Modelling pathogen load dynamics to elucidate mechanistic determinants of host-*Plasmodium falciparum* interactions

Athina Georgiadou[#1], Hyun Jae Lee[#,2a], Michael Walther[3b], Anna E. van Beek[4,5], Fadlila Fitriani[1], Diana Wouters[4c], Taco W. Kuijpers[5,6], Davis Nwakanma[3], Umberto D'Alessandro[3], Eleanor M. Riley[7,8], Thomas D. Otto[9], Azra Ghani[10], Michael Levin[1], Lachlan J. Coin[2], David J. Conway[11], Michael T. Bretscher[†,10d], Aubrey J. Cunnington[†,1]*

[1]Section of Paediatrics, Imperial College, London, United Kingdom

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

[3]Medical Research Council Unit, Fajara, The Gambia at the London School of Hygiene and Tropical Medicine, Fajara, The Gambia.

[4]Department of Immunopathology, Sanquin Research and Landsteiner Laboratory, Amsterdam University Medical Centre, University of Amsterdam, Amsterdam, the Netherlands

[5]Department of Pediatric Hematology, Immunology and Infectious Diseases, Emma Children's Hospital, Amsterdam University Medical Centre, Amsterdam, the Netherlands

[6]Department of Blood Cell Research, Sanquin Research and Landsteiner Laboratory, Amsterdam University Medical Centre, University of Amsterdam, Amsterdam, the Netherlands

[7]The Roslin Institute and the Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, United Kingdom

[8]Department of Immunology and Infection, London School of Hygiene and Tropical Medicine, London, United Kingdom

[9]Centre of Immunobiology, Institute of Infection, Immunity & Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom

[10]MRC Centre for Global Infectious Disease Analysis, Imperial College, London, United Kingdom

[11]Department of Pathogen Molecular Biology, London School of Hygiene and Tropical Medicine, United Kingdom

*Corresponding author: a.cunnington@imperial.ac.uk (AJC)

[#,†] Denote equal contributions

[a]Current address: QIMR Berghofer Medical Research Institute, Brisbane, Australia

[b]Current address: Untere Grabenstraße 10, 88299 Leutkirch, Germany

[c]Current address: Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, the Netherlands

[d]MTB performed the work related to this publication at Imperial College and is now an employee of F. Hoffmann-La Roche Ltd.

Summary:

During infection, increasing pathogen load stimulates both protective and harmful aspects of the host response. The dynamics of this interaction are hard to quantify in humans, but doing so could improve understanding of mechanisms of disease and protection. We sought to model the contributions of parasite multiplication rate and host response to observed parasite load in individual subjects with *Plasmodium falciparum* malaria, using only data obtained at the time of clinical presentation, and then to identify their mechanistic correlates. We predicted higher parasite multiplication rates and lower host responsiveness in severe malaria cases, with severe anemia being more insidious than cerebral malaria. We predicted that parasite growth-inhibition was associated with platelet consumption, lower expression of *CXCL10* and type-1 interferon-associated genes, but increased cathepsin G and matrix metallopeptidase 9 expression. We found that cathepsin G and matrix metallopeptidase 9 directly inhibit parasite invasion into erythrocytes. Parasite multiplication rate was associated with host iron availability and higher complement factor H levels, lower expression of gametocyte-associated genes but higher expression of translation-associated genes in the parasite. Our findings demonstrate the potential of using explicit modelling of pathogen load dynamics to deepen understanding of host-pathogen interactions and identify mechanistic correlates of protection.

Introduction:

Improved methods are needed to identify mechanisms which protect against human infectious diseases in order to develop better vaccines and therapeutics[1,2]. Pathogen load is associated with the severity of many infections[3], and is a consequence of how fast the pathogen can replicate, how long the infection has been ongoing, and the inhibition or killing of pathogen by the host response (Fig. 1a). The contribution of these factors varies within an individual over the course of infection, as well as between individuals. Identifying mechanistic correlates of the processes which determine pathogen load is likely to be more informative than identifying correlates of pathogen load *per se*. However, in humans the timing of infection is rarely known and treatment cannot usually be withheld to observe the natural dynamics of pathogen load and host response. Here we present an approach to estimate the latent determinants of parasite load dynamics. We use these estimates to better understand severe malaria phenotypes and to identify mechanisms inhibiting parasite growth and controlling parasite multiplication during *Plasmodium falciparum* malaria in Gambian children.

Results:

Estimating determinants of parasite load and host response dynamics in humans

To estimate the determinants of parasite load dynamics in naturally-infected malaria patients we calibrated a statistical prediction model using outputs from a mechanistic simulation which combined information from two datasets. A historical dataset of the longitudinal course of untreated infection in 97 patients who were deliberately inoculated with *P. falciparum* as a treatment for neurosyphilis (malariatherapy dataset) (Supplementary Fig.1)

was used as a reference for changes in parasite load over time[4]. A dataset from 139 naturally infected Gambian children with malaria (Gambian dataset, Supplementary Table 1, Supplementary Dataset 1) was used for subsequent discovery of the determinants of parasite load dynamics. We used an existing mathematical model for the malariatherapy data (the Dietz model[4]), which estimated latent variables thought to determine changes in parasite load over time in each individual, and modified the model to better represent the features of the Gambian dataset. We used the modified model to simulate a large number of *in-silico* Gambian patients, with all latent variables and course of infection fully known, and then trained a statistical model to learn from these simulations the relationships between variables available in the real Gambian patient data and the unobservable, latent variables.

In the models[4], the increase in parasite load up to the first peak is determined by two individual-specific latent variables (Fig. 1b, see Methods): the within-host multiplication rate, m, which is the initial rate of increase in parasite load before any constraint by the host response; and the parasite load required to stimulate a host response that reduces parasite growth by 50%, P_c,[4]. When m, P_c, and parasite load are known, parasite growth inhibition (PGI) by the host response can be calculated (see Methods). We allowed rescaling of P_c values between the malariatherapy and Gambian datasets, and incorporated plasma Tumour Necrosis Factors (TNF) concentrations as an indicator of the protective host response[5,6], using a maximum-likelihood approach (see Methods and Supplementary Fig. 2). These modifications resulted in higher P_c values in the Gambian population than malariatherapy subjects, consistent with epidemiological data showing higher fever thresholds in *P. falciparum* infected children than in adults[7]. Other model assumptions and definitions are shown in Supplementary Table 2.

To test whether combination of a mechanistic simulation model with statistical learning of the relationships between latent and directly observable variables was better at predicting the determinants of parasite load than using observable variables alone, we simulated 2000 Gambian children with malaria with known values of m, P_c, parasite biomass, duration of illness and plasma TNF (Fig. 1c and Supplementary Fig. 3) and then fit general additive models (GAMs) to predict values of m and P_c for individual children (Supplementary Table 3). The resulting models produced more accurate predictions of of m and P_c than using individual variables alone (Fig. 1d).

Next we used the GAMs to predict values of P_c and m for each of the 139 individuals in the Gambian dataset (Fig. 1e-k, Supplementary Fig. 4). Children with the most severe manifestations of malaria (SM2) had the highest parasite load, TNF, predicted m, and predicted P_c values, intermediate values were seen in those with prostration as the only manifestation of severe disease (SM1), and values were lowest in uncomplicated malaria (UM), whilst duration of illness did not differ significantly by clinical phenotype (Fig. 1e-i). These observations suggest that high parasite load and severe disease are most likely in individuals with either fast replicating parasites (high m) or less immune responsiveness (high P_c).

Since age can be a major determinant of malaria severity and naturally acquired immunity[8], we examined whether age was associated with m or P_c. Age was not significantly correlated with m but was significantly negatively correlated with P_c (Fig. 1j,k). This implies little age-related acquisition of constitutive resistance (for example, naturally-acquired antibody-mediated immunity) in these children, as might be expected from the relatively low malaria transmission in this region of The Gambia[9]. However, these data also indicate that a lower

parasite load would be needed to provoke an equivalent host response in older individuals without significant naturally acquired immunity.

Predicting severe malaria phenotype from within-host dynamics

We next asked whether individual estimates of m and P_c could be used to predict pathophysiological features malaria which had not been used in our model derivation. Severe malarial anemia (hemoglobin concentration <5g/dL), is most common in the youngest children in high transmission settings, but rare in lower transmission settings such as Greater Banjul region of The Gambia, where cerebral malaria was relatively more common[10]. Severe malarial anemia is characterised by a higher parasite biomass[10-12], lower levels of both TNF and interleukin-10 (IL-10), but an elevated ratio of TNF:IL-10[13,14] when compared to cerebral malaria. In our Gambian subjects, hemoglobin concentration could be predicted from estimated P_c, m and age; IL-10 concentration could be predicted from m and P_c (Supplementary Table 4, Fig. 2a-b). We simulated a population of Gambian infants, selected those predicted to have hemoglobin <5 g/dL, and compared their characteristics to real Gambian subjects with cerebral malaria. The simulated severe anemia cases had lower m but similar P_c, higher parasite biomass and longer duration of illness than the cerebral malaria patients (Fig 2c-f). Both TNF and IL-10 concentrations were predicted to be lower in severe anemia than in cerebral malaria (Fig 2g-h), whereas the TNF:IL-10 ratio was predicted to be higher in severe anemia (Fig 2i), supporting the biological plausibility of relationships defined in our model and illuminating a potential explanation for these distinct severe malaria phenotypes.

Estimating parasite growth inhibition reveals the protective effect of platelets

The role of the host response in restricting parasite load is often unclear in human malaria because the strongest host responses are often seen in patients with the highest parasite loads and most severe disease[15,16]. For example platelets directly inhibit parasite growth[16,17], and the reduction in platelet count typically seen in malaria is partly a consequence of the protective mechanism of platelet adhesion to infected red cells[16]. However the reduction in platelet count is greatest in individuals with the highest parasite load and most severe disease[18], which seems counterintuitive if the low platelet counts indicate parasite killing. In Gambian children, estimated PGI did not differ significantly by clinical phenotype (Fig. 3a) indicating that the components of the host response which restrain parasite growth are similarly activated in severe and uncomplicated disease groups at the time of hospital presentation, but implying that this response developed too late to prevent high parasite load in the severe cases. Subjects with severe disease had the lowest platelet counts (Fig. 3b and Supplementary Table 1) and the highest parasite loads (Fig. 1d), but the protective role of platelets was evident through the significant (P=0.0001) correlation with PGI (Fig. 3c). Thus considering differences between individuals in observed parasite load and host response as part of a dynamic rather than static process can resolve counterintuitive associations.

