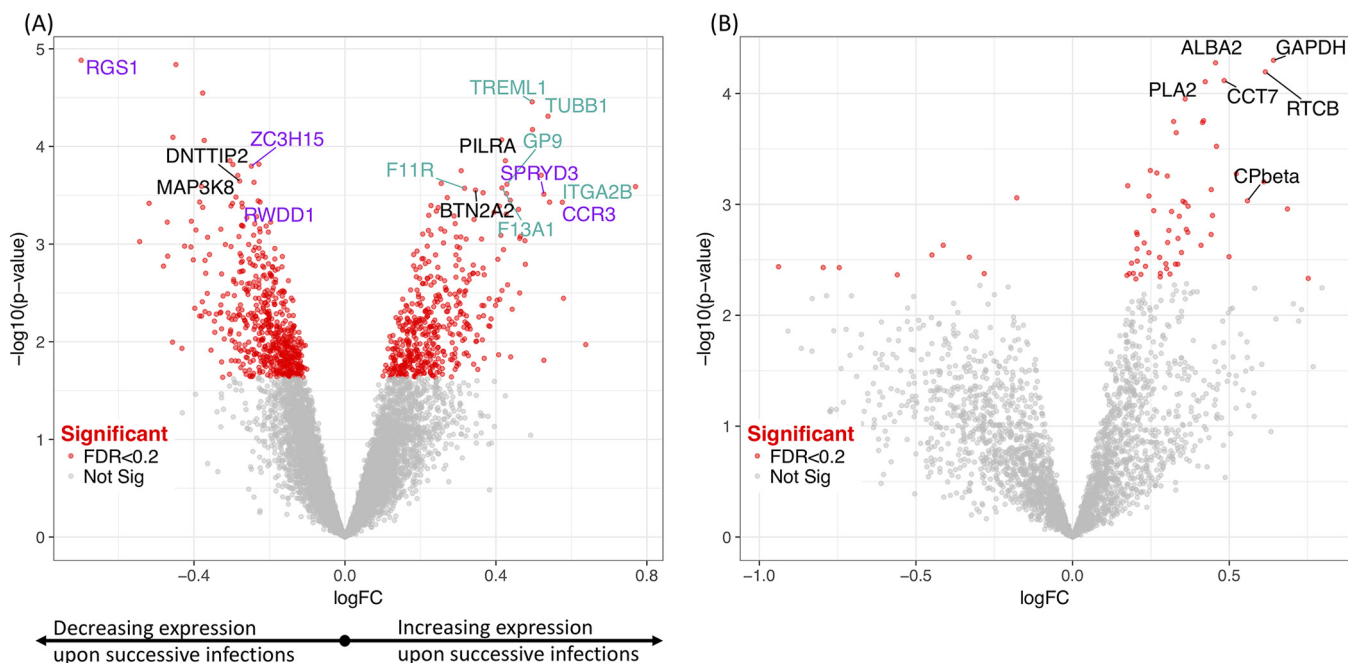


FIG 2 Volcano plot showing the results of the differential gene expression according to the number of successive infections for the host (A) and parasite (B) genes. Each dot represents one gene and is displayed according to the log fold change in expression (x axis) and the statistical significance of the association (y axis, in $-\log_{10}$ of the P value). Red dots indicate genes deemed to be differentially expressed (FDR = 0.2). Genes that increased in expression over the course of the five successive infections are shown by positive log fold change values, and those that decreased in expression are shown by negative log fold change values. Selected genes discussed in the text are labeled and, for the host, are color coded based on their functional annotation (immunoregulatory functions shown in black, platelet aggregation in turquoise, and G-protein signaling in purple).

