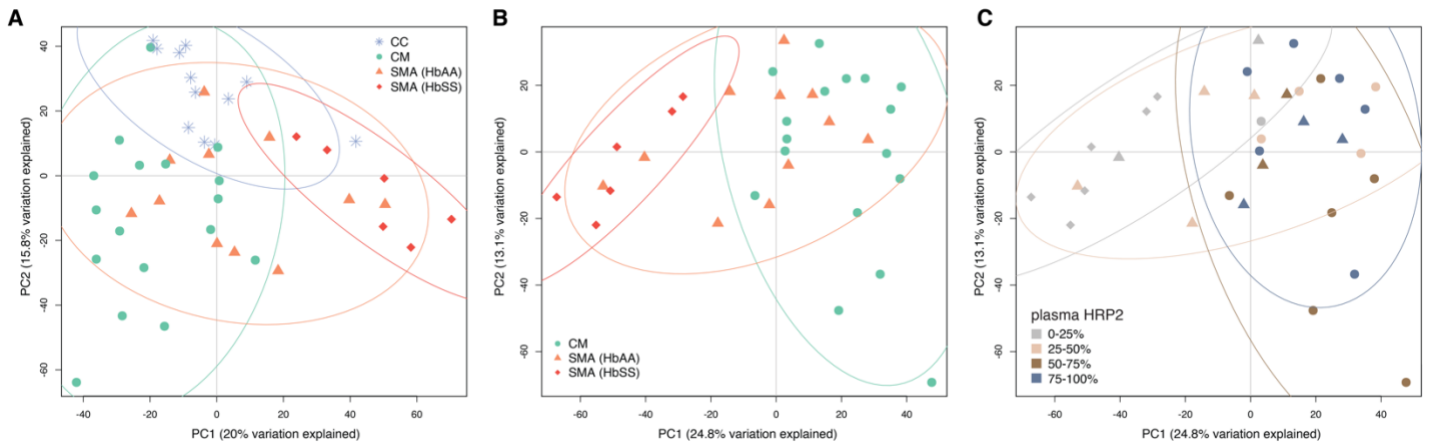


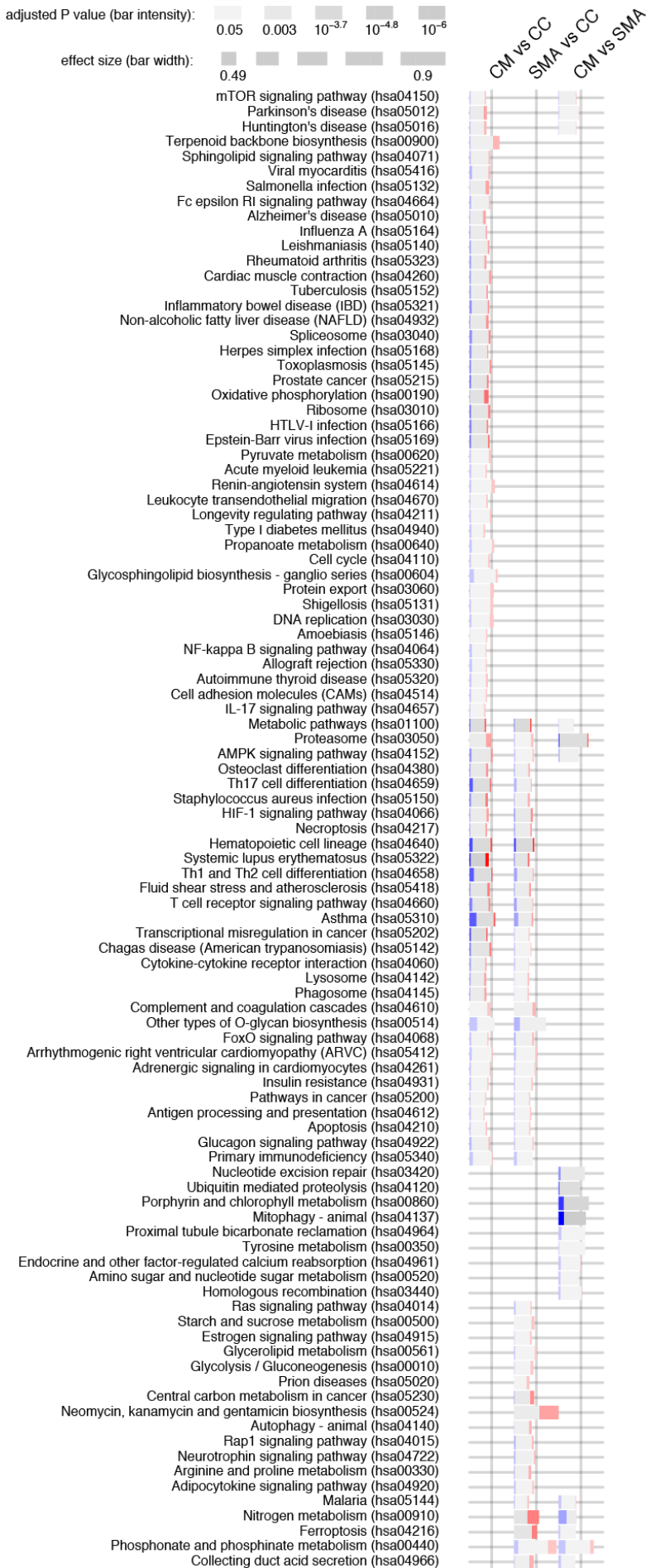
**Supplementary Figure 1.** Parent cohort and selection of samples for microarray analysis. CM = cerebral malaria; SMA = severe malarial anemia; CC = community children; RIN = RNA integrity number.