Predicting mechanistic correlates of parasite growth inhibition

To determine whether our model-derived estimates could be used to discover aspects of host-parasite interaction we sought to identify mechanistic correlates of protection and susceptibility. We analysed human whole blood gene expression, with gene signature-based deconvolution to adjust for leukocyte-mixture[19], from samples of 24 children at the time of presentation (13 with UM, 11 with SM, Supplementary Table 5). Of 11702 detected human genes, 51 were significantly correlated (26 positively, 25 negatively) with estimated PGI after

adjustment for false discovery rate (Benjamini-Hochberg adjusted P<0.05, Fig. 4a, Supplementary Table 6). We reasoned that genes positively correlated with PGI should be enriched for effector mechanisms which act to reduce parasite load, whilst genes negatively correlated with PGI should be enriched for mechanisms which favour increase in parasite load. Eight of these genes were also correlated with parasite biomass and three with TNF (Supplementary Table 6).

Genes positively correlated with PGI (Fig 4a) showed limited canonical pathway enrichments (Supplementary Table 7) but 16 (62%) were linked together in a network around extracellular signal-regulated kinases ERK1/2 and AKT serine/threonine kinase (Fig. 4b). These kinases integrate cellular inflammatory and metabolic responses to control innate defence mechanisms such as cytokine secretion, phagocytosis and degranulation[20,21]. The 25 genes negatively correlated with PGI were strongly enriched in immune response pathways (Supplementary Table 7). Network analysis showed 15 (60%) of the negatively correlated genes were linked through a network focussed around type 1 interferon (Fig. 4c), consistent with observations that sustained type 1 interferon signalling is associated with higher parasitemia in mice[22-25] and potentially in humans[22,26]. C-X-C motif chemokine ligand 10 (*CXCL10*, also known as IFN-γ-inducible protein of 10 kDa, IP-10) expression had the greatest log-fold change of the genes negatively correlated with PGI (Fig. 4c), consistent with findings that CXCL10 deletion and neutralisation decrease parasite load in mice[27].

We investigated whether associations with PGI were dependent on assumptions we made about the true severity rate in Gambian children, which we assumed to be 5% based on published data in other settings[28,29]. Varying this to credible extremes of 1% and 10% and repeating the process of calibration between datasets, fitting of models to predict m and P_c, and estimating new values for PGI, resulted in little difference in the genes identified as

significantly associated with PGI, or the significance of individual genes (Supplementary Table 8).

Cathepsin G and MMP9 directly inhibit parasite growth

The 26 genes positively correlated with PGI have not, to our knowledge, previously been described as having anti-parasitic effects so we sought direct biological evidence, focussing on two encoding secreted proteins as the best candidates: *CTSG* (cathepsin G) and *MMP9* (matrix metallopeptidase 9, also known as matrix metalloproteinase 9 and gelatinase B), which both encode neutrophil granule proteins[30]. We tested whether these proteases could inhibit parasite growth *in vitro*. Cathepsin G and MMP9 both inhibited growth of *P. falciparum* 3D7 strain (Fig. 5a). Addition of cathepsin G to schizont cultures produced a dramatic reduction in invasion of new erythrocytes, and pretreatment of erythrocytes with cathepsin G before adding them to schizont cultures produced a similar reduction in their invasion (Fig. 5b), indicating that cathepsin G acts primarily on the erythrocyte. Addition of MMP9 to schizont cultures produced a more modest reduction, whilst pretreatment of erythrocytes did not reduce invasion, implying that MMP9 likely acts against schizonts or free merozoites rather than preventing invasion at the erythrocyte surface (Fig. 5b).

In order to identify biologically relevant concentrations of cathepsin G and MMP9 we measured their concentrations in whole blood from healthy donors, before and after stimulating degranulation, and in plasma from children with malaria at the time of clinical presentation (Fig. 5c). Local concentrations which might occur *in vivo*, adjacent to degranulating neutrophils, could be at least an order of magnitude higher[31]. MMP9 is also known to be released from other cell types in response to *P. falciparum*, including vascular endothelial cells[32]. MMP9 dose-dependently inhibited parasite growth over a physiological range of concentrations (Fig. 5d). Similarly, parasite invasion was dose-dependently inhibited

by cathepsin G pre-treatment of erythrocytes, with similar effects in each of four parasite strains with different invasion phenotypes[33] (Fig. 5e). Combined treatment with low doses of MMP9 and cathepsin G – in the range detected in patient plasma – showed an additive effect (Fig. 5f).

Cathepsin G has previously been reported to cleave red cell surface glycophorins[34], therefore we asked whether it might also cleave other RBC surface proteins which are used as invasion receptors by *P. falciparum*[35]. Consistent with its broad inhibition of parasite invasion, cathepsin G dose-dependently cleaved the majority of *P. falciparum* invasion receptors including glycophorins A, B, and C, CD147 (basigin), CD108 (semaphorin 7A), and complement receptor 1 (CR1), but not CD55 (DAF) (Fig. 5g). MMP9 did not cleave any of these surface receptors (Supplementary Fig. 5). PMA stimulation of healthy donor whole blood recapitulated the loss of erythrocyte surface glycophorins A and B, CD108 and CD147 in all donors, decreased glycophorin C expression in 6 of 8 healthy donors, but did not consistently reduce CR1 (Fig. 5h) (as might be expected from the dose-response curves, Fig 5g). In samples from Gambian children on the day of presentation with *P. falciparum* malaria, the proportions of erythrocytes with detectable expression of glycophorins A and B and CD147 were significantly lower than in convalescent samples (28 days after treatment), and there was a trend to lower expression of CD108 and glycophorin C (Fig. 5i). These results would be consistent with cleavage of these surface molecules *in vivo* during acute infection. The variable expression seen at day 28 (Fig. 5i) may indicate the persistence of modified erythrocytes in the circulation. The importance of glycophorins and basigin in RBC invasion and genetic susceptibility to severe malaria is well established[36-38], and so it is highly likely that the cleavage of these erythrocyte receptors by cathepsin G would have a protective effect *in vivo*.

Host and parasite factors associated with parasite multiplication rate

In our model, *m* is influenced by constitutive host and parasite factors but independent of any parasite load-dependent responses. We sought to confirm associations with two constitutive host factors known to influence parasite growth: iron[39] and complement factor H (FH)[40,41] (Supplementary Dataset 1). Since we did not have premorbid blood samples we used convalescent blood as a proxy for pre-infection status, with samples collected 28 days after treatment when the host response was quiescent (median C-reactive protein 1.1mcg/mL (IQR 0.5-5.1, n=70), similar to healthy West African population levels[42]).

Iron deficiency is protective against malaria[43] and reduces parasite multiplication *in vitro*[39]. Consistent with this, levels of hepcidin (a regulator of iron metabolism and marker of iron sufficiency or deficiency[44]) were significantly correlated with *m* (r_s=0.21, P=0.049) in 92 children who had not received blood transfusion.

FH is a constitutive negative regulator of complement activation which protects host cells from complement mediated lysis[45] but many pathogens including *P. falciparum* have evolved FH binding proteins to benefit from this protection[40,41]. FH protects blood-stage parasites from complement mediated killing *in vitro*[40,41] and higher plasma levels are associated with susceptibility and severity of malaria[46]. In the 14 children with residual day 28 plasma available, FH correlated with *m* (r_s=0.75, P=0.002), providing further support that the quantitative estimates from our model exhibit expected relationships with known determinants of parasite growth.

We investigated whether we could identify any parasite processes associated with *m*, through correlation with parasite gene expression. Of 3704 parasite genes detected by RNA-Seq, adjusted for developmental stage distribution[19], no individual genes passed the FDR adjusted

P-value threshold of <0.05. Therefore we used weighted gene correlation network analysis to reduce dimensionality[47], generating 17 modules of co-expressed parasite genes. Module eigengene values[19] of two modules correlated with *m* (unadjusted Spearman correlation *P*<0.05); their hub-genes were *PF3D7_1136000* (a conserved *Plasmodium* protein of unknown function) and *PF3D7_1238300* (putative pre-mRNA-splicing factor CWC22). The *PF3D7_1136000* module was negatively correlated (r_s=-0.5, *P*=0.01) with *m* and contained 140 genes with greatest gene ontology enrichment in microtubule-based movement (Supplementary Tables 9 & 10). The *PF3D7_1136000* module genes have high tolerance to insertional mutagenesis (Fig. 6a) and high parasite fitness following mutation (Fig. 6b), characteristics of winning mutants in competitive growth assays[48], supporting the concept that lower expression of these genes may favour more rapid growth. 77 (55%) of the genes in this module exhibit greatest expression during gametocyte development[49], consistent with the concept that increased sexual-stage commitment results in reduced asexual replication[50]. In contrast, the *PF3D7_1238300* module was positively correlated with *m* (r_s=0.46, *P*=0.03), and contained 45 genes enriched in translation functions (Supplementary Tables 9 & 10), plausible determinants of *m*, with mutagenesis tolerance typical of essential genes (Fig. 6a,b). Parasite genes differentially expressed between severe and uncomplicated malaria cases[19] were highly over-represented in this module (16 of 45 (36%), *P*=1.2x10^{-8}, Fisher exact test).

Discussion:

Using a model-based approach to estimate the within-host dynamics of pathogen load and its determinants in human infection we could estimate the relative contributions of parasite multiplication and host response to parasite load measured at a single point in time, and we have used these predictions to identify mechanistic determinants of parasite load in malaria. Our approach is based on clearly defined assumptions, but as with any attempt to model

complex biology, alternative approaches are possible. We cannot, at present, propagate uncertainty throughout the sequential stages of the model fitting, prediction of parameter estimates in individual subjects, and association of these parameter estimates with real variables. However, estimating the dynamics of parasite load allows us to make inferences about disease biology and mechanisms associated with PGI which could not have been made using only direct measurements. Our mechanistic validation suggests that the relative estimates of latent variables are accurate enough to be useful, providing proof-of -principle that pathogen load dynamics can be estimated in humans. This approach could be refined and expanded to identify additional genetic and serological determinants of pathogen load dynamics. The latter should be identified prospectively, since use of convalescent samples may introduce confounding.