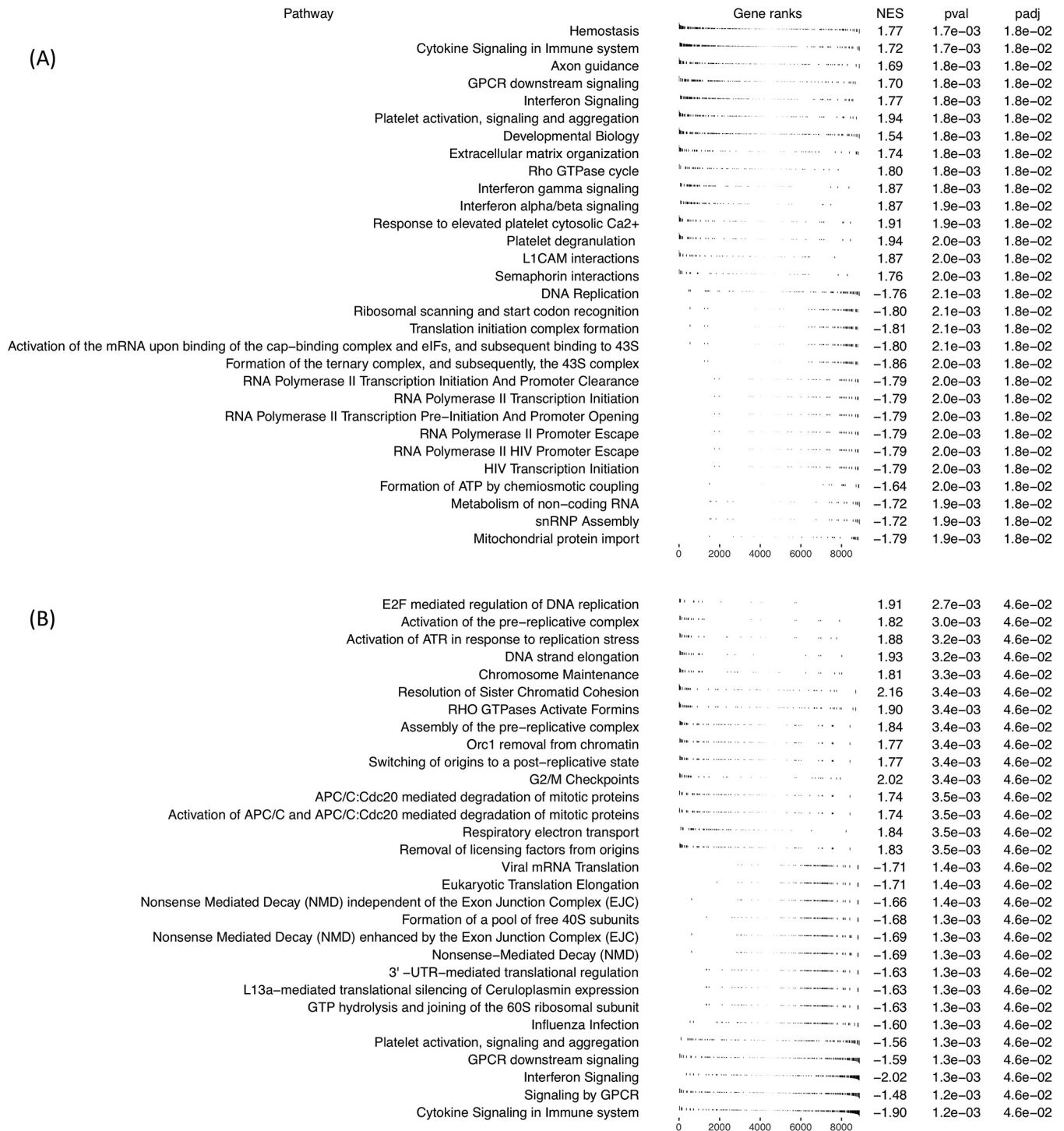


FIG 3 GSEA analysis of the human transcriptome by infection number (A) and patient 1 versus patient 2 (B). (A) The top 15 plots show the top 15 pathways upregulated over successive infection numbers, and the bottom 15 plots show the top 15 pathways downregulated over successive infection numbers. (B) The top 15 plots show the top 15 pathways upregulated in patient 2, and the bottom 15 plots show the top 15 pathways downregulated in patient 2.

of the three children, showed enrichment of platelet-related and cytokine-signaling related pathways compared to both of the other children (although child 1 versus child 3 did not reach significance), reflecting the findings of enrichment by infection number.

Coexpression of host and parasite genes. Joint characterization of host and parasite gene expression profiles from the same blood sample provides an opportunity to look for interactions, either directly between host and pathogen proteins, or indi-

rectly as one molecular pathway in one organism may regulate a separate process in the other organism. We searched for putative interactions by measuring the correlation between the expression levels of each pair of host gene-parasite gene across all 15 infections. We identified 2,690 pairs with a Spearman's coefficient of correlation $R^2 > 0.9$ (see, e.g., Fig. S1 in the supplemental material). This high extent of correlation observed between host and parasite gene expression was much greater than one would expect solely by chance ($P = 0.024$, based on 500 permutations), and indeed, only 709 gene pairs should display such high correlations by chance (corresponding to a FDR of 0.26, see Materials and Methods). Thus, despite the small sample size of the current study, our analyses demonstrate that dual RNA-seq can identify statistically significant host/pathogen correlations at the transcript level and could provide a framework to rigorously assess interactions occurring during an infection (though larger sample sizes would be needed to lower the false discovery rate and pinpoint biologically relevant interactions).

Gene expression deconvolution allows determination of the relative proportions of WBC subsets and parasite developmental stages. Host gene expression data generated from whole blood can be difficult to interpret as the samples contain a variable proportion of cell types, each with their own specific regulation, and gene expression differences between samples could simply reflect differences in cell composition. Similarly, parasite gene expression profiles will be influenced by the relative proportions of different parasite developmental stages. To overcome these limitations and determine the proportions of WBC subsets and parasite developmental stages in each sample, we used gene expression deconvolution analysis (24). First, we used transcriptome profiles from sorted WBCs (25–28), as well as *P. falciparum* developmental stage transcriptome profiles obtained from single-cell RNA-seq (29) to generate the gene expression signature profiles of each cell type and parasite stage. We then used these signature profiles to deconvolute the complex gene expression profiles generated from whole blood and statistically separate the transcriptional signal from each cell and parasite stage (Fig. 4).

Overall, the proportions of the different white blood cell subsets inferred from the RNA-seq data matched those expected in human whole blood (30), except for sample 3C, which displayed a low proportion of granulocytes and relatively high proportions of T cells, B cells, and myeloid dendritic cells. Interestingly, the proportion of NK cells seemed to decrease with the infection number ($P = 2.0 \times 10^{-3}$), though the small proportion of NK cells in each sample warrants caution. Similarly, the proportion of transcripts derived from myeloid dendritic cells and NK cells seemed to differ significantly among individuals (P values of 0.03 and 1.8×10^{-5} , respectively) (Table S6).

In contrast, the proportion from different parasite developmental stages did not seem to change between individuals ($P > 0.06$) or as a function of the number of infections ($P > 0.11$) (Table S7). Note that the small number of samples in the current study prevented us from correcting the differential expression analyses described above for these variations in composition, but larger studies could easily integrate this information to correct for differences among samples and distinguish whether the differential expression is caused by differences in cell composition or genuine differences in specific transcript regulation.

Complexity of infection and genotyping. In addition to the mixture of parasite stages, *Plasmodium* infections often simultaneously contain multiple, genetically distinct clones. Since *Plasmodium* parasites are haploid in the human host, identification of multiple alleles throughout the genome is indicative of a polyclonal infection. To evaluate whether RNA-seq data distinguishes monoclonal from polyclonal infections, we analyzed allelic variations, within each infection, at nucleotide sites highly sequenced ($>50\times$) using the sequences generated by RNA-seq. While most infections displayed a single allele at each transcribed position, allelic variation patterns in six infections were suggestive of the presence of two or more clones (Fig. 5). These observations were consistent with the *Fws* value (47), an estimate of polyclonality akin