**Supplementary Figure 2.** Principal components analysis of whole-blood expression data for 47,323 genes with grouping by clinical presentation with (A) and without (B) healthy community children (CC; stars) or by (C) quartiles for plasma HRP2 level. For C, only parasitemic samples are shown. CM = cerebral malaria (circles); SMA HbAA = severe malarial anemia without sickle cell (triangles); SMA HbSS = severe malarial anemia with sickle cell (diamonds). The ellipses represent 95% confidence intervals for each cluster. Batch effects were removed from expression data prior to generating plots.

**Supplementary Table 1.** Analysis of variance to determine differences in means between sample clusters in Figure 2 for the indicated variables. Diff= difference in observed means; LWR = lower end point of interval; UPR = upper end point of interval. Intervals are based on Tukey's Honest Significant Difference method with P values adjusted for multiple comparisons. Significant P values (<0.05) are bolded.

Variable	Cluster 2 – Cluster 1				Cluster 3 – Cluster 1				Cluster 3 – Cluster 2			
	Diff	LWR	UPR	adj. P value	Diff	LWR	UPR	adj. P value	Diff	LWR	UPR	adj. P value
log <sub>10</sub> (parasites/ul)	4.32	3.20	5.43	<b>1.65E-11</b>	3.26	1.91	4.61	<b>1.65E-06</b>	-1.06	-2.24	0.13	0.089
HRP2	4.94	3.42	6.47	<b>2.06E-09</b>	1.62	-0.22	3.47	0.094	-3.32	-4.94	-1.70	<b>3.22E-05</b>
Age (years)	-1.13	-2.74	0.49	0.22	-0.33	-2.29	1.63	0.911	0.79	-0.93	2.52	0.510
Weight (kg)	-3.09	-6.98	0.81	0.14	-0.90	-5.62	3.81	0.888	2.19	-1.96	6.33	0.414
WBC count	6.24	-2.34	14.82	0.19	12.81	2.36	23.27	<b>0.013</b>	6.57	-2.87	16.02	0.220
Neutrophil count	5.21	1.45	8.97	<b>4.75E-03</b>	5.08	0.13	10.02	<b>0.043</b>	-0.13	-4.67	4.40	0.997
Lymphocyte count	0.68	-3.39	4.74	0.91	4.69	-0.43	9.82	0.078	4.02	-0.65	8.68	0.103
Monocyte count	0.82	-0.15	1.79	0.11	0.67	-0.56	1.89	0.391	-0.15	-1.27	0.97	0.941
Hemoglobin	-5.10	-6.81	-3.40	<b>1.50E-08</b>	-7.00	-9.06	-4.94	<b>6.28E-10</b>	-1.90	-3.71	-0.09	<b>0.038</b>
Platelet count	-195.52	-288.93	-102.12	<b>2.57E-05</b>	-25.97	-136.49	84.55	0.836	169.56	70.38	268.73	<b>4.70E-04</b>



**Supplementary Figure 3.** CERNO gene set enrichment testing in Tmod [1] using KEGG pathways for the indicated comparisons between the clinical groups cerebral malaria (CM), several malarial anemia (SMA), and healthy community children (CC) with adjustment for HbSS status. Only modules with a Benjamini-Hochberg-adjusted *P* value <0.05 are shown. Effect size (bar width) represents the area under the curve in the CERNO test. Red indicates the proportion of genes upregulated in the first group versus the second group, whereas blue indicates the proportion of genes downregulated in the first group. Gray indicates that proportion of genes for which expression direction could not be assigned based on a pre-set significance threshold (FDR<20%).