Parasite load is only one of the factors associated with severe malaria and its interpretation is dependent on epidemiological context[10,15,29]. Variations in the host response, naturally acquired immunity, and the expression of *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) variants are also important determinants of severity and of disease phenotype[10,15]. We have previously suggested that variation in the dynamics of parasite load may explain why cerebral malaria and severe anaemia occur with parasites expressing the same PfEMP1 variants[10], and our model-based approach predicted that slower growth and longer duration of illness may distinguish severe anemia from cerebral malaria.

The importance of pathogen load and the dynamic nature of host-pathogen interactions are often omitted from studies of life-threatening infectious diseases in humans[3]. Much of our understanding of the host-pathogen interactions comes from comparisons between individuals at the point of clinical presentation, despite the fact that they may be at different stages in the dynamic process of infection. This can result in seemingly paradoxical

observations such as high levels of TNF or low levels of platelets associated with severe malaria[15,16], whilst evidence also indicates that TNF and platelets mediate defense against malaria parasites[5,6,15-17]. Considering the dynamic nature of the host-parasite interaction may explain these paradoxes and identify protective mechanisms more efficiently.

We identified several mechanisms which might be considered as prototypes for host-directed therapy in malaria. Inhibition of type-1 interferon or CXCL10 signalling with inhibitory antibodies or small molecules might be strategies to enhance control of parasite load. The therapeutic potentials of cathepsin G and MMP9 may be counterbalanced by risk of collateral tissue damage, but selective targeting of receptors on the erythrocyte surface may be a useful paradigm for both treatment and prevention of malaria.

Our approach could be applied to some other infectious diseases in which pathogen load can be measured and for which we do not have effective treatments, including emerging viral infections like Ebola, and possibly highly resistant bacterial pathogens, for which host-directed therapies may life-saving[2].

METHODS

Subjects and laboratory assays

We used data from all of the malariatherapy patients reported by Dietz et al.[4] and from all 139 Gambian subjects reported in our previous studies[11,51,52] who had all of the following data available: age, parasite biomass estimate, plasma TNF concentration, duration of illness and severity of illness. No subjects were excluded after this selection, and all available data was included in analyses, with the exception that one outlier was excluded from parasite gene expression analysis. As described previously[11,51,52], Gambian children (<16 years old) were recruited with parental consent from three peri-urban health centres in the Greater Banjul region, from August 2007 through January 2011 as part of a study approved by the Gambia Government/MRC Laboratories Joint Ethics Committee, and the Ethics Committee of the London School of Hygiene and Tropical Medicine. *P. falciparum* malaria was defined by compatible clinical symptoms in the presence of ≥5000 asexual parasites/μL blood, and any children suspected or proven to have bacterial co-infection were excluded. Severe malaria was specifically defined by the presence of prostration (SM1) or any combination of three potentially overlapping syndromes (cerebral malaria (CM), severe anemia (SA, hemoglobin <5 g/dL), and hyperlactatemia (blood lactate >5 mmol/L) - collectively SM2)[11,51-53]. Clinical laboratory assays, measurements of plasma TNF and IL-10 by Luminex, measurements of gene expression by RT-PCR, and estimation of total parasite biomass from *Pf*HRP2 ELISA have been previously described[11,52]. Subject-level data from these Gambian children is available as **Supplementary Dataset 1**.

Statistical analyses

Statistical analyses were undertaken using the R statistical software[54] and GraphPad Prism (GraphPad Software, Inc.). Directly measured continuous variables were compared between groups using unpaired or paired student's t-test (when normally distributed) and the Mann-Whitney or Wilcoxon matched pairs tests (when normal distribution could not be assumed), and ANOVA or Kruskal-Wallis test for comparison across multiple groups. Associations between measured variables and latent variables were assessed using generalised additive models (GAM[55], with the R package "mgcv"); the generalised cross-validation score and explained variance were used to select the best GAM once all model terms had significant effects ($P<0.05$). It was not possible to propagate uncertainty estimates through all stages from model development, calibration to the Gambian data, and prediction of latent variables in individual subjects, and so statistical analyses of latent variable were undertaken using their predicted values without any measure of uncertainty, and using non-parametric methods. Correlations between predicted values of latent variables and measured variables were done using Spearman correlation.

All hypothesis tests were two-sided with alpha = 0.05 unless specifically stated otherwise. One-sided testing was only used when justified by small sample size and a strong *a priori* hypothesis for the direction of effect. We did not adjust for multiple hypothesis testing, except in the case of gene expression analyses where false-discovery rate was controlled using the Benjamini-Hochberg method. Dose-response curves were fitted using asymmetrical sigmoidal five-parameter logistic equation in GraphPad Prism.

Model relating parasite multiplication, host response and parasite load

A process-based, stochastic simulation model was devised to reproduce the clinical data collected from the Gambian children. This was achieved by combining the information in the Gambian data with a model describing the first wave of parasitemia in non-immune adults who were deliberately infected with *P. falciparum* malaria to treat neurosyphilis ("malariatherapy")[4]. These malariatherapy data, from the pre-antibiotic era, are the main source of information on the within-host dynamics and between-host variation in the course of parasitemia in untreated malaria infections. The model of Dietz et al.[4] was modified and extended in order to be applied to the Gambian data, and we made the assumption that the Gambian children presented to hospital prior to the first peak of parasitemia.

Model of ascending parasitemia in malariatherapy subjects. The model relates parasite density after each 2-day asexual blood stage cycle ($P_{(t+2)}$) to the parasite density at the end of the previous cycle ($P_{(t)}$) by the following equation:

$$P_{(t+2)} = P_{(t)}.m.S_{c(t)}$$

The host-specific parasite multiplication rate, *m*, is a property of both parasite and host, allowing for growth-inhibition by constitutive factors; the proportion of parasites that will survive the effects of the density-dependent host response in the present cycle is S_c:

$$S_{c(t)} = \frac{1}{1 + \left(\frac{P_{(t)}}{P_c}\right)}$$

, where P_c is the host-specific parasite load threshold at which the host response is strong enough to inhibit 50% of parasite growth in that cycle. Parasite growth inhibition ($PGI_{(t)}$) is defined as 1-$S_{c(t)}$.

Consistent with the original Dietz model, $P_{(0)}$ was set to 0.003 parasites/µl[4].

The original Dietz model included an additional parameter, S_m, to help describe the decline in parasitemia after the peak of the first wave. S_m is the proportion of isogenic parasites surviving an additional density- and time-dependent host response, which might represent adaptive immunity (*4*). Estimates of the range of values of S_m in the Dietz dataset and model were used when simulating data but since this parameter has little influence on parasite densities prior to the peak it was not used to make subsequent predictions of m and P_c in individual Gambian subjects.

At the explicit request of Klaus Dietz and Louis Molineaux, we hereby communicate the following correction regarding their assertion that the malariatherapy patients had not received any treatment[4]: it was later found that 47 of these patients had indeed received subcurative treatment, and that those patients had significantly higher parasite densities. This is unlikely to influence our analysis, because treatment would only be provided when malariatherapy patients became very unwell, presumably at maximum parasitemia, whereas we assume that most patients with naturally acquired infection likely present prior to the peak parasitemia that might occur in the absence of treatment.

Fitting of the malariatherapy model to data from Gambian children. Individual-level parameter estimates for the malariatherapy dataset were kindly provided by Klaus Dietz. The logarithms of these 97 estimates of m and P_c were well described by a multivariate normal distribution, providing a quantitative description of inter-individual variation in the dynamics

of the first wave of parasitemia. In order to use the Dietz model to simulate the Gambian data, a number of modifications and extensions were made. Some of these required estimation of additional parameters by comparing the model simulations with the Gambian data. Dietz et al. provided a statistical description of the parasite density at which first fever occurred (the "fever threshold") in the form of the distribution of the ratio of threshold density to peak parasitemia. The median density at first fever was at 1.4% of peak density. We introduced the assumption that the onset of fever occurs at a particular threshold value of S_c, because fever is dependent on the production of cytokines like interleukin-6 and TNF, both components of the host response. This constitutes a process-based model for the onset of fever rather than a purely statistical one. Because individuals differ in their response to parasite load (captured through variation in P_c), this results in variation of parasite densities at first fever but ignores any potential variation among individuals with respect to magnitude of host response necessary to generate fever. The host response threshold for the onset of fever $S_c^f = 0.86$ was determined as the value of S_c calculated at 1.4% of the peak density of a simulated individual with the median parameter values. This yielded a distribution of fever ratios similar to the one described by Dietz et al.[4], albeit with less variation.

To simulate the time between onset of fever and clinical presentation we made use of the self-reported duration of symptoms in the Gambian data. The model which was most consistent with these values assumed a gamma-distributed duration of symptoms in non-severe cases, and a possibility to present earlier in the case of more severe disease. Since parasite biomass is related to likelihood of having severe malaria[11,12,56] the probability of early presentation on any day after onset of fever was set proportional to the (density-dependent) probability of having severe disease on that day. Scale (ζ) and shape (κ) parameters of the gamma distribution as well as the factor (ξ) for determining the probability of early presentation were estimated from the Gambian data.

We assumed that TNF production $\tau_{(t)}$ increases monotonically with density dependent host response (1-S_c) and represented this relationship using a heuristic function of the form

$$\tau_{(t)} = a + b\left(1 - \frac{1}{1 + \left(\frac{-\log\left(S_{c(t)}\right)}{\lambda^*}\right)^{\gamma}}\right)$$

, with free parameters a, b, λ^* and γ estimated from the Gambian data.