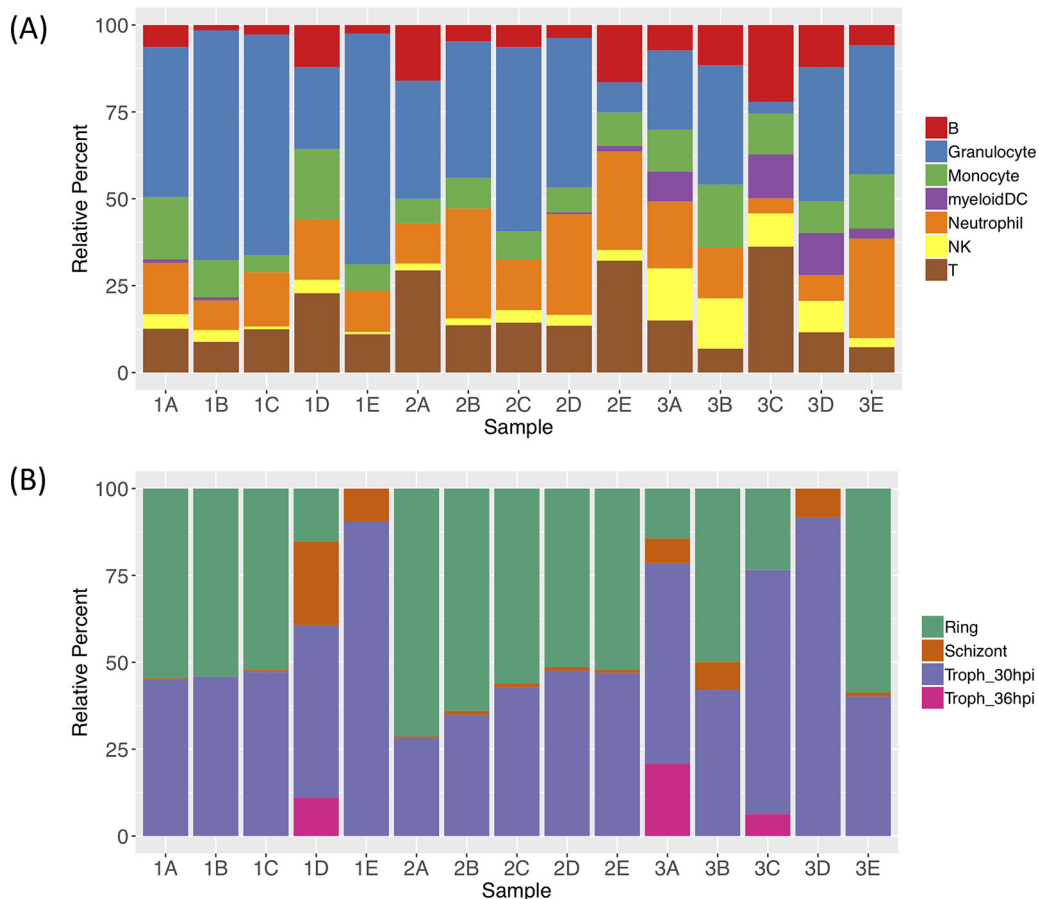


FIG 4 Gene expression deconvolution results. (A) Relative proportions of the different white blood cell subsets determined from the host transcriptomes. (B) Relative proportions of the different *P. falciparum* developmental stages determined from the parasite transcriptomes (hpi, hours postinfection).

to Wright’s inbreeding coefficient and calculated by comparing the heterozygosity within and between infections, determined from each infection: six samples displayed an *F_{ws}* of <0.95, indicative of multiple clones present in these infections (Fig. 5). We could hypothesize that, as the patients acquire immunity over successive *P. falciparum* infections, they would be infected with fewer clones but we did not observe any association between polyclonality and the number of infections ($P = 0.19$) (nor with the patient identifier [ID], $P = 0.5$), although more samples will be required to rigorously evaluate this hypothesis.

We also used genetic information extracted from the RNA-seq data to examine relationships among the dominant *P. falciparum* clone of each infection. All clones appeared equally distant from each other (Fig. S2), regardless of whether they were observed in successive infections of the same child or in different children. This analysis is consistent with successful drug treatment following each infection and indicates that consecutive infections in the same individual were caused by new infections rather than by recrudescence of resistant parasites.

DISCUSSION

Here, we applied dual RNA-seq to analyze whole-blood samples collected from three Malian children over five successive *P. falciparum* clinical infections. We successfully obtained more than one million reads from each sample to characterize both host and parasite transcriptomes, allowing robust analysis of differential gene expression, discovery of extensive host and parasite gene coexpression, determination of the propor-