**Supplementary Table 2.** List of differentially expressed genes for the CM versus CC comparison.

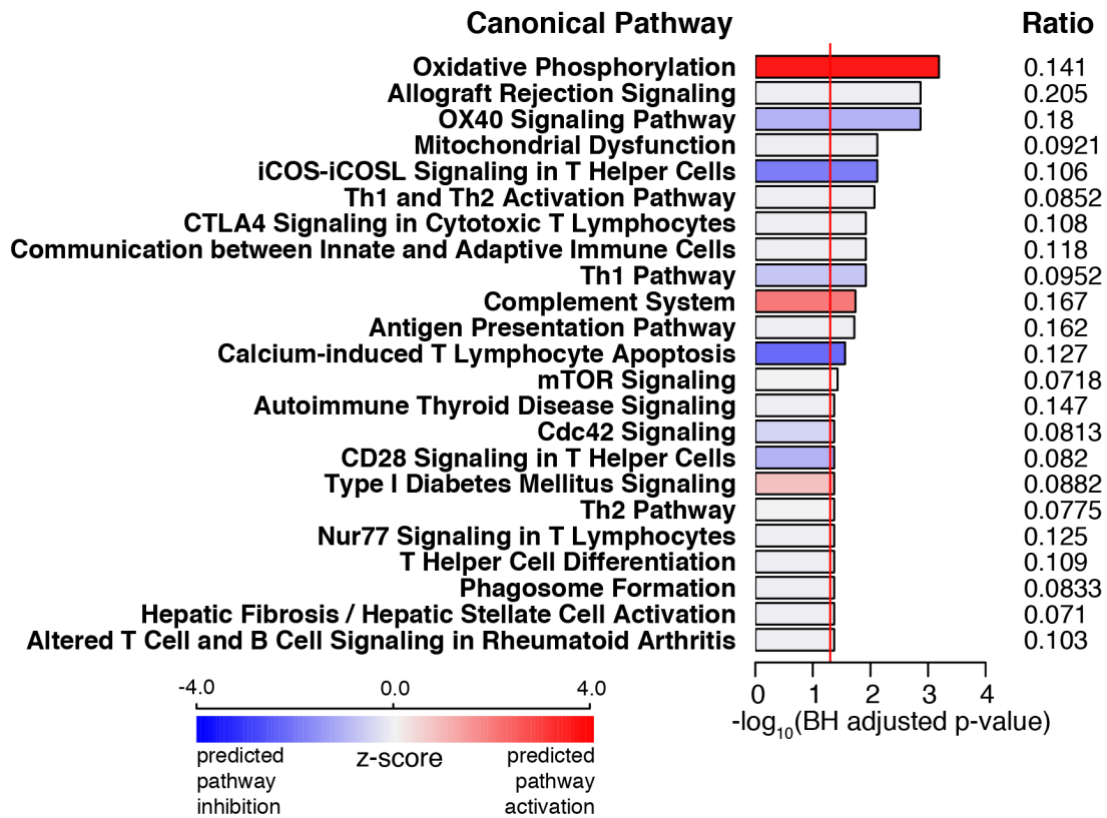
<https://iu.box.com/v/tableS2>

**Supplementary Table 3.** List of differentially expressed genes for the SMA versus CC comparison.

<https://iu.box.com/v/tableS3>

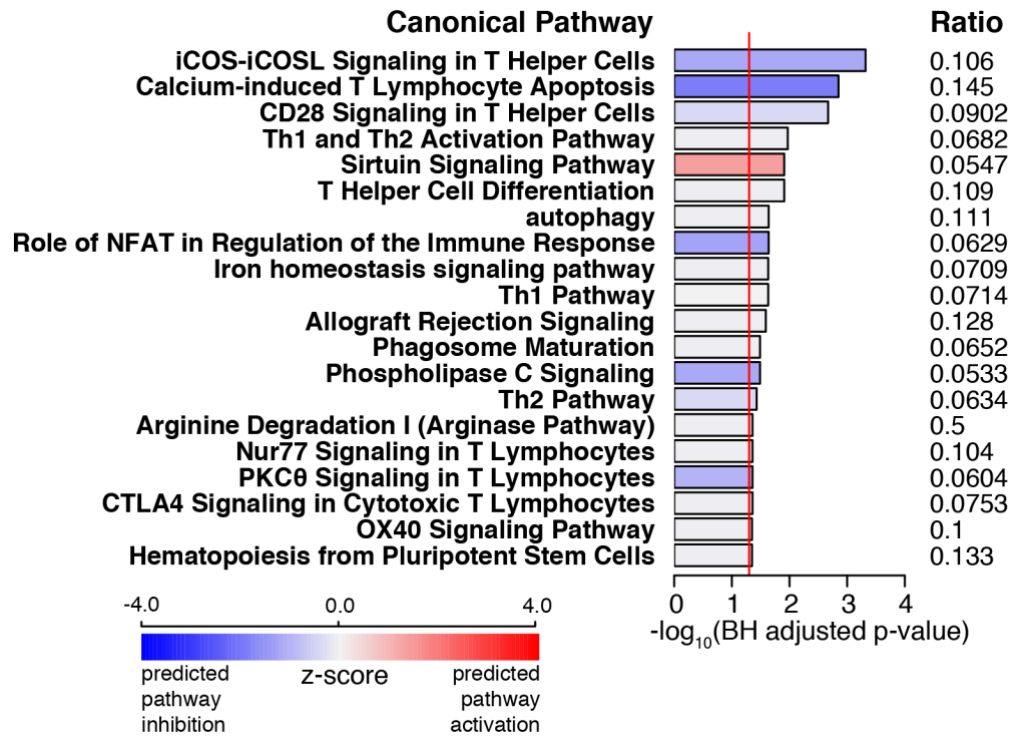
**Supplementary Table 4.** List of differentially expressed genes for the CM versus SMA comparison.