The Gambian children had on average higher parasite densities than the malariatherapy patients, which led to a bad fit of the original model to the Gambian data. This was overcome by introducing the assumption that the Gambian children had a different range of values of P_c to the adult malariatherapy patients. A factor π was therefore estimated by which the ln P_c value from the Dietz model was multiplied. This led to overall higher parasite densities upon presentation. However, our model uses parasite biomass and its relationship with disease severity to predict day of presentation, and there is an interaction between the mean ln P_c and the variation in ln P_c, as well as the proportion of severe malaria in the simulated Gambian population. Based on the relatively low malaria transmission in the Banjul area of The Gambia, we assumed that severe cases (defined by the presence of any of: prostration, hyperlactatemia, severe anemia or cerebral malaria) were over-represented by hospital-based recruitment and that in an unselected population of children of similar age to those in our dataset only approximately 5% of all malaria infections would be severe[28,29]. Therefore we estimated a factor δ by which the variance of ln P_c should be multiplied such that both rate of severity as well as the distribution of parasite biomass matched well after fitting our simulation to the Gambian data.

The free parameters ζ , κ, ξ, a, b, λ^*, γ, π and δ (Supplementary table 11), together summarized as θ, were estimated by fitting model simulations to the information on TNF, parasite density, and duration of symptoms, for any given candidate parameterization, a total of 139 clinically presenting individuals were simulated from the model, which corresponds to the size of the Gambian dataset. An objective function $L(\theta)$ was calculated, and a simulated annealing algorithm (provided by the "optim" function in R) determined the value for θ which maximizes this function. The log-likelihood $L(\theta)$ was comprised of three separate objectives. The first objective represented the log-probability that the frequency of severe cases in the simulation was equal to an assumed 5%, employing a binomial likelihood, given the actual number of severe cases sampled in 139 simulated individuals. The second objective considered the overlap between the bivariate distribution of ln parasite density vs. ln TNF concentration in the simulated data compared to the Gambian dataset. An approximate numerical value for this partial log-likelihood was obtained as the log probability of the Gambian data (density and TNF) given a two-dimensional kernel density estimate of the simulation output as a likelihood model. Kernel density estimates were obtained using the "kde2d" function in the "MASS" package in R. In this calculation, the TNF/density data points of severe or prostrated Gambian patients entered the partial likelihood with a weight of 1/11, to account for the oversampling of severe cases in the Gambian data. The third objective concerned the two-dimensional distribution of log density and duration since first fever. This partial log-likelihood was obtained using the same kernel-based approach described above, with weights of 1/11 for severe and prostrated cases. The overall log-likelihood $L(\theta)$ was calculated as a weighted sum of the three partial log-likelihoods, with the log-probability of having the desired true severity rate weighted with a factor of 68, which ensured similar magnitude of the three partial log-likelihoods at the optimum.

The results of the fitting algorithm were visually confirmed to yield a good overlap of the joint distributions of density and biomass, the duration of symptoms, TNF and biomass between simulation and the Gambian children. Approximate confidence intervals for the parameter estimates were determined by employing a 2nd degree polynomial to estimate the curvature of the maximum simulated likelihood surface in the vicinity of the parameter point estimate, assuming independence of parameters.

As in the original model of Dietz et al.[4], peripheral parasite densities were used to determine the dynamic changes in parasitemia, implying a correlation between peripheral densities and total parasite biomass. Total parasite biomass per kg was calculated from the predicted parasite density by the equation 70,000 x 1.09 x predicted parasite density in parasites/µL, as has been determined previously for uncomplicated malaria cases in the Gambian dataset[11].

Deterministic relationships between observable and latent variables. The range of values of m and ln P_c in a simulated population of 2000 patients were determined and each divided into 50 equally spaced increments in order to generate 2500 possible combinations of m and ln P_c for which all model outcomes were determined in order to visualize their relationships. For the purpose of this analysis, the time-dependent adaptive immune response parameters (which comprise S_m) were set for all subjects at their respective population median values. The model of Dietz *et al*. makes use of discrete 2 day time intervals[4], corresponding to the duration of the intraerythrocytic cycle in a highly synchronised infection. However, naturally acquired infections are rarely this synchronous and the time since infection of our Gambian patients is an unknown continuous variable. In order to cope with this we assumed that the relationship between predicted outcome variables (parasite biomass, duration of illness and TNF concentration) and explanatory variables (m and P_c) could be approximated by smoothed GAM. We used the GAM to estimate values of m, P_c and parasite growth

inhibition (PGI, 1-S_c) in the Gambian children, based on their known total parasite biomass, duration of symptoms and TNF concentration.

Predicting severe anemia and IL-10 concentrations

We used the data from the Gambian children to predict hemoglobin and IL-10 concentrations as continuous variables, using GLM with predicted P_c, predicted m, and age as explanatory variables. We then simulated a population of 50,000 1-year olds with malaria, allowing for normal variation in baseline hemoglobin concentration[57], and adjusting P_c values according to a linear relationship between predicted ln P_c and age in the Gambian children. To predict the occurrence of severe anemia, we calculated the proportion of subjects estimated to have hemoglobin <5g/dL, and for these we calculated IL-10 concentrations as a continuous outcome.

RNA-sequencing and data analysis

We used RNA-sequencing data from all 24 subjects who were included in our previously reported study[19] and had data to allow estimation of parasite growth inhibition and multiplication rate. RNA extraction, library preparation, sequencing and downstream analysis, including adjustment for leukocyte and parasite developmental stage mixture, have all been previously described[19].

The association of gene expression with m and PGI was determined using a generalized linear model approach in edgeR, allowing adjustment for leukocyte and parasite developmental stage mixture. Coefficients and P-values were calculated for the relationships between adjusted log gene expression and PGI for all detected genes. False discovery rate (FDR) was then computed using the Benjamini-Hochberg approach and FDR below 0.05 was considered to be significant in the initial analysis. FDRs between 0 and 0.1 were considered to indicate

consistent findings when comparing associations obtained under different model assumptions. Gene ontology (GO) terms were obtained from Bioconductor packages "org.Hs.eg.db" and "org.Pf.plasmo.db". Fisher's exact test was used to identify significantly over-represented GO terms from gene lists. The background gene sets consisted of all expressed genes detected in the data set. Enrichment analysis for biological process terms was carried out using the "goana()" function in edgeR. Using groups of genes significantly positively or negatively correlated with PGI, Ingenuity Pathway Analysis (Qiagen) was used to identify networks of genes functionally linked by regulators, interactions or downstream effects, which were visualized as radial plots centered around the most connected network member. The weighted gene co-expression network analysis (WGCNA) tool[47] was used to construct modules of highly co-expressed parasite genes, based on analysis of 23 samples (sample HL_478 was removed as an outlier in parasite RNA-seq analysis) as described previously[19]. Module eigengene values for each subject were correlated with predicted *m*, using Spearman correlation.

Parasite culture, growth and invasion assays

P. falciparum 3D7 strain was used in continuous culture for all of the experiments unless otherwise stated. Asexual blood stage parasites were cultured in human blood group A red cells, obtained from the National Blood Service, at 1-5% hematocrit, 37°C, 5% CO_2 and low oxygen (1% or 5%) as described previously[58,59]. Growth medium was RPMI-1640 (without L-glutamine, with HEPES) (Sigma) supplemented with 5 g/L Albumax II (Invitrogen), 147 µM hypoxanthine, 2 mM L-glutamine, and 10 mM D-glucose. Parasite developmental stage synchronization was performed using 5% D-sorbitol to obtain ring stage parasites or Percoll gradients for schizont stage enrichment[58,60]. For growth assays, schizonts were mixed at <1% parasitemia with uninfected erythrocytes at 2% final hematocrit. Cathepsin G (Abcam) or

recombinant active MMP9 (Enzo) were added for 72 hour incubation to allow two replication cycles. Growth under each condition was calculated relative to the average growth in untreated samples. Invasion assays were performed by adding parasites synchronised at the schizont stage to target erythrocytes and incubating for 24 hours. Cathepsin G and MMP9 were either pre-incubated with the target cells overnight followed by four washes with RPMI to completely remove them, or they were added directly to the culture of schizonts with target erythrocytes for 24 hours. The same protocol was followed for other *P. falciparum* strains except Dd2, for which magnetic purification was used to purify schizonts[61]. For combined treatments, cathepsin G was added to target erythrocytes and MMP9 was added at the same time as schizonts.

Flow cytometry for parasitemia and erythrocyte surface receptor expression

Flow cytometry was performed using a BD LSR Fortessa machine and analysis was conducted using FlowJo v10 (TreeStar Inc.), and gating strategies are show in Supplementary Figure 5. To assess parasitemia, 1µl of sample at 50% hematocrit was stained with Hoechst 33342 (Sigma) and dihydroethidium (Sigma) and then fixed with 2% paraformaldehyde (PFA) before flow cytometry as previously described[62]. Erythrocyte surface receptor expression was assessed by median fluorescence intensity of erythrocytes labelled with monoclonal antibodies or by comparison with isotype control antibodies (Supplementary Table 12). Briefly, erythrocytes were washed twice before resuspending at 50% haematocrit, of which 1-2µl was stained in 100µl of antibody cocktail in FACS buffer (2% fetal bovine serum, 0.01% sodium azide in PBS) for 30 minutes in the dark on ice. Samples were washed twice in FACS buffer and then fixed in 300µl FACS buffer with 2% paraformaldehyde. Surface receptor loss was calculated from the difference between the treated and untreated

sample median fluorescent intensities after the isotype control antibody fluorescence had been subtracted.

Whole blood stimulation and Cathepsin G and MMP9 ELISA

Whole blood was collected from 8 healthy adult donors and plated at 25% hematocrit, and incubated overnight with or without 1µM PMA (Sigma). Supernatant was collected to perform Cathepsin G (CTSG ELISA Kit-Human, Aviva Systems Biology) and MMP9 (Legend Max Human MMP-9, Biolegend) ELISAs, and erythrocytes were collected for assessment of surface receptor expression. The same ELISA kits were used to measure cathepsin G and MMP9 in acute (day 0) plasma samples from children with malaria.

C-reactive protein, Hepcidin, and complement Factor H ELISA

Using plasma samples collected 28 days after infection, CRP was measured using the human Simple Step ELISA kit (Abcam) and hepcidin concentration was measured in subjects who had not received blood transfusion using the Hepcidin-25 bioactive ELISA kit (DRG), both according to the manufacturer's instructions, with duplicate measurements when sufficient plasma was available. Complement Factor H assays were performed using an in-house ELISA as described[63].