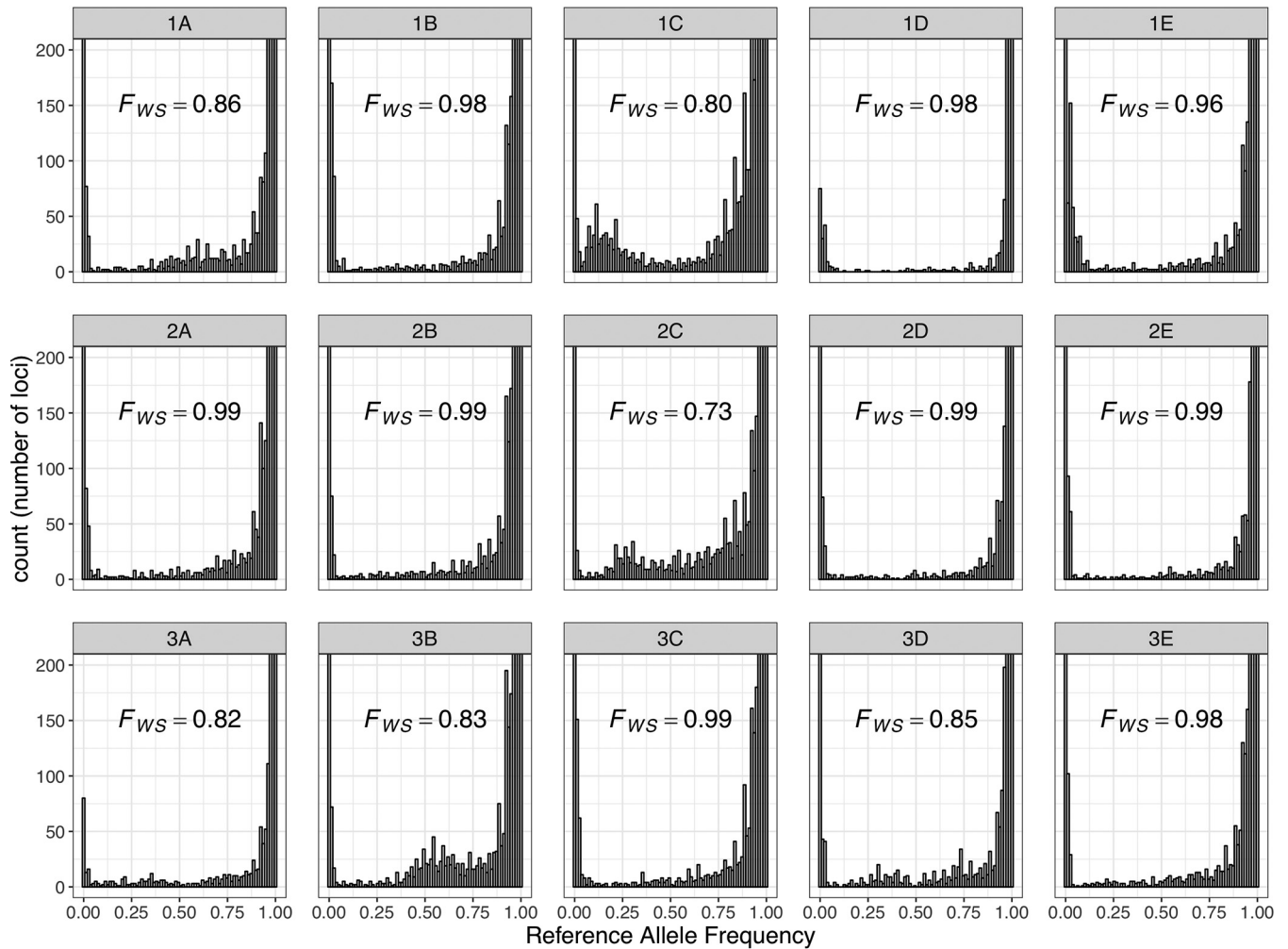


FIG 5 Complexity of infection analysis. The reference allele frequency distributions show, for each sample, the number of nucleotide positions (y axis) with a given proportion of reads carrying the reference allele (x axis). Note that while most infections show a clear U-shaped distribution consistent with the presence of a single (haploid) clone, infections 1A, 1C, 2C, and 3B display clear multimodal distributions consistent with the presence of multiple, genetically different parasites. The corresponding F_{WS} values are indicated in each plot (with $F_{WS} < 0.95$ indicative of polyclonal infections).

tions of the WBC subsets and parasite developmental stages, assessment of the complexity of infection, and parasite genotyping.

One striking result from this analysis was the different patterns of clustering of the host and parasite transcriptomes generated from the same infections: host gene expression profiles appeared to be quantitatively more affected by the individual than by the number of previous infections, while the parasite transcriptomes tended to separate early from late infections. This pattern, which was observed using unsupervised clustering and gene-by-gene analysis, could indicate that transcriptional changes occur in *P. falciparum* parasites in order to successfully infect more malaria-experienced hosts (although the number of genes identified in our analyses remained small, and additional samples would be required to rigorously validate this hypothesis). Similarly, and despite the larger quantitative interindividual variations, many host genes were statistically associated with sequential clinical infections and could hint at the molecular mechanisms involved in the acquisition of immunity against falciparum malaria. Thus, we observed differential host expression of several immunoregulatory genes (Fig. 2A), including $PILR\alpha$ and $BTN2A2$, that were upregulated in successive infections, and $DNTTIP2$ and $MAP3K8$, that were downregulated. $PILR\alpha$ is one member of an immunoglobulin-like receptor gene pair and acts as an innate immune system signaling inhibitor (32). $BTN2A2$ inhibits T-cell metabolism, IL-2 and gamma interferon (IFN- γ)

secretion, and CD4 and CD8 T-cell proliferation (33). MAP3K8 induces production of NF- κ B, a potent inducer of proinflammatory genes (34). These findings are consistent with a progressive dampening of the host inflammatory response over successive infections and mirror some of the gene expression changes described in malaria-experienced hosts compared to malaria-naïve hosts (15). Table S5, displaying GSEA results and leading edge genes driving enrichment, shows that within the enriched pathways there are immunoregulatory genes that increase in expression with infection number such as suppressors of cytokine signaling and cytokine-inducible SH2-containing protein, and suppressors of interferon such as interferon regulatory factor 2 (IRF2) which competitively inhibits IRF1-mediated activation of interferons alpha and beta (35). Interleukin 6 receptor and IL-6 signal transducer genes are also present on significantly enriched pathways. Interleukin 6 has both pro- and anti-inflammatory roles, and inhibits the proinflammatory IL-1 as well as activates the anti-inflammatory IL-10, and the latter has previously been suggested to be involved in antidisease immunity to malaria (18). The identification of genes involved in platelet regulation as differentially regulated upon successive infections is interesting, as platelets have been shown to be involved in parasite killing and clumping of *P. falciparum*-infected erythrocytes, which leads to thrombocytopenia (one complication of malaria) (7). GSEA has previously been used to analyze transcriptional changes during controlled human malaria infection (CHMI), and it is interesting to note the similarities in enriched pathways, including platelet activation and GTPase-mediated signaling found over successive infections in this study compared to days postinoculation versus baseline in *P. falciparum* CHMI (36), and *P. vivax* CHMI of naïve versus semi-immune individuals (37).

Note here that it is possible that the children had malaria episodes prior to enrollment in our study and that infection 1 does not correspond to the child's first malaria infection (although given the young age of the children studied, it is probably one of their first). Techniques used herein, such as differential expression, GSEA, gene signature-based deconvolution, and correlation of host and parasite gene expression, have been used elsewhere for human and *Plasmodium* transcriptomic analysis (15, 16, 19, 36, 38). However, as highlighted in a recent review (38), there is extensive variability in the human subjects compared and techniques used to understand development of malaria immunity, a lack of guidance on methodology to aid defining and characterizing naturally acquired immunity, and absence of detailed time course or infection number transcriptional changes within the same individual. In addition, very little is known about parasite adaptations across successively more malaria-experienced hosts. Overall, while the small sample size of the current study prevents drawing definitive conclusions, our study demonstrates that dual RNA-seq over successive infections can provide a solid framework to better understand transcriptional changes in the parasite and the host accompanying the development of acquired immunity in malaria patients.

Beyond testing for gene expression differences, we leveraged the RNA-seq data to determine the relative proportions of WBC subsets and parasite developmental stages in each sample using gene expression deconvolution (24). Our findings demonstrate that whole-blood RNA-seq is not critically hampered by the cell heterogeneity of each sample but, in contrast, can provide important information and facilitate measurement of changes in WBC subsets over time, and if sample size is sufficient, to correct differential gene expression analyses for these changes to distinguish changes in cell proportions from a difference in gene regulation in a specific cell population. However, we noted that, using gene expression deconvolution, it was difficult to accurately differentiate and quantify cell populations that have similar transcriptional profiles. In particular, we were not able to reliably differentiate CD4⁺ and CD8⁺ T-cell subsets in our analyses and, despite their different biological roles, had to combine these two populations into a single category, though recent progress in gene expression deconvolution methods could address this issue (39).