<https://iu.box.com/v/tableS4>



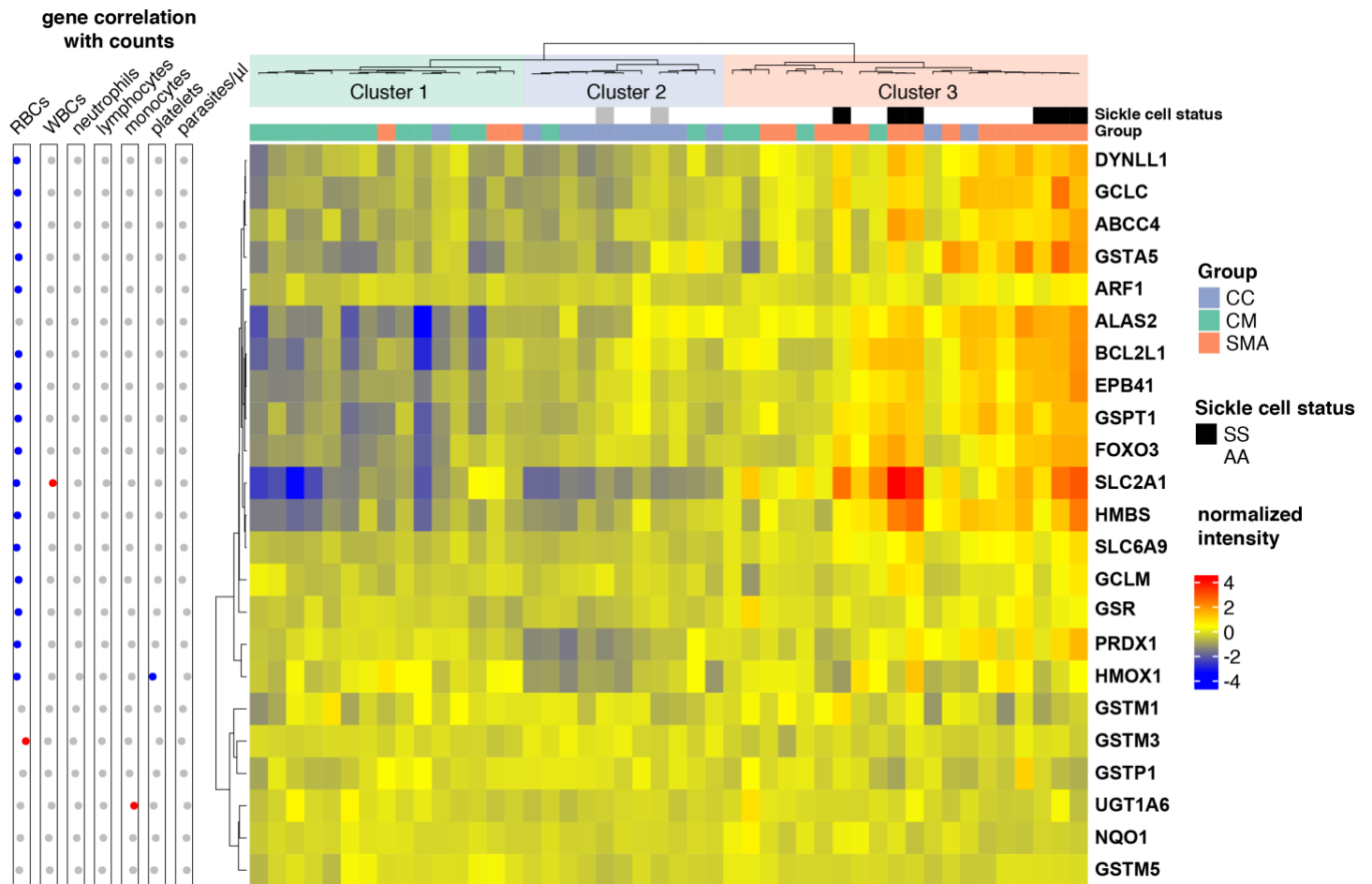
**Supplementary Figure 4.**

Overrepresented Ingenuity canonical pathways among DEGs with significance thresholds of  $|\log_2 \text{ fold change}| > 0.585$  and false discovery rate  $< 10\%$  for the CM versus CC comparison. Differential gene expression analysis included HbSS and HRP2 levels as a co-variates. Gray indicates no activity pattern available. Ratio indicates the proportion of DEGs that overlap with genes in the indicated pathway.



**Supplementary Figure 5**

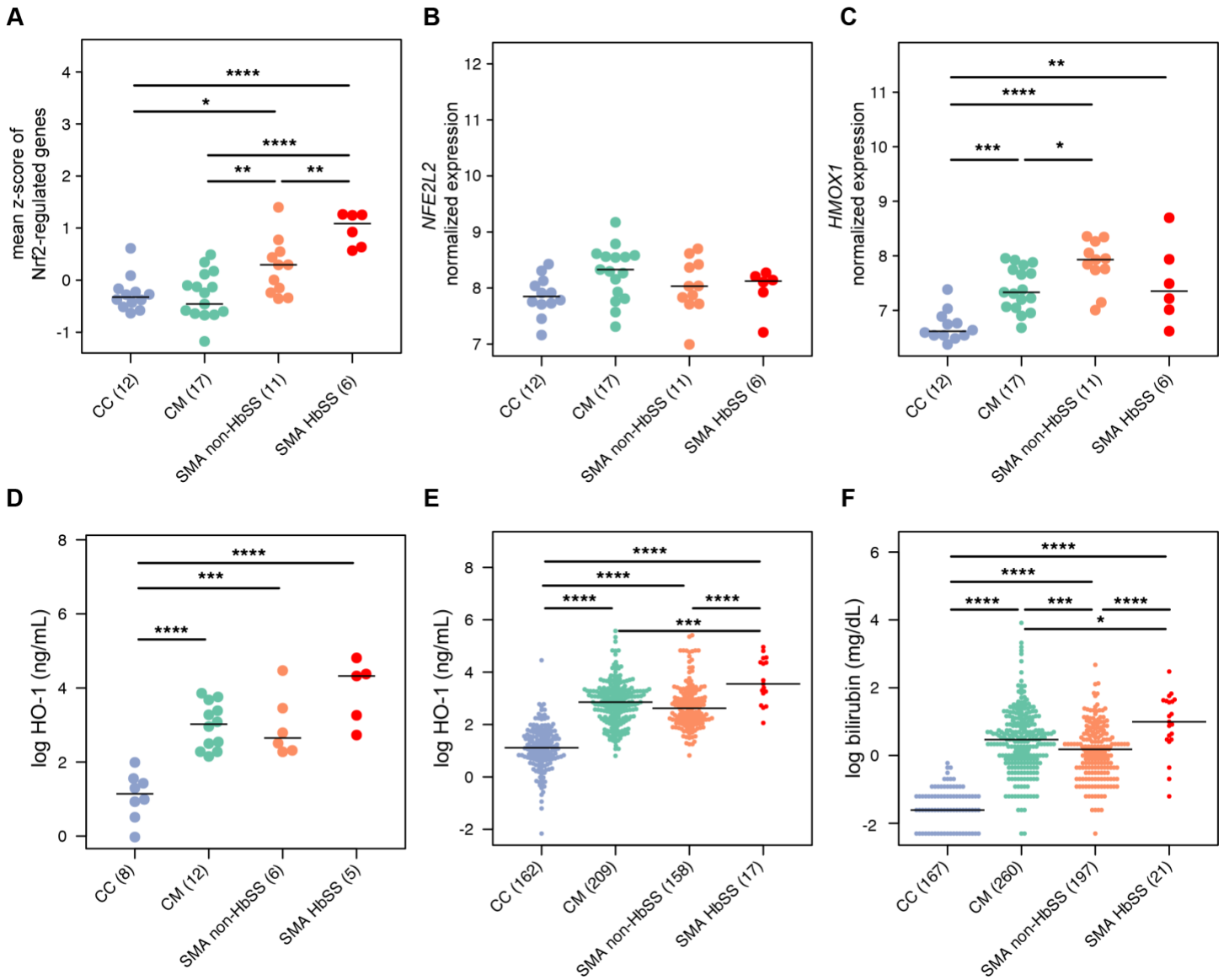
Overrepresented Ingenuity canonical pathways among DEGs with significance thresholds of  $|\log_2 \text{fold change}| > 0.585$  and false discovery rate  $< 10\%$  for the SMA versus CC comparison. Differential gene expression analysis included HbSS status and HRP2 levels as co-variates. Gray indicates no activity pattern available. Ratio indicates the proportion of DEGs that overlap with genes in the indicated pathway.