Data availability

Estimates of parameters determining within-host dynamics in the malariatherapy dataset were obtained from reference 4, whose corresponding author may be contacted at klaus.dietz@uni-tuebingen.de. RNA-seq data have been deposited in the ArrayExpress database at EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-6413. Individual subject-

level data is available within the paper and its supplementary information files. All other data that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability

The source code for the model simulating Gambian child subjects and examples of its use are presented as Supplementary Library File, Supplementary Example File.

FIGURE LEGENDS

Fig 1. Estimating the dynamics of parasite load and host response in malaria.

(**a**) Conceptual model of determinants of parasite load. (**b**) Schematic of relationships between parasite load, multiplication rate (*m*), P_c, and parasite growth inhibition (PGI) derived from the longitudinal malariatherapy dataset. (**c**) Correlation matrix for P_c, *m*, parasite biomass, duration of illness and TNF concentrations in 2000 simulated Gambian children (Spearman correlation, LOWESS fit lines). (**d**) Performance in simulated subjects of the best models to predict ln P_c and *m*, compared with predictions made using individual variables only. Boxes show median and interquartile range, whiskers extend 1.5-times the interquartile range or to limit of range, n=100 simulated datasets (each of 139 subjects). (**e-i**) Comparisons of parasite biomass (**e**), TNF (**f**), duration of illness (**g**), predicted *m* (**h**), predicted P_c (**i**), in 139 Gambian children with uncomplicated (UM, *n*=64) or severe malaria (SM1, prostration, *n*=36; SM2, any combination of cerebral malaria, hyperlactatemia or severe anemia, *n*=39). Box and whiskers as in **d**; *P* for Kruskal-Wallis (above plots) and Mann-Whitney tests (UM vs SM2, within plot). (**j, k**) Correlation of predicted *m* (**j**) or P_c (**k**) with age, *P* for Spearman correlation, n=139.

Fig 2. Contribution of parasite load dynamics to severe malaria phenotype. (**a**, **b**) Comparison of predicted and actual hemoglobin (**a**, n=136) and IL-10 (**b,** n=139) concentrations in the Gambian children. Pearson correlation, shaded region, 95% CI of regression line. (**c-i**) Comparisons of *m*, P_c, parasite biomass, days of illness, plasma TNF, plasma IL-10, and plasma TNF:IL-10 ratio, in Gambian children with cerebral malaria (CM, n=12) and simulated Gambian infants with severe anemia (SA, n=24). Boxes show median and interquartile range, whiskers extend 1.5-times the interquartile range or to limit of range.

Fig 3. The protective effect of platelets is revealed by estimating parasite growth inhibition.

(**a,b**) Comparisons of PGI (**a**) and platelet count (**b**) in 139 Gambian children with uncomplicated (UM, *n*=64) or severe malaria (SM1, prostration, *n*=36; SM2, any combination of cerebral malaria, hyperlactatemia or severe anemia, *n*=39 (platelet data missing from 4 subjects)). (**c**) Correlation between platelet count and PGI (n=135) shows that low platelet count is associated with greater parasite growth inhibition. Boxes show median and interquartile range, whiskers extend 1.5-times the interquartile range or to limit of range; *P* for Kruskal-Wallis (above plots) test (**a**, **b**) and for Spearman correlation (**c**).

Fig. 4 Transcriptional correlates of parasite growth inhibition

(**a**) Volcano plot showing association between gene expression and parasite growth inhibition after adjustment for leukocyte mixture in a linear model. Log fold change (log FC) is the coefficient of log adjusted gene expression vs. parasite growth inhibition. Positive log FC indicates that increasing gene expression is associated with increasing parasite growth inhibition. Negative log FC indicates that increasing gene expression is associated with decreasing parasite growth inhibition. *P* calculated using two-sided likelihood ratio test, adjusted for multiple testing using the Benjamini-Hochberg method: false discovery rate adjusted $P < 0.05$ (FDR) is considered significant (above dashed line, colored circles). The 10 significant genes with greatest positive and negative log FC are labelled. Analyses based on data from n=24 subjects. (**b,c**) Primary networks derived from the genes significantly associated with PGI, with positive (**b,** n=26) and negative (**c,** n=25) log FC.

Fig. 5 Effects of cathepsin G and MMP9 on parasite growth and expression of erythrocyte invasion receptors

(**a**) Effect of cathepsin G (18µg/mL, n=5) and MMP9 (16µg/mL, n=3) or no treatment (n=8) on *in vitro* growth of *P. falciparum* 3D7 (n are biological replicates, results representative of two independent experiments). (**b**) Effect of cathepsin G (18µg/mL) and MMP9 (18µg/mL) on erythrocyte invasion of *P. falciparum* 3D7 when added directly to schizonts and donor red cells, or when pre-incubated (PT) with donor red cells before washing and adding to schizonts (n=3 biological replicates per condition, representative of two independent experiments). (**a, b**) Show mean (95% CI) and *P* for two-sided unpaired t-test. (**c**) Cathepsin G and MMP9 concentrations in plasma from healthy donor whole blood (n=8) unstimulated or stimulated with 1µM PMA, and from Gambian children with *P. falciparum* malaria (n=34). Bars show median, *P* for two-sided Wilcoxon matched pairs test. (**d-e**) Dose effects on growth inhibition by MMP9 against *P. falciparum* 3D7 (**d**), and invasion inhibition by cathepsin G pre-treatment against four parasite strains (**e)** (n=3 biological replicates per dose, mean (95% CI) and *P* for linear trend, each result representative of two independent experiments). (**f**) Additive effect of Cathepsin G 1µg/mL and MMP9 1µg/mL against *P. falciparum* 3D7 invasion (n=4 biological replicates per condition, mean (95% CI) and *P* for ANOVA, representative of three independent experiments). (**g**) Dose response for erythrocyte surface receptor cleavage by cathepsin G (n=3 biological replicates per dose, mean +/- standard error, asymmetric 5-parameter logistic regression fit lines, representative of two experiments). (**h**) Effect of PMA stimulation of healthy donor (n=8) whole blood on erythrocyte surface receptor expression assessed by fluorescence intensity (*P* for two-sided Wilcoxon matched pairs test). (**i**) Comparison of proportion of erythrocytes with detectable receptor expression in acute (day 0) and convalescent (day 28) samples from Gambian children with malaria (n=6, *P* for one-sided Wilcoxon matched pairs test).

Figure 6. Parasite gene expression modules associated with predicted multiplication rate. (a,b) Violin plots showing comparison of mutation insertion scores (**a**) and mutation fitness scores (**b**) between modules associated with predicted multiplication rate (*PF3D7_1136000*, n=138 genes; *PF3D7_1238300* n=42 genes) and all other genes (n=3421). (Violin plots indicate distribution of data (kernel density estimates) and median (red circle); *P* for comparison between each module and all other genes using a two-sided Mann-Whitney test).

References:

1 Plotkin, S. A. Complex correlates of protection after vaccination. *Clin Infect Dis* **56**, 1458-1465 (2013).

2 Zumla, A. *et al.* Host-directed therapies for infectious diseases: current status, recent progress, and future prospects. *Lancet Infect Dis* **16**, e47-63 (2016).

3 Cunnington, A. J. The importance of pathogen load. *PLoS Pathog* **11**, e1004563 (2015).

4 Dietz, K., Raddatz, G. & Molineaux, L. Mathematical model of the first wave of *Plasmodium falciparum* asexual parasitemia in non-immune and vaccinated individuals. *Am J Trop Med Hyg* **75**, 46-55 (2006).

5 Kumaratilake, L. M. *et al.* A synthetic tumor necrosis factor-alpha agonist peptide enhances human polymorphonuclear leukocyte-mediated killing of *Plasmodium falciparum in vitro* and suppresses *Plasmodium chabaudi* infection in mice. *J Clin Invest* **95**, 2315-2323 (1995).

6 Kumaratilake, L. M., Ferrante, A. & Rzepczyk, C. M. Tumor necrosis factor enhances neutrophil-mediated killing of *Plasmodium falciparum*. *Infect Immun* **58**, 788-793 (1990).

7 Smith, T. *et al.* Relationships between *Plasmodium falciparum* infection and morbidity in a highly endemic area. *Parasitology* **109 (Pt 5)**, 539-549 (1994).

8 Crompton, P. D. *et al.* Malaria immunity in man and mosquito: insights into unsolved mysteries of a deadly infectious disease. *Annu Rev Immunol* **32**, 157-187 (2014).

9 Okebe, J. *et al.* School-based countrywide seroprevalence survey reveals spatial heterogeneity in malaria transmission in the Gambia. *PLoS One* **9**, e110926 (2014).

10 Cunnington, A. J., Walther, M. & Riley, E. M. Piecing together the puzzle of severe malaria. *Sci Transl Med* **5**, 211ps218 (2013).

11 Cunnington, A. J., Bretscher, M. T., Nogaro, S. I., Riley, E. M. & Walther, M. Comparison of parasite sequestration in uncomplicated and severe childhood *Plasmodium falciparum* malaria. *J Infect* **67**, 220-230 (2013).

12 Hendriksen, I. C. *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* **207**, 351-361 (2013).

13 Kurtzhals, J. A. *et al.* Low plasma concentrations of interleukin 10 in severe malarial anaemia compared with cerebral and uncomplicated malaria. *Lancet* **351**, 1768-1772 (1998).

14 Perkins, D. J. *et al.* Severe malarial anemia: innate immunity and pathogenesis. *Int J Biol Sci* **7**, 1427-1442 (2011).

15 Cunnington, A. J., Riley, E. M. & Walther, M. Stuck in a rut? Reconsidering the role of parasite sequestration in severe malaria syndromes. *Trends Parasitol* **29**, 585-592 (2013).

16 Kho, S. *et al.* Platelets kill circulating parasites of all major *Plasmodium* species in human malaria. *Blood* **132**, 1332-1344 (2018).

17 McMorran, B. J. *et al.* Platelets kill intraerythrocytic malarial parasites and mediate survival to infection. *Science* **323**, 797-800 (2009).