Finally, we show that data generated by RNA-seq enable determination of the complexity of each infection and comparison of the genotype of the clones in different

samples. This information is critical for studies of successive infections to ensure that the samples analyzed truly represent new infections and not recrudescence, from previous infections, of parasites that are resistant to antimalarial drugs or have been incompletely cleared. This approach could also allow assessment of the role of polyclonality, and possibly of specific parasite genetic polymorphisms, in the response to successive infections. Note however that the determination of allelic variants from RNA might fail to identify polyclonal infections if the different clones in one infection are present at different developmental stages. If this is the case, analyses of genomic DNA might be necessary to avoid misclassifying possible asynchronous polyclonal infections as monoclonal.

Overall, we show that RNA-seq data generated from whole-blood samples collected from children with malaria can provide a wide variety of information to better understand host and parasite changes accompanying the acquisition of immunity against malaria. In addition to the analysis of differential gene expression of the host and parasite associated with successive clinical infections, our study demonstrates that the RNA-seq data can enable identifying host/pathogen interactions, measuring (and correcting for) the proportion of the white blood cell subsets and parasite developmental stages, and determining the clone genotypes and the number of clones present in each infection. The biological complexity of clinical malaria infections involves interactions between a large number of host, parasite, and environmental factors, which would require analyses on a much larger sample size than presented here. A greater number of samples would, for example, enable a rigorous analysis of the interaction between the sex of the host and gene expression of both the host and parasite across successive infections. While the current study is limited by its sample size, application of the approaches implemented here to a larger cohort could provide a novel and comprehensive perspective on the dynamic changes in host and parasite regulation and their interactions during the acquisition of immunity to the disease and could highlight key molecular processes that could then be leveraged to develop more efficient treatment and prevention approaches against malaria.

MATERIALS AND METHODS

Sample collection. Whole-blood samples were collected from five successive symptomatic, uncomplicated infections in three Malian children aged ~1 to 2 years using PAXgene blood RNA tubes (PreAnalytiX). The presence of *P. falciparum* in each sample was confirmed via light microscopic examination of thick blood smears, with no detectable presence of other parasitic species.

Ethics approval and consent. The study protocol and informed consent/assent process were reviewed and approved by the institutional review boards of the Faculty of Medicine, Pharmacy and Dentistry of the University of Sciences, Techniques and Technologies of Bamako and the University of Maryland, Baltimore (IRB numbers HCR-HP-00041382 and HP-00085882). Individual written informed consent was obtained from parents or guardians.

Generation of RNA-seq data. RNA was extracted from PAXgene tubes using the Blood RNA kit (Qiagen) and used to prepare stranded libraries after rRNA and globin depletion using the TruSeq Stranded RNA kit (Illumina) and poly(A) selection using the TruSeq RNA sample preparation v2 kit (Illumina). cDNA libraries were sequenced on an Illumina HiSeq 4000 to generate paired-end reads of 75 bp. To test whether any infection contains more than one *Plasmodium* species, we first randomly subsampled 2,500,000 reads from each fastq file using seqtk v1.3 (<https://github.com/lh3/seqtk>) and aligned those reads using hisat2 v2.0.4 (40) to a fasta file containing the *P. falciparum* 3D7, *P. vivax* PvP01, *P. cynomolgi* M version 2, *P. knowlesi* H strain, *P. malariae* UG01, and *P. ovale* GH01 genomes from PlasmoDB v36 (31). We then counted the number of reads mapped uniquely to each genome using samtools view. We aligned all reads using hisat2 (v2.0.4) (40) to (i) the *P. falciparum* 3D7 genome (PlasmoDB v36 [31]) (with the default parameters except for --max-intronlen 5000, --score-min L,0,-0.4) and (ii) to nonredundant autosomal sequences from the human hg38 genome. We then filtered out any reads mapping to both genomes (always less than 0.17%) and removed potential PCR duplicates with samtools v1.7 markdup. We calculated read counts per gene using the gene annotations downloaded from PlasmoDB (plasmodb.org, for *Plasmodium* genes) and NCBI (for the human genes) and custom python scripts (available at <https://github.com/kbradwell/malaria-dualTranscriptomics>).

Gene expression analysis. The read counts per gene were normalized into count per million reads mapped separately for the human and parasite genes. Unsupervised clustering was performed after calculating Euclidean distances between transcriptomes using the R functions dist() and hclust() (v3.3.1). Statistical assessment of differential gene expression was performed using EdgeR v3.16.5 (41) using simultaneously the number of successive infections and patient ID as covariates (without interactions) and a quasiliikelihood negative binomial generalized model. For these analyses, we considered only genes with >10 counts per million in seven or more samples as expressed and tested a total of 8,896

human genes (out of 17,137 human genes) and 2,822 parasite genes (out of 5,558 parasite genes). Inclusion of parasitemia as a covariate did not notably change the results. All results were corrected for multiple testing by FDR (42).

The PANTHER overrepresentation test (release no. 20190308) was performed using Fisher's exact test with differentially expressed genes (FDR = 0.1) as the test gene set and all 8,896 expressed genes as the reference gene set. GSEA was performed with the R package fgsea v1.0.2 (43), using genes ranked via multiplication of the log fold change with $-\log_{10}(P \text{ value})$, 1,000 permutations, and the reactomePathways() function, which uses NCBI stable ID mappings to pathways, to generate normalized enrichment scores and adjusted *P* values for pathway enrichment.

Gene coexpression analysis. To determine the extent of coexpression between host and parasite genes, we measured the Spearman correlation coefficient between each pair of human and *P. falciparum* genes across all samples using the R function cor.test() with method=spearman. To assess significance of the findings, we determined the number of pairwise correlations with a Spearman's correlation above different R^2 thresholds when randomizing the host and parasite transcriptomes (i.e., by randomly matching the human gene expression profiles and parasite gene expression profiles) and conducting 500 such random permutations. We then determined the significance of the experimental results by calculating the proportion of random permutations with a greater number of pairwise correlations than the number observed at each R^2 threshold) and calculating the enrichment by comparing the number observed experimentally to the average number obtained across all 500 permutations.