**Supplementary Figure 6**

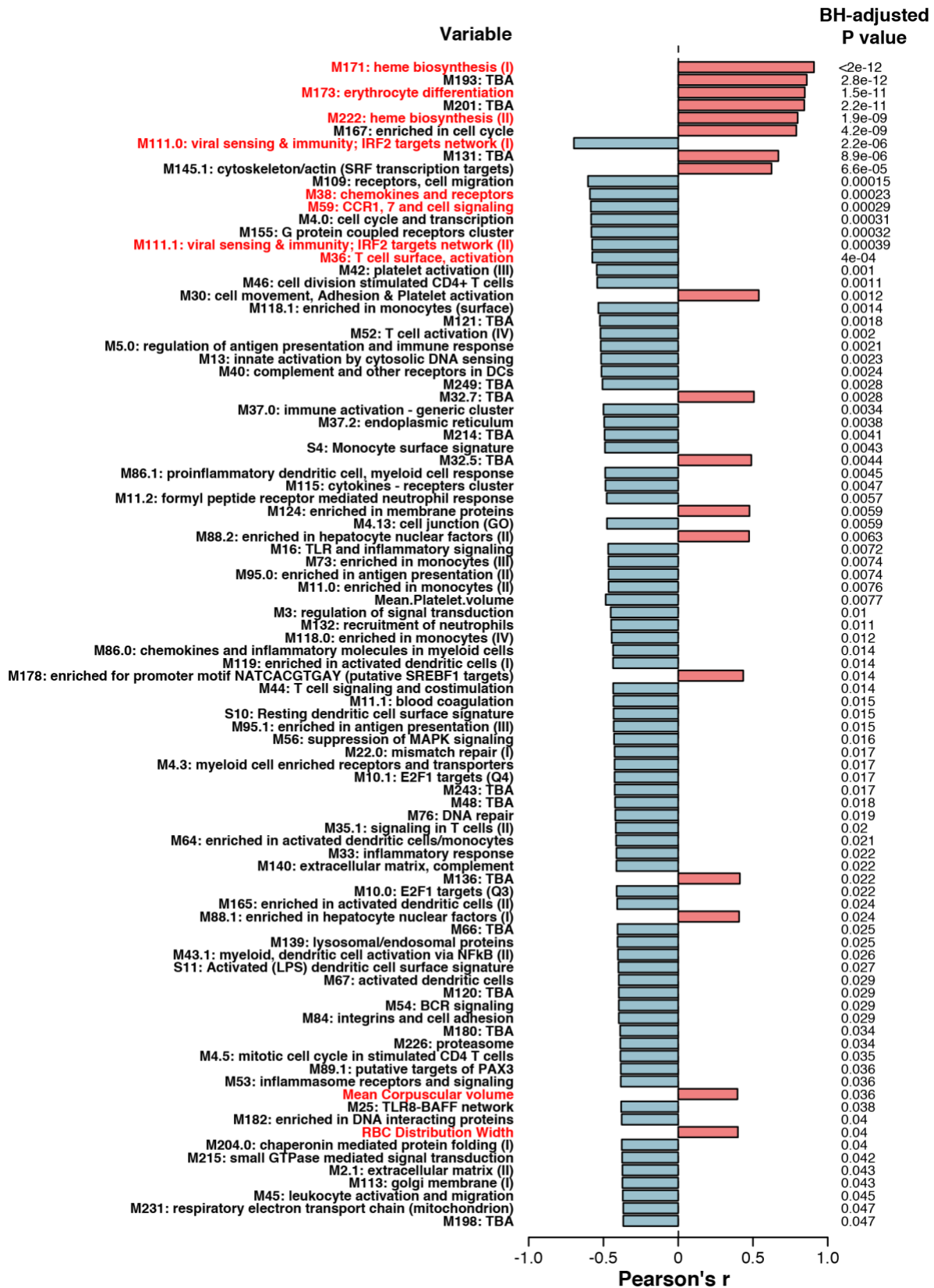
Heat map of 23 genes known to be regulated by Nrf2 using the Ingenuity upstream regulator database. “Correlation with cell count” row annotations show Pearson’s coefficients for correlations between expression for each gene with indicated cell count using data from all samples. Unsupervised row and column clustering was performed using Pearson correlation and Ward.D2. Significance of sample clustering of groups within Clusters 1-3 was determined using the Fisher’s exact test ( $P = 4.1 \times 10^{-7}$ ).