18 Cserti-Gazdewich, C. M. *et al.* Inter-relationships of cardinal features and outcomes of symptomatic pediatric *Plasmodium falciparum* malaria in 1,933 children in Kampala, Uganda. *Am J Trop Med Hyg* **88**, 747-756 (2013).

19 Lee, H. J. *et al.* Integrated pathogen load and dual transcriptome analysis of systemic host-pathogen interactions in severe malaria. *Sci Transl Med* **10,** eaar3619 (2018).

20 Arthur, J. S. & Ley, S. C. Mitogen-activated protein kinases in innate immunity. *Nat Rev Immunol* **13**, 679-692 (2013).

21 Manning, B. D. & Toker, A. AKT/PKB Signaling: Navigating the Network. *Cell* **169**, 381-405 (2017).

22 Spaulding, E. *et al.* STING-Licensed Macrophages Prime Type I IFN Production by Plasmacytoid Dendritic Cells in the Bone Marrow during Severe *Plasmodium yoelii* Malaria. *PLoS Pathog* **12**, e1005975 (2016).

23 Zander, R. A. *et al.* Type I Interferons Induce T Regulatory 1 Responses and Restrict Humoral Immunity during Experimental Malaria. *PLoS Pathog* **12**, e1005945 (2016).

24 Haque, A. *et al.* Type I IFN signaling in CD8- DCs impairs Th1-dependent malaria immunity. *J Clin Invest* **124**, 2483-2496 (2014).

25 Haque, A. *et al.* Type I interferons suppress CD4(+) T-cell-dependent parasite control during blood-stage *Plasmodium* infection. *Eur J Immunol* **41**, 2688-2698 (2011).

26 Montes de Oca, M. *et al.* Type I Interferons Regulate Immune Responses in Humans with Blood-Stage *Plasmodium falciparum* Infection. *Cell Rep* **17**, 399-412 (2016).

27 Ioannidis, L. J. *et al.* Monocyte- and Neutrophil-Derived CXCL10 Impairs Efficient Control of Blood-Stage Malaria Infection and Promotes Severe Disease. *J Immunol* **196**, 1227-1238 (2016).

28 Griffin, J. T. *et al.* Gradual acquisition of immunity to severe malaria with increasing exposure. *Proc Biol Sci* **282** (2015).

29 Goncalves, B. P. *et al.* Parasite burden and severity of malaria in Tanzanian children. *N Engl J Med* **370**, 1799-1808 (2014).

30 Cowland, J. B. & Borregaard, N. Granulopoiesis and granules of human neutrophils. *Immunol Rev* **273**, 11-28 (2016).

31 Weksler, B. B., Jaffe, E. A., Brower, M. S. & Cole, O. F. Human leukocyte cathepsin G and elastase specifically suppress thrombin-induced prostacyclin production in human endothelial cells. *Blood* **74**, 1627-1634 (1989).

32 D'Alessandro, S., Basilico, N. & Prato, M. Effects of *Plasmodium falciparum*-infected erythrocytes on matrix metalloproteinase-9 regulation in human microvascular endothelial cells. *Asian Pac J Trop Med* **6**, 195-199 (2013).

33 Binks, R. H. & Conway, D. J. The major allelic dimorphisms in four *Plasmodium falciparum* merozoite proteins are not associated with alternative pathways of erythrocyte invasion. *Mol Biochem Parasitol* **103**, 123-127 (1999).

34 Bykowska, K., Duk, M., Kusnierz-Alejska, G., Kopec, M. & Lisowska, E. Degradation of human erythrocyte surface components by human neutrophil elastase and cathepsin G: preferential digestion of glycophorins. *Br J Haematol* **84**, 736-742 (1993).

35 Satchwell, T. J. Erythrocyte invasion receptors for *Plasmodium falciparum*: new and old. *Transfus Med* **26**, 77-88 (2016).

36 Crosnier, C. *et al.* Basigin is a receptor essential for erythrocyte invasion by *Plasmodium falciparum*. *Nature* **480**, 534-537 (2011).

37 Leffler, E. M. *et al.* Resistance to malaria through structural variation of red blood cell invasion receptors. *Science* (2017).

38 Pasvol, G., Wainscoat, J. S. & Weatherall, D. J. Erythrocytes deficiency in glycophorin resist invasion by the malarial parasite *Plasmodium falciparum*. *Nature* **297**, 64-66 (1982).

39 Clark, M. A. *et al.* Host iron status and iron supplementation mediate susceptibility to erythrocytic stage *Plasmodium falciparum*. *Nat Commun* **5**, 4446 (2014).

40 Rosa, T. F. *et al.* The *Plasmodium falciparum* blood stages acquire factor H family proteins to evade destruction by human complement. *Cell Microbiol* **18**, 573-590 (2016).

41 Kennedy, A. T. *et al.* Recruitment of Factor H as a Novel Complement Evasion Strategy for Blood-Stage *Plasmodium falciparum* Infection. *J Immunol* **196**, 1239-1248 (2016).

42 Eriksson, U. K., van Bodegom, D., May, L., Boef, A. G. & Westendorp, R. G. Low C-reactive protein levels in a traditional West-African population living in a malaria endemic area. *PLoS One* **8**, e70076 (2013).

43 Gwamaka, M. *et al.* Iron deficiency protects against severe *Plasmodium falciparum* malaria and death in young children. *Clin Infect Dis* **54**, 1137-1144 (2012).

44 Pasricha, S. R. *et al.* Expression of the iron hormone hepcidin distinguishes different types of anemia in African children. *Sci Transl Med* **6**, 235re233 (2014).

45 Parente, R., Clark, S. J., Inforzato, A. & Day, A. J. Complement factor H in host defense and immune evasion. *Cell Mol Life Sci* **74**, 1605-1624 (2017).

46 van Beek, A. E. *et al.* Complement Factor H Levels Associate With *Plasmodium falciparum* Malaria Susceptibility and Severity. *Open Forum Infect Dis* **5**, ofy166 (2018).

47 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

48 Zhang, M. *et al.* Uncovering the essential genes of the human malaria parasite *Plasmodium falciparum* by saturation mutagenesis. *Science* **360** (2018).

49 Lopez-Barragan, M. J. *et al.* Directional gene expression and antisense transcripts in sexual and asexual stages of Plasmodium falciparum. *BMC Genomics* **12**, 587 (2011).

50 Josling, G. A. & Llinas, M. Sexual development in *Plasmodium* parasites: knowing when it's time to commit. *Nat Rev Microbiol* **13**, 573-587 (2015).

51 Walther, M. *et al.* HMOX1 gene promoter alleles and high HO-1 levels are associated with severe malaria in Gambian children. *PLoS Pathog* **8**, e1002579 (2012).

52 Walther, M. *et al.* Distinct roles for FOXP3 and FOXP3 CD4 T cells in regulating cellular immunity to uncomplicated and severe *Plasmodium falciparum* malaria. *PLoS Pathog* **5**, e1000364 (2009).

53 Marsh, K. *et al.* Indicators of life-threatening malaria in African children. *N Engl J Med* **332**, 1399-1404 (1995).

54 R Core Development Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, <URL http://www.R-project.org/> (2014).

55 Wood, S. N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J R Stat Soc Series B Stat Methodol* **73**, 3-36 (2011).

56 Dondorp, A. M. *et al.* Estimation of the total parasite biomass in acute falciparum malaria from plasma PfHRP2. *PLoS Med* **2**, e204 (2005).

57 Adetifa, I. M. *et al.* Haematological values from a Gambian cohort--possible reference range for a West African population. *Int J Lab Hematol* **31**, 615-622 (2009).

58 Maier, A. G. & Rug, M. In vitro culturing *Plasmodium falciparum* erythrocytic stages. *Methods Mol Biol* **923**, 3-15 (2013).

59 Trager, W. & Jensen, J. B. Human malaria parasites in continuous culture. *Science* **193**, 673-675 (1976).

60 Dluzewski, A. R., Ling, I. T., Rangachari, K., Bates, P. A. & Wilson, R. J. A simple method for isolating viable mature parasites of *Plasmodium falciparum* from cultures. *Trans R Soc Trop Med Hyg* **78**, 622-624 (1984).

61 Ribaut, C. *et al.* Concentration and purification by magnetic separation of the erythrocytic stages of all human *Plasmodium* species. *Malar J* **7**, 45 (2008).