Gene expression deconvolution. Reference transcriptome profiles for WBC populations were obtained from FACS-sorted RNA-seq studies (25–28) (Table S8). Reference transcriptome profiles for the different *P. falciparum* developmental stages were obtained from a single-cell RNA-seq study (29). Sufficient male and female gametocyte data were unavailable, and this stage was thus absent from the analysis. Deconvolution was then performed using CIBERSORT v1.06 (24) as described in reference 44. Associations between the proportions of WBC subsets or the parasite developmental stages and successive infections and the child ID were tested by analysis of variance (ANOVA) using the aov() function in Rstudio (v1.0.136).

Complexity of infection and genotyping. Reference allele frequency plots were generated for each sample by measuring the proportion of reads carrying the reference *P. falciparum* allele at each genomic position sequenced $>50\times$. A subset of 3,411,387 positions covered by $>50\times$ in at least two samples was used to determine pairwise differences between the dominant clone of each infection (45), and the resulting distance matrix was used to reconstruct a neighbor-joining tree in MEGA v7170509 (46). *Fws* values were determined by the R package moimix, using curated sites (47) with $>50\times$ coverage.

Data availability. All scripts used in this study are freely available at <https://github.com/kbradwell/malaria-dualTranscriptomics>. All sequence data are available through NCBI Sequence Read Archive under BioProject accession no. PRJNA591657.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, TIF file, 0.6 MB.

FIG S2, TIF file, 0.4 MB.

TABLE S1, XLSX file, 0.01 MB.

TABLE S2, XLSX file, 0.02 MB.

TABLE S3, XLSX file, 2.8 MB.

TABLE S4, XLSX file, 0.9 MB.

TABLE S5, XLSX file, 0.4 MB.

TABLE S6, XLSX file, 0.01 MB.

TABLE S7, XLSX file, 0.01 MB.

TABLE S8, XLSX file, 0.01 MB.

ACKNOWLEDGMENTS

We thank the participants and their families for participating in this study, as well as the community of Bandiagara.

Funding to support this study was provided to the Institute for Genome Sciences, the Malaria Research Program of the Center for Vaccine Development and Global Health, and the Malaria Research and Training Center in Bamako, Mali, by NIH grants R21AI146853, U01AI065683, R01HL130750, R01AI099628, and D43TW001589, a Burroughs Wellcome Fund/American Society of Tropical Medicine and Hygiene Postdoctoral Fellowship, and a Passano Foundation Clinician-Investigator Award.

We declare that we have no competing interests.

D.C., M. A. Travassos, M.B.L., A.D., Y.T., A.K.K., K.T., A.N., and M. A. Thera coordinated sample collection. K.R.B. extracted the RNA and prepared the RNA-seq libraries. K.R.B. and D.S. developed the methodology. K.R.B. performed the analyses. K.R.B., M. A. Travassos, D.S., D.C., M.B.L., A.A.B., B.K., K.E.L., S.T.-H., O.K.D., C.V.P. and M. A. Thera

designed the study. K.R.B., M. A. Travassos, and D.S. wrote the manuscript. All authors read and approved the final manuscript.