**Supplementary Figure 7**

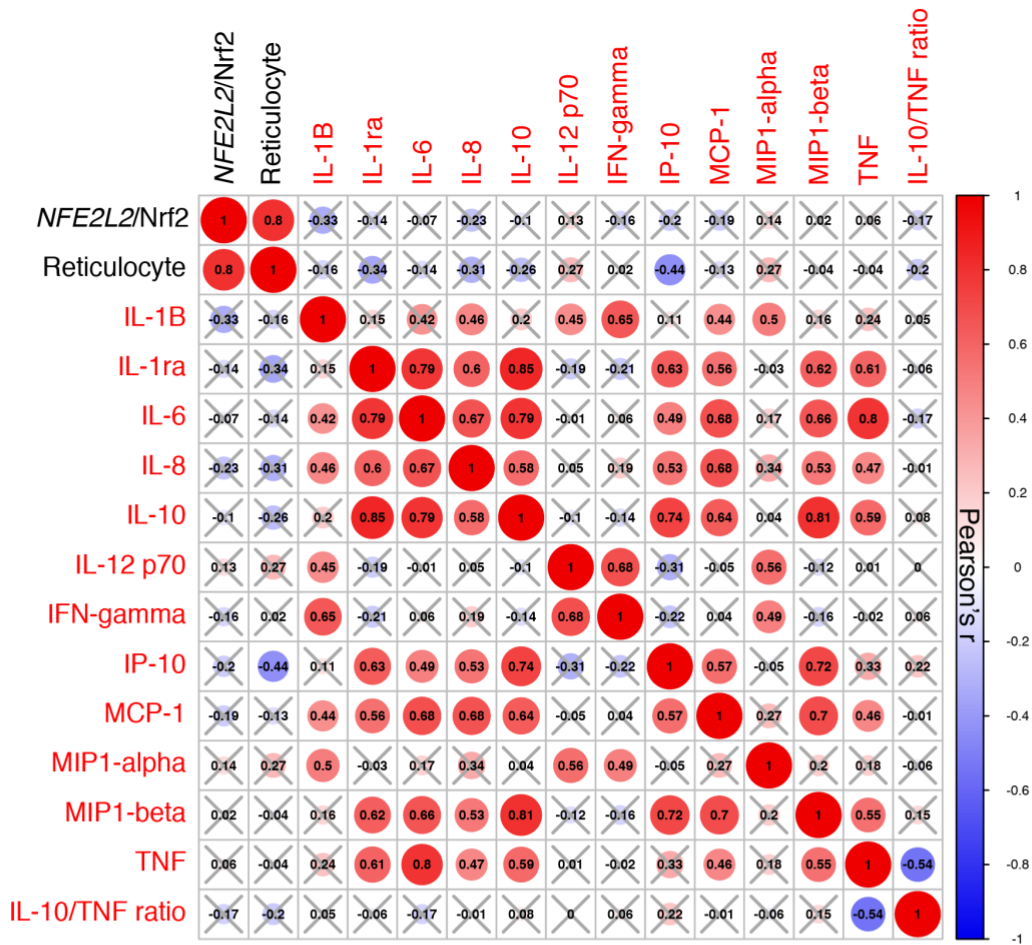
Normalized expression of (A) Nrf2-regulated genes, (B) *NFE2L2* and (C) *HMOX1*. Plasma concentrations of heme oxygenase-1 (HO-1) for (D) children with microarray data and (E) children within the larger cohort for whom plasma was available. (F) Serum total bilirubin levels for subjects within the larger cohort. Sample sizes for each group indicated in parentheses. Bars within beeswarms represent group medians. Pairwise significance was determined by analysis of variance with Tukey's Honest Significance Differences. Significance indicated by \* <0.05, \*\* <0.01, \*\*\* <0.001, \*\*\*\*<0.0001; all other comparisons were not statistically significant. CC = community controls, CM = cerebral malaria, SMA = severe malarial anemia, HbSS = sickle cell disease.