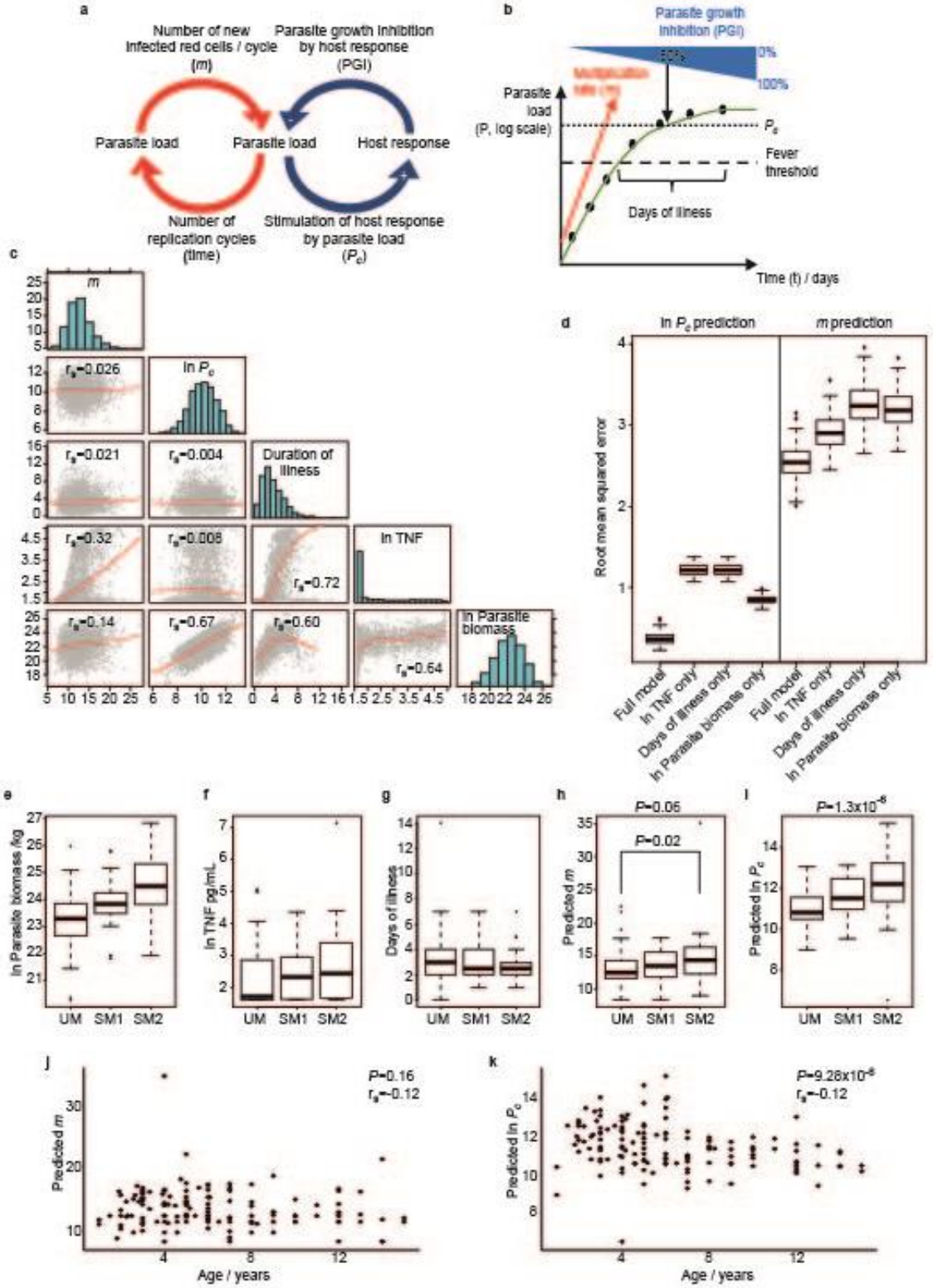
62 Malleret, B. *et al.* A rapid and robust tri-color flow cytometry assay for monitoring malaria parasite development. *Sci Rep* **1**, 118 (2011).

63 Pouw, R. B. *et al.* Complement Factor H-Related Protein 3 Serum Levels Are Low Compared to Factor H and Mainly Determined by Gene Copy Number Variation in CFHR3. *PLoS One* **11**, e0152164 (2016).

Correspondence and Requests for Materials: should be sent to Dr Aubrey Cunnington, a.cunnington@imperial.ac.uk. **Acknowledgments:** We are grateful to Klaus Dietz for providing the original data and parameter estimates from malariatherapy patients and his model, and to the St. Mary's NHLI FACS core facility and Yanping Guo for support and instrumentation. **Funding**: This work was supported by the Medical Research Council (MRC) UK via core funding to the malaria research programme at the MRC Unit, The Gambia; by the UK MRC and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement and is also part of the EDCTP2 program supported by the European Union (MR/L006529/1 to A.J.C.); by a Wellcome Trust Value In People Award to A.J.C; by European Union's seventh Framework program under EC-GA no. 279185 (EUCLIDS; www.euclids-project.eu). **Author contributions**: A.J.C., A.G., A.E.vB., F.F., D.J.C., D.N., and M.W. collected the data used in the study; A.J.C., E.M.R., M.T.B., M.W. and D.J.C designed the study; A.J.C. and M.T.B. developed the mathematical model; A.J.C., M.T.B., H.J.L., F.F., T.D.O and A.E.vB. analysed the data; D.W., T.W.K, D.N., U.D., E.M.R., M.L., L.J.C., A.G., D.J.C., and A.J.C supervised aspects of the project; all authors contributed to interpretation of the results and drafting the manuscript. **Competing interests**: The authors declare that they have no competing financial interests.

854 **FIGURES**

855 Figure 1.



856

Figure 2.

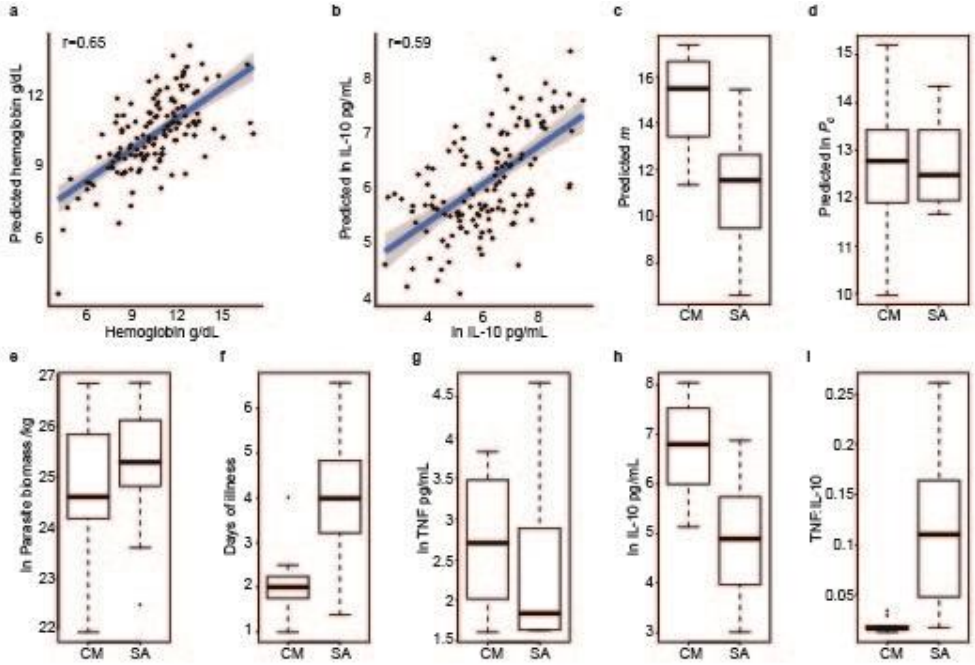

a
r=0.65
Predicted hemoglobin g/dL
Hemoglobin g/dL
b
r=0.59
Predicted ln IL-10 pg/mL
ln IL-10 pg/mL
c
Predicted m
CM
SA
d
Predicted ln P_0
CM
SA
e
ln Parasite biomass /kg
CM
SA
f
Days of illness
CM
SA
g
ln TNF pg/mL
CM
SA
h
ln IL-10 pg/mL
CM
SA
i
TNF:IL-10
CM
SA

Figure 3.

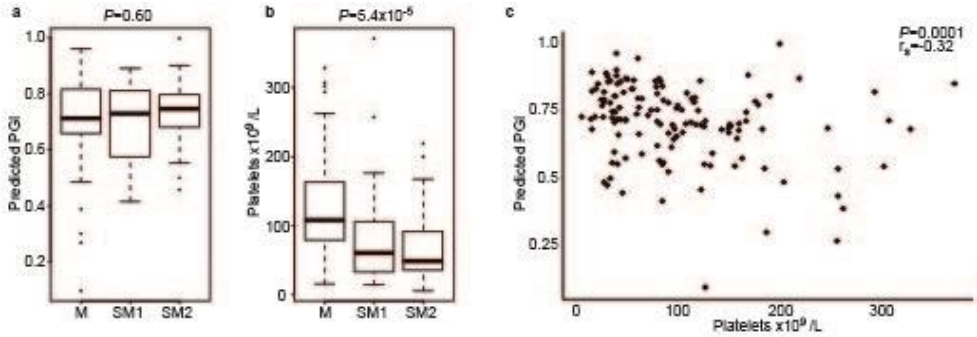

a
P=0.60
Predicted PGI
1.0
0.8
0.6
0.4
0.2
M
SM1
SM2
b
P=5.4x10-5
Platelets x10⁹ /L
300
200
100
0
M
SM1
SM2
c
P=0.0001
rs=-0.32
Predicted PGI
1.0
0.75
0.5
0.25
0
100
200
300
Platelets x10⁹ /L

Figure 4.