REFERENCES

- World Health Organization. 2018. World malaria report 2018. World Health Organization, Geneva, Switzerland.
- UNICEF. 2019. UNICEF under-five mortality, cause of death 2018 report. UNICEF Data, UNICEF, New York, NY. data.unicef.org/wp-content/uploads/2018/09/Cause_of_death_2018-1.xlsx.
- Coulibaly D, Travassos MA, Tolo Y, Laurens MB, Kone AK, Traore K, Sissoko M, Niangaly A, Diarra I, Daou M, Guindo B, Rebaudet S, Kouriba B, Dessay N, Piarroux R, Plowe CV, Doumbo OK, Thera MA, Gaudart J. 2017. Spatio-temporal dynamics of asymptomatic malaria: bridging the gap between annual malaria resurgences in a Sahelian environment. *Am J Trop Med Hyg* 97:1761–1769. <https://doi.org/10.4269/ajtmh.17-0074>.
- Coulibaly D, Travassos MA, Kone AK, Tolo Y, Laurens MB, Traore K, Diarra I, Niangaly A, Daou M, Dembele A, Sissoko M, Guindo B, Douyon R, Guindo A, Kouriba B, Sissoko MS, Sagara I, Plowe CV, Doumbo OK, Thera MA. 2014. Stable malaria incidence despite scaling up control strategies in a malaria vaccine-testing site in Mali. *Malar J* 13:374. <https://doi.org/10.1186/1475-2875-13-374>.
- Langhorne J, Ndungu FM, Sponaas A-M, Marsh K. 2008. Immunity to malaria: more questions than answers. *Nat Immunol* 9:725–732. <https://doi.org/10.1038/nri.205>.
- Ademolue TW, Awandare GA. 2018. Evaluating antidiarrhoeal immunity to malaria and implications for vaccine design. *Immunology* 153:423–434. <https://doi.org/10.1111/imm.12877>.
- Deroost K, Pham T-T, Opendakker G, Van den Steen PE. 2016. The immunological balance between host and parasite in malaria. *FEMS Microbiol Rev* 40:208–257. <https://doi.org/10.1093/femsre/fuv046>.
- Daou M, Kouriba B, Ouédraogo N, Diarra I, Arama C, Keita Y, Sissoko S, Ouologuem B, Arama S, Bousema T, Doumbo OK, Sauerwein RW, Scholzen A. 2015. Protection of Malian children from clinical malaria is associated with recognition of multiple antigens. *Malar J* 14:56. <https://doi.org/10.1186/s12936-015-0567-9>.
- Stevenson MM, Riley EM. 2004. Innate immunity to malaria. *Nat Rev Immunol* 4:169–180. <https://doi.org/10.1038/nri1311>.
- Marsh K, Kinyanjui S. 2006. Immune effector mechanisms in malaria. *Parasite Immunol* 28:51–60. <https://doi.org/10.1111/j.1365-3024.2006.00808.x>.
- Doolan DL, Dobano C, Baird JK. 2009. Acquired immunity to malaria. *Clin Microbiol Rev* 22:13–36. <https://doi.org/10.1128/CMR.00025-08>.
- Griffin JT, Hollingsworth TD, Reyburn H, Drakeley CJ, Riley EM, Ghani AC. 2015. Gradual acquisition of immunity to severe malaria with increasing exposure. *Proc Biol Sci* 282:20142657. <https://doi.org/10.1098/rspb.2014.2657>.
- de Mendonça VR, Barral-Netto M. 2015. Immunoregulation in human malaria: the challenge of understanding asymptomatic infection. *Mem Inst Oswaldo Cruz* 110:945–955. <https://doi.org/10.1590/0074-02760150241>.
- Abel A, Steeg C, Aminkiah F, Addai-Mensah O, Addo M, Gagliani N, Casar C, Yar DD, Owusu-Dabo E, Jacobs T, Mackroth MS. 2018. Differential expression pattern of co-inhibitory molecules on CD4(+) T cells in uncomplicated versus complicated malaria. *Sci Rep* 8:4789. <https://doi.org/10.1038/s41598-018-22659-1>.
- Tran TM, Jones MB, Ongoiba A, Bijker EM, Schats R, Venepally P, Skinner J, Doumbo S, Quinten E, Visser LG, Whalen E, Presnell S, O'Connell EM, Kayentao K, Doumbo OK, Chaussabel D, Lorenzi H, Nutman TB, Ottenhoff THM, Haks MC, Traore B, Kirkness EF, Sauerwein RW, Crompton PD. 2016. Transcriptomic evidence for modulation of host inflammatory responses during febrile *Plasmodium falciparum* malaria. *Sci Rep* 6:31291. <https://doi.org/10.1038/srep31291>.
- Lee HJ, Georgiadou A, Walther M, Nwakanma D, Stewart LB, Levin M, Otto TD, Conway DJ, Coin LJ, Cunningham AJ. 2018. Integrated pathogen load and dual transcriptome analysis of systemic host-pathogen interactions in severe malaria. *Sci Transl Med* 10:eaar3619. <https://doi.org/10.1126/scitranslmed.aar3619>.
- Jagannathan P, Kim CC, Greenhouse B, Nankya F, Bowen K, Eccles-James I, Muhindo MK, Arinaitwe E, Tappero JW, Kamya MR, Dorsey G, Feeney ME. 2014. Loss and dysfunction of Vdelta2(+) gammadelta T cells are associated with clinical tolerance to malaria. *Sci Transl Med* 6:251ra117. <https://doi.org/10.1126/scitranslmed.3009793>.
- Portugal S, Moebius J, Skinner J, Doumbo S, Doumbo D, Kone Y, Dia S, Kanakabandi K, Sturdevant DE, Virtaneva K, Porcella SF, Li S, Doumbo OK, Kayentao K, Ongoiba A, Traore B, Crompton PD. 2014. Exposure-dependent control of malaria-induced inflammation in children. *PLoS Pathog* 10:e1004079. <https://doi.org/10.1371/journal.ppat.1004079>.
- Yamagishi J, Natori A, Tolba MEM, Mongan AE, Sugimoto C, Katayama T, Kawashima S, Makalowski W, Maeda R, Eshita Y, Tuda J, Suzuki Y. 2014. Interactive transcriptome analysis of malaria patients and infecting *Plasmodium falciparum*. *Genome Res* 24:1433–1444. <https://doi.org/10.1101/gr.158980.113>.
- Lee HJ, Georgiadou A, Otto TD, Levin M, Coin LJ, Conway DJ, Cunningham AJ. 2018. Transcriptomic studies of malaria: a paradigm for investigation of systemic host-pathogen interactions. *Microbiol Mol Biol Rev* 82:e00071-17. <https://doi.org/10.1128/MMBR.00071-17>.
- Niangaly A, Gunalan K, Ouattara A, Coulibaly D, Sá JM, Adams M, Travassos MA, Ferrero J, Laurens MB, Kone AK, Thera MA, Plowe CV, Miller LH, Doumbo OK. 2017. *Plasmodium vivax* infections over 3 years in duffy blood group negative Malians in Bandiagara, Mali. *Am J Trop Med Hyg* 97:744–752. <https://doi.org/10.4269/ajtmh.17-0254>.
- Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* 47:D419–D426. <https://doi.org/10.1093/nar/gky1038>.
- Greenwalt D, Lipsky R, Ockenhouse C, Ikeda H, Tandon N, Jamieson G. 1992. Membrane glycoprotein CD36: a review of its roles in adherence, signal transduction, and transfusion medicine. *Blood* 80:1105–1115. <https://doi.org/10.1182/blood.V80.5.1105.1105>.
- Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. 2015. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 12:453–457. <https://doi.org/10.1038/nmeth.3337>.
- Hoek KL, Samir P, Howard LM, Niu X, Prasad N, Galassie A, Liu Q, Allos TM, Floyd KA, Guo Y, Shyr Y, Levy SE, Joyce S, Edwards KM, Link AJ. 2015. A cell-based systems biology assessment of human blood to monitor immune responses after influenza vaccination. *PLoS One* 10:e0118528. <https://doi.org/10.1371/journal.pone.0118528>.
- Hosokawa K, Kajigaya S, Keyvanfar K, Qiao W, Xie Y, Biancotto A, Townsley DM, Feng X, Young NS. 2017. Whole transcriptome sequencing identifies increased CXCR2 expression in PNH granulocytes. *Br J Haematol* 177:136–141. <https://doi.org/10.1111/bjh.14502>.
- Linsley PS, Speake C, Whalen E, Chaussabel D. 2014. Copy number loss of the interferon gene cluster in melanomas is linked to reduced T cell infiltrate and poor patient prognosis. *PLoS One* 9:e109760. <https://doi.org/10.1371/journal.pone.0109760>.
- Wong JJ-L, Ritchie W, Ebner OA, Selbach M, Wong JWH, Huang Y, Gao D, Pinello N, Gonzalez M, Baidya K, Thoeng A, Khoo T-L, Bailey CG, Holst J, Rasko J. 2013. Orchestrated intron retention regulates normal granulocyte differentiation. *Cell* 154:583–595. <https://doi.org/10.1016/j.cell.2013.06.052>.
- Poran A, Nötzel C, Aly O, Mencia-Trinchant N, Harris CT, Guzman ML, Hassane DC, Elemento O, Kafack B. 2017. Single-cell RNA sequencing reveals a signature of sexual commitment in malaria parasites. *Nature* 551:95–99. <https://doi.org/10.1038/nature24280>.
- Palmer C, Diehn M, Alizadeh AA, Brown PO. 2006. Cell-type specific gene expression profiles of leukocytes in human peripheral blood. *BMC Genomics* 7:115. <https://doi.org/10.1186/1471-2164-7-115>.
- Aurrecochea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, Gao J, Gingle A, Grant G, Harb OS, Heiges M, Innamorato F, Iodice J, Kissinger JC, Kraemer E, Li W, Miller JA, Nayak V, Pennington C, Pinney DF, Roos DS, Ross C, Stoeckert C, Treatman C, Wang H. 2009. PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res* 37:D539–D543. <https://doi.org/10.1093/nar/gkn814>.
- Mousseau DD, Banville D, L'Abbé D, Bouchard P, Shen SH. 2000. PILRalpha, a novel immunoreceptor tyrosine-based inhibitory motif-bearing protein, recruits SHP-1 upon tyrosine phosphorylation and is paired with the truncated counterpart PILRbeta. *J Biol Chem* 275:4467–4474. <https://doi.org/10.1074/jbc.275.6.4467>.