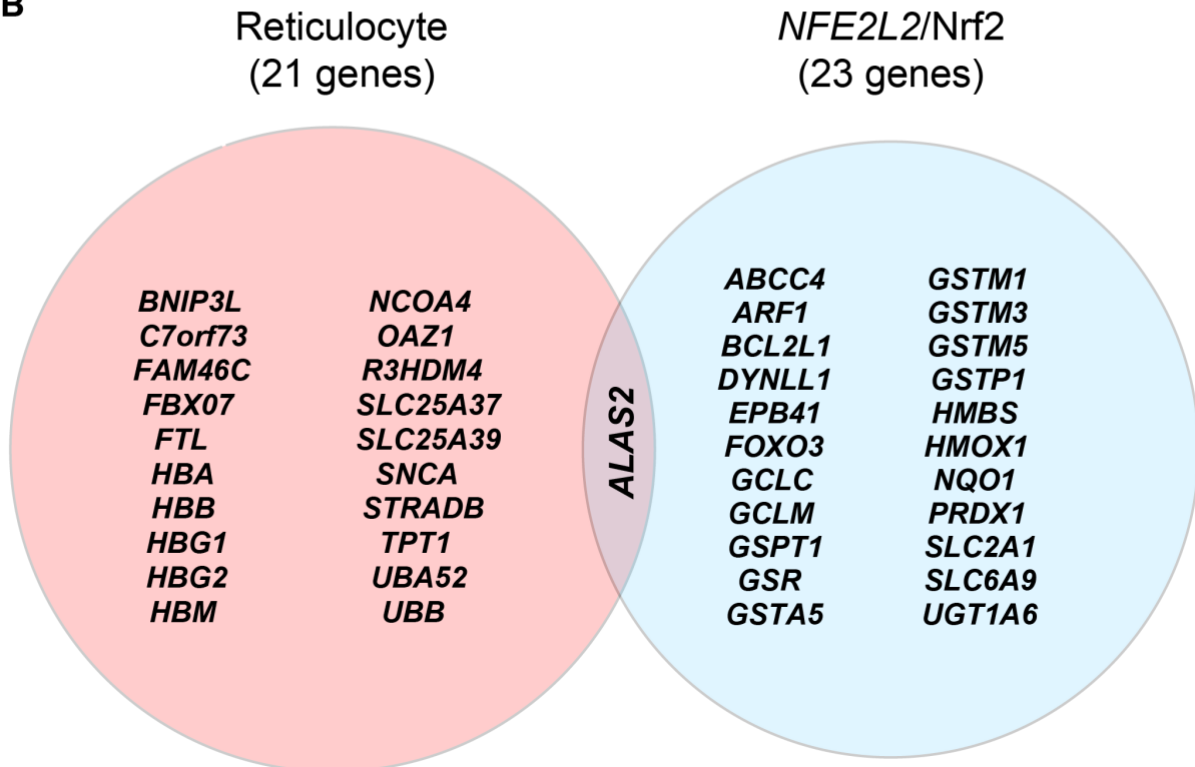
### Supplementary Figure 8

Pearson correlation between reticulocyte-specific gene signature expression and each of the 20 hematologic indices and 346 blood transcription modules. Only correlations with a Benjamini-Hochberg (BH) adjusted  $P$  value of  $<0.05$  are shown. Variables highlighted in red are referenced in main text.

**A**



**B**



**Supplementary Figure 9**

(A) Correlation matrix for pair-wise comparisons between expression values of Nrf2-regulated and reticulocyte-specific gene signatures (black labels), log-transformed pro-inflammatory plasma cytokine levels, and IL-10/TNF ratio (red labels). Squares marked with an “X” do not meet significance criteria of  $<0.05$  for unadjusted  $P$  value (below the diagonal) or Benjamini-Hochberg-adjusted  $P$  value  $<0.05$  (above the diagonal). (B) Venn diagram of genes within the reticulocyte-specific and *NFE2L2/Nrf2* gene sets.

## **Detailed Methods**

### **Study Site and Participants**

From November 2008 through December 2010, 437 children between 18 months to 12 years of age with CM (n=166) or SMA (n=152) were enrolled as part of a larger study conducted at Mulago Hospital, Kampala, Uganda (Supplementary Figure 1). SMA was defined as presence of *P. falciparum* on blood smear and a hemoglobin level  $\leq 5$  g/dL. CM was defined as: 1) coma (Blantyre Coma Score  $\leq 2$ ); 2) *P. falciparum* on blood smear; and 3) no other known cause of coma (e.g., meningitis, a prolonged postictal state or hypoglycemia-associated coma). Community children (CC; n=140) were healthy participants recruited from the households of children with CM or SMA. Children with malaria were managed according to the Ugandan Ministry of Health treatment guidelines current at the time of the study. Exclusion criteria included known chronic illness; known developmental delay; or prior history of coma, head trauma, hospitalization for malnutrition, or cerebral palsy. Additional exclusion criteria for children with SMA included impaired consciousness on exam; other clinical evidence of central nervous system disease; or  $>1$  seizure prior to admission. Additional exclusion criteria for CC included illness requiring medical care within the previous month or major medical/neurological abnormalities on screening physical exam. For the present study, to provide the most rigorous definition of CM, we further restricted CM to include only children with retinopathy confirmed by ophthalmological evaluation and hemoglobin values  $>5$  g/dL (thus excluding CM children with concurrent SMA). Children with sickle cell disease (HbSS), sickle cell trait, or HIV were not excluded.

### **Blood collection**

Peripheral blood was collected by venipuncture on hospital admission (CM and SMA) or as an outpatient (CC) for thin and thick smears, hematological testing (complete blood count and erythrocyte indices), plasma for cytokine and biomarker assays, and filter paper blood spots. Additional whole blood was collected for RNA using the PAXgene Blood RNA System (PreAnalytiX; Hombrechtikon, Switzerland) with a modified procedure that utilized a blood volume smaller than the manufacturer's instructions while maintaining the appropriate blood-to-PAXgene additive ratio. Plasma and PAXgene whole-blood samples were stored at  $-80^{\circ}\text{C}$  until further processing

and analysis. No power outages or other complications affected sample stability during the time of storage. Additional laboratory testing was performed as clinically indicated.

### **Laboratory testing**

Complete blood count with differential and platelet count was determined using a Coulter Counter (Beckman Coulter, Miami, FL). The presence of *Plasmodium* was determined by microscopy of Giemsa-stained thick blood smears. For positive samples, species was determined by thin smear, and asexual parasite density was determined by enumerating the number of parasites/200 leukocytes and converting to parasites/ $\mu$ l using a leukocyte density of 8,000 cells/ $\mu$ l of blood. Two independent readings were conducted. Discordant readings were resolved by a third reading. Plasma erythropoietin (EPO) levels were tested by high-sensitivity radioimmunoassay, as previously described [3]. PfHRP-2 levels were determined using the Malaria Antigen Cellabs Enzyme-Linked Immunosorbent Assay (Cellabs, Brookvale, Australia). Heme oxygenase-1 levels were measured using the human HO-1 StressXpress ELISA kit per the manufacturer's instructions (Stressgen/ENZO Lifesciences, Ann Arbor, MI). HbS genotyping was determined by PCR amplification of DNA extracted from dried blood spots using beta hemoglobin specific primers with confirmation by hemoglobin electrophoresis.

### **RNA processing and BeadChip microarray procedures**

Total RNA was purified from PAXgene-stabilized whole blood and depleted of globin mRNA transcripts using GLOBINclear Kits (Life Technologies, Grand Island, NY). RNA quality was assessed at the University of Minnesota Genomics Center (UMGC) with either an Agilent 2100 Bioanalyzer or a Perkin Elmer LabChip GX and quantity was assessed with a NanoDrop 1000A Spectrophotometer. After RNA purification and quality assessment, computerized random sampling was used to select 20 samples from each group for transcriptome analyses based on sufficient RNA yield and RNA integrity numbers ( $>6.1$ ). Approximately 300 ng of total RNA was used in Illumina TotalPrep RNA Amplification Kits to produce cRNA. Whole genome expression analyses were performed using HumanHT-12 v4 Expression BeadChips (Illumina, San Diego, CA) and an Illumina BeadArray Reader. Gene expression profiling experiments were conducted as three batches consisting of approximately the same number of samples from each group per batch. Expression data was deposited in the Gene

Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) and is accessible through the accession number GSE117613.