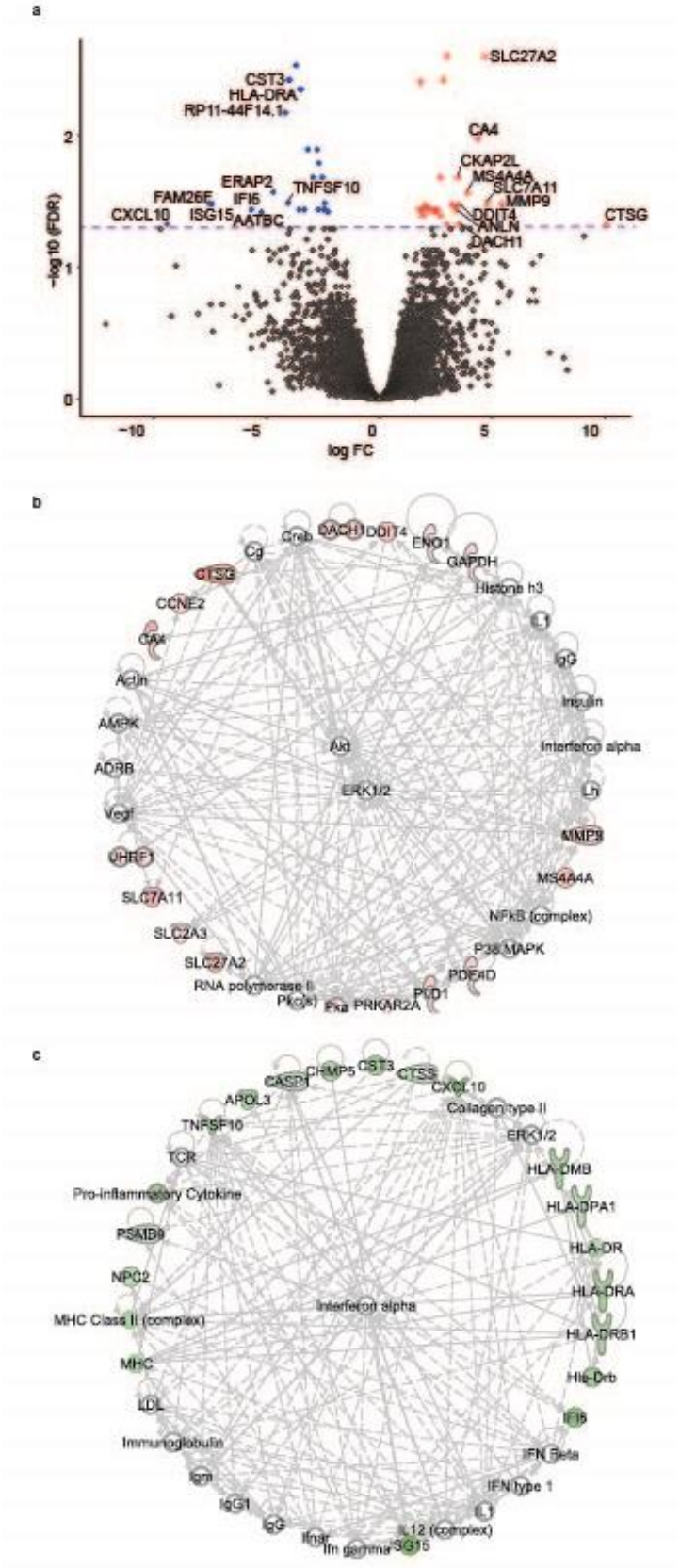

a
SLC27A2
CST3
HLA-DRA
RP11-44F14.1
CA4
CKAP2L
MS4A4A
SLC7A11
MMP9
DDIT4
ANLN
DACH1
CTSG
ERAP2
IFI6
AATBC
TNFSF10
FAM26E
CXCL10
ISG15
−log10 (FDR)
log FC
b
Crab
DACH1
DDIT4
ENO1
GAPDH
Histone h3
Cg
CTSG
CCNE2
CA4
IL1
IgG
Actin
Insulin
AMPK
Akt
Interferon alpha
ADRB
ERK1/2
Lh
Vegf
MMP9
UHRF1
MS4A4A
SLC7A11
NFkB (complex)
SLC2A3
P38 MAPK
SLC27A2
PDE4D
RNA polymerase II
PKD1
Pka
PRKAR2A
c
CHMP5
CST3
CASP1
CTSS
CXCL10
APOL3
Collagen type II
TNFSF10
ERK1/2
TCR
HLA-DMB
Pro-inflammatory Cytokine
HLA-DPA1
PSMB9
HLA-DR
NPC2
HLA-DRA
Interferon alpha
MHC Class II (complex)
HLA-DRB1
MHC
Hla-Drb
LDL
IFI6
Immunoglobulin
IFN Beta
Igm
IFN type 1
IgG1
IL1
IgG
IL12 (complex)
Ifnar
Ifn gamma
ISG15

Figure 5.

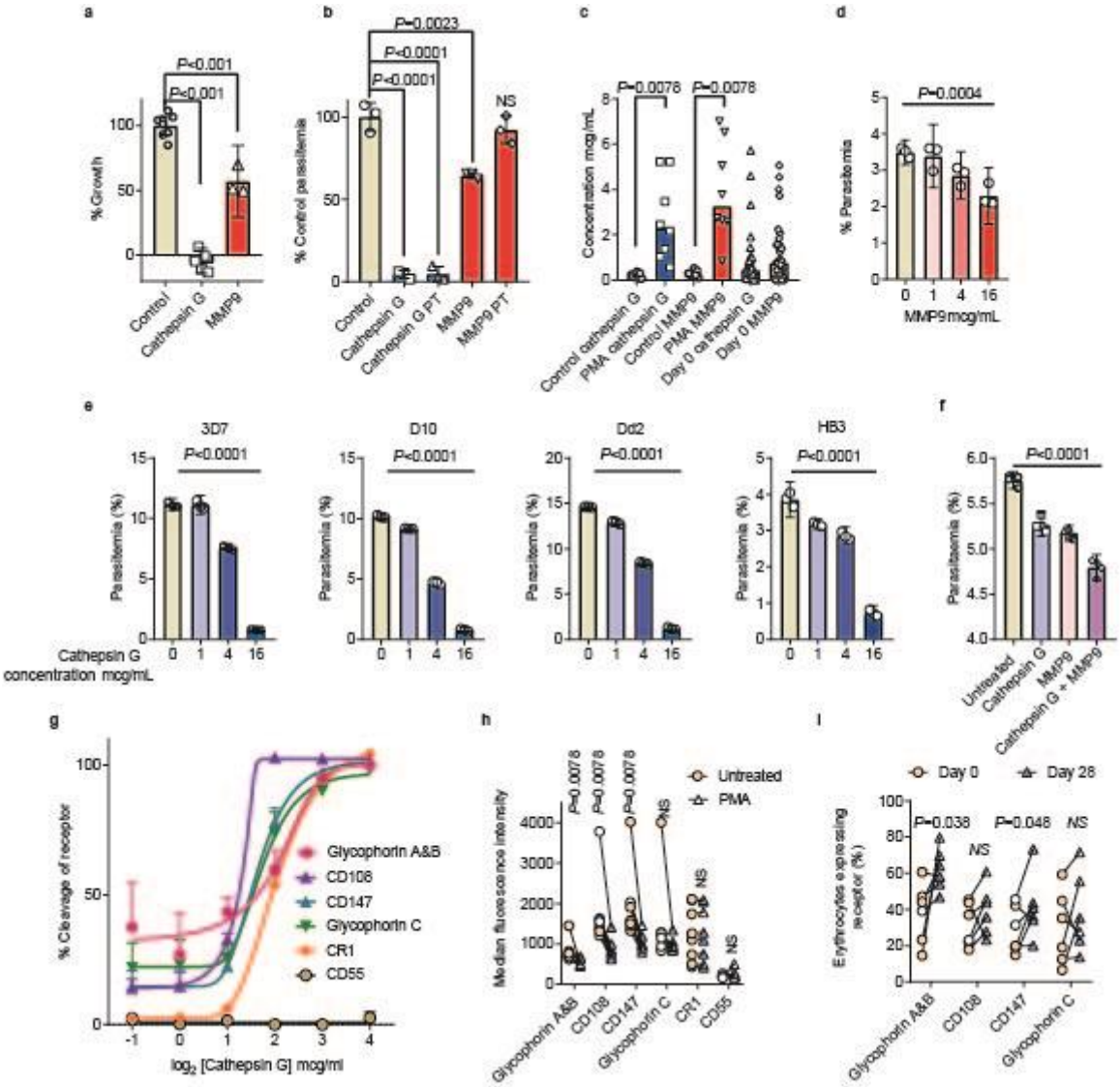
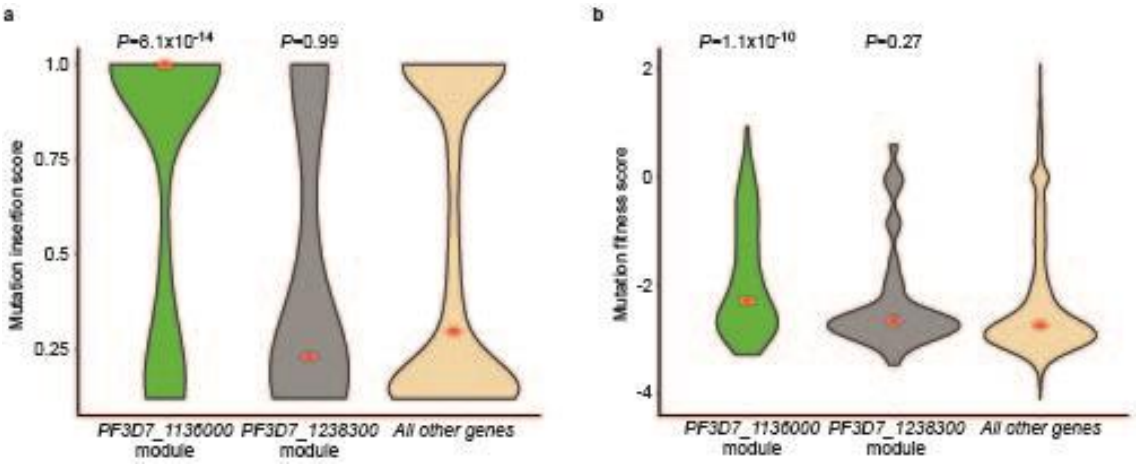

a
P<0.001
P<0.001
% Growth
Control
Cathepsin G
MMP9
b
P=0.0023
P<0.0001
P<0.0001
NS
% Control parasitemia
Control
Cathepsin G
Cathepsin G PT
MMP9
MMP9 PT
c
P=0.0078
P=0.0078
Concentration mcg/mL
Control cathepsin G
PMA cathepsin G
Control MMP9
PMA MMP9
Day 0 cathepsin G
Day 0 MMP9
d
P=0.0004
% Parasitemia
MMP9 mcg/mL
e
3D7
P<0.0001
D10
P<0.0001
Dd2
P<0.0001
HB3
P<0.0001
Parasitemia (%)
Cathepsin G concentration mcg/mL
f
P<0.0001
Parasitemia (%)
Untreated
Cathepsin G
MMP9
Cathepsin G + MMP9
g
% Cleavage of receptor
Glycophorin A&B
CD108
CD147
Glycophorin C
CR1
CD55
log2 [Cathepsin G] mcg/ml
h
Median fluorescence intensity
Untreated
PMA
P=0.0078
P=0.0078
P=0.0078
NS
NS
NS
Glycophorin A&B
CD108
CD147
Glycophorin C
CR1
CD55
i
Day 0
Day 28
Erythrocytes expressing receptor (%)
P=0.038
P=0.048
NS
NS
Glycophorin A&B
CD108
CD147
Glycophorin C

Figure 6.


a
P=6.1x10-14
P=0.99
1.0
0.75
0.5
0.25
Mutation insertion score
PF3D7_1136000 module
PF3D7_1238300 module
All other genes
b
P=1.1x10-10
P=0.27
2
0
-2
-4
Mutation fitness score
PF3D7_1136000 module
PF3D7_1238300 module
All other genes