33. Sarter K, Leimgruber E, Gobet F, Agrawal V, Dunand-Sauthier I, Barras E, Mastelic-Gavillet B, Kamath A, Fontannaz P, Guery L, Duraes FDV, Lipens C, Ravn U, Santiago-Raber M-L, Magistrelli G, Fischer N, Siegrist C-A, Hugues S, Reith W. 2016. Btn2a2, a T cell immunomodulatory molecule coregulated with MHC class II genes. *J Exp Med* 213:177–187. <https://doi.org/10.1084/jem.20150435>.
34. Chorzalska A, Ahsan N, Rao RSP, Roder K, Yu X, Morgan J, Tepper A, Hines S, Zhang P, Treaba DO, Zhao TC, Olszewski AJ, Reagan JL, Liang O, Gruppuso PA, Dubielecka PM. 2018. Overexpression of Tpl2 is linked to imatinib resistance and activation of MEK-ERK and NF-kappaB pathways in a model of chronic myeloid leukemia. *Mol Oncol* 12:630–647. <https://doi.org/10.1002/1878-0261.12186>.
35. Han K-J, Jiang L, Shu H-B. 2008. Regulation of IRF2 transcriptional activity by its sumoylation. *Biochem Biophys Res Commun* 372:772–778. <https://doi.org/10.1016/j.bbrc.2008.05.103>.
36. Rothen J, Murie C, Carnes J, Anupama A, Abdulla S, Chemba M, Mpina M, Tanner M, Lee Sim BK, Hoffman SL, Gottardo R, Daubenberger C, Stuart K. 2018. Whole blood transcriptome changes following controlled human malaria infection in malaria pre-exposed volunteers correlate with parasite prepatent period. *PLoS One* 13:e0199392. <https://doi.org/10.1371/journal.pone.0199392>.
37. Gardinassi LG, Arévalo-Herrera M, Herrera S, Cordy RJ, Tran V, Smith MR, Johnson MS, Chacko B, Liu KH, Darley-Usmar VM, Go Y-M, MaHPIC Consortium, Jones DP, Galinski MR, Li S. 2018. Integrative metabolomics and transcriptomics signatures of clinical tolerance to *Plasmodium vivax* reveal activation of innate cell immunity and T cell signaling. *Redox Biol* 17:158–170. <https://doi.org/10.1016/j.redox.2018.04.011>.
38. Hodgson SH, Muller J, Lockstone HE, Hill AVS, Marsh K, Draper SJ, Knight JC. 2019. Use of gene expression studies to investigate the human immunological response to malaria infection. *Malar J* 18:418–418. <https://doi.org/10.1186/s12936-019-3035-0>.
39. Monaco G, Lee B, Xu W, Mustafah S, Hwang YY, Carré C, Burdin N, Visan L, Ceccarelli M, Poidinger M, Zippelius A, Pedro de Magalhães J, Larbi A. 2019. RNA-Seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. *Cell Rep* 26:1627–1640.e7. <https://doi.org/10.1016/j.celrep.2019.01.041>.
40. Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12:357–360. <https://doi.org/10.1038/nmeth.3317>.
41. Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140. <https://doi.org/10.1093/bioinformatics/btp616>.
42. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* 57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
43. Korotkevich G, Sukhov V, Sergushichev A. 2019. Fast gene set enrichment analysis. *bioRxiv* <https://doi.org/10.1101/060012>.
44. Kim A, Popovici J, Menard D, Serre D. 2019. *Plasmodium vivax* transcriptomes reveal stage-specific chloroquine response and differential regulation of male and female gametocytes. *Nat Commun* 10:371. <https://doi.org/10.1038/s41467-019-08312-z>.
45. Chan ER, Menard D, David PH, Ratsimbao A, Kim S, Chim P, Do C, Witkowski B, Mercereau-Puijalon O, Zimmerman PA, Serre D. 2012. Whole genome sequencing of field isolates provides robust characterization of genetic diversity in *Plasmodium vivax*. *PLoS Negl Trop Dis* 6:e1811. <https://doi.org/10.1371/journal.pntd.0001811>.
46. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874. <https://doi.org/10.1093/molbev/msw054>.
47. Manske M, Miotto O, Campino S, Auburn S, Almagro-Garcia J, Maslen G, O'Brien J, Djimde A, Doumbo O, Zongo I, Ouedraogo J-B, Michon P, Mueller I, Siba P, Nzila A, Borrmann S, Kiara SM, Marsh K, Jiang H, Su X-Z, Amaratunga C, Fairhurst R, Socheat D, Nosten F, Imwong M, White NJ, Sanders M, Anastasi E, Alcock D, Drury E, Oyola S, Quail MA, Turner DJ, Ruano-Rubio V, Jyothi D, Amenga-Etego L, Hubbart C, Jeffreys A, Rowlands K, Sutherland C, Roper C, Mangano V, Modiano D, Tan JC, Ferdig MT, Amambua-Ngwa A, Conway DJ, Takala-Harrison S, Plowe CV, Rayner JC, Rockett KA, Clark TG, Newbold CI, Berriman M, MacInnis B, Kwiatkowski DP. 2012. Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* 487:375–379. <https://doi.org/10.1038/nature11174>.