### **Differential gene expression analysis**

Data processing, quality control, and differential gene expression analysis were performed in R (version 3.5.1). Raw data from three batches were imported into R from the BeadStudio. BeadArray-specific variance stabilization, normalization, and gene annotation were performed using the *lumi* package. Data from all three batches were combined prior to quantile normalization using the `normalizeBetweenArrays` function from the *limma* package. Low-expression features with a normalized intensity <6.13 in at least 12 samples were removed. The remaining 33,546 features were used to perform the differential expression analysis. The design matrix was set to adjust for presence of sickle cell disease (HbSS genotype), HRP2 levels, and batch effects using the following model formula:

$$\sim \text{Class} + \text{HbSS} + \text{HRP2} + \text{Batch}$$

where Class is a factor variable of CM, SMA, or CC with CC as the reference level; HRP2 is a numeric variable; and HbSS and Batch are factor variables. Differential gene expression analysis was determined using the empirical Bayes moderated T-test in the *limma* package to obtain log<sub>2</sub> fold-change (LFC) values and Benjamini-Hochberg false discovery rates (FDR) for each of the following contrasts: CM vs. CC, SMA vs. CC, and CM vs. SMA. Differential enrichment of functional pathways was assessed by CERNO testing [4] in the *tmod* package [1] using blood transcription modules [5, 6] and KEGG biological pathways as genesets and the same contrasts as in the differential gene expression analysis. Differentially expressed genes with an absolute LFC > 0.585 (1.5-fold in linear space) and FDR < 5% were applied to Ingenuity Pathways Analysis to determine over-represented canonical pathways and upstream regulators. Expression values for blood transcription modules [5] and the reticulocyte signature [7] were calculated as the mean of z-score-transformed expression values for all genes within a module/gene set.

### **Plasma cytokines**

For study participants who had available plasma samples, we measured plasma concentrations of using the Bio-Plex Pro Human Cytokine 27-plex Assay (Bio-Rad, Hercules, CA) in plasma diluted 1:4, according to

manufacturer's instruction. We selected ten analytes from the 27-plex assay based on their relevance to malaria-induced inflammation for analysis: IL-1 $\beta$ , IL-1 receptor antagonist (IL-1Ra), IL-10, IL-12p70, IFN $\gamma$ , IP-10, MCP-1, MIP1- $\alpha$ , and MIP1- $\beta$ . In addition, plasma levels of TNF and IL-6 were measured by magnetic cytometric bead assay (EMD-Millipore, Billerica, MA) according to the manufacturer's instructions. All the testing was performed with a BioPlex-200 system (Bio-Rad, Hercules, CA). To assess intra-assay reproducibility, 10% of samples were randomly selected from each assay plate for re-testing. The mean coefficient of variance for these samples for all analytes was 25.8%.

### **Statistical analyses.**

Statistical analyses were performed in R (version 3.5.1). Significance testing and linear regression models were performed as indicated in figure legends with the following additions. To determine the effect of EPO levels and clinical group on RDW or reticulocyte-specific gene expression, a model in which an interaction between EPO and clinical group was initially explored. Since interaction terms were non-significant, we used parallel slopes linear regression models in which RDW or reticulocyte-specific gene expression was the response variable and EPO levels and group were predictor variables for the final analyses. For the multiple regression analysis that examined the effect of M111.1 and IP-10 on reticulocyte specific-gene expression (Table 2), the following formula was used:

$$\text{Reticulocyte-specific module expression} = \text{M111.1 expression} + \log(\text{IP-10 levels}) + \text{absolute white blood count} + \log(\text{hemoglobin level}) + \text{HRP2} + \text{clinical group}$$

where expression of the reticulocyte-specific and M111.1 modules were calculated as the mean of z-score-transformed expression values of genes within each respective gene set. Clinical group is a factor consisting of four groups: CM (reference), CC, SMA HbAA, and SMA HbSS. Results were considered significant when  $P < 0.05$ , with adjustment for multiple comparisons when appropriate. The multiple regression analysis was limited to 41 children for whom data was available for all variables.

## References in Supplementary Materials

1. Weiner 3rd J, Domaszewska T. tmod: an R package for general and multivariate enrichment analysis. *PeerJ Preprints* **2016**; 4:e2420v1.
2. Ssenkusu JM, Hodges JS, Opoka RO, et al. Long-term Behavioral Problems in Children With Severe Malaria. *Pediatrics* **2016**; 138.
3. Georgieff MK, Landon MB, Mills MM, et al. Abnormal iron distribution in infants of diabetic mothers: spectrum and maternal antecedents. *J Pediatr* **1990**; 117:455-61.
4. Yamaguchi KD, Ruderman DL, Croze E, et al. IFN-beta-regulated genes show abnormal expression in therapy-naive relapsing-remitting MS mononuclear cells: gene expression analysis employing all reported protein-protein interactions. *J Neuroimmunol* **2008**; 195:116-20.
5. Li S, Roupael N, Duraisingam S, et al. Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat Immunol* **2014**; 15:195-204.
6. Kazmin D, Nakaya HI, Lee EK, et al. Systems analysis of protective immune responses to RTS,S malaria vaccination in humans. *Proc Natl Acad Sci U S A* **2017**; 114:2425-30.
7. Goh SH, Josleyn M, Lee YT, et al. The human reticulocyte transcriptome. *Physiol Genomics* **2007**; 30:172-8.