Title

Identifying serological markers of recent P. falciparum exposure for precision malaria surveillance

Author

Vinden, Joanna

Publication Date

2024

Peer reviewed|Thesis/dissertation

Identifying serological markers of recent *P. falciparum* exposure for precision malaria surveillance

By

Joanna Vinden

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor in Philosophy

in

Infection Diseases & Immunity

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Dr. Bryan Greenhouse, Co-chair
Dr. Eva Harris, Co-chair
Dr. Laurent Coscoy
Dr. John Marshall

Summer 2024

Abstract

Identifying serological markers of recent *P. falciparum* exposure for precision malaria surveillance

By

Joanna Vinden

Doctor of Philosophy in Infectious Diseases & Immunity

University of California, Berkeley

Dr. Bryan Greenhouse, Co-chair
Dr. Eva Harris, Co-chair

Surveillance plays a key role in malaria control and elimination efforts by allowing for informed and effective allocation of often limited resources. Current methods for estimating malaria exposure are limited either by cost or accuracy. Serological data offers the potential to provide inexpensive and accurate estimates of exposure, but to date there is no consensus on which antibody responses are informative and how they can be interpreted. Previous studies have been limited to single age groups, transmission intensities, and geographic locations that are not generalizable across populations. This dissertation describes an investigation into serologic biomarkers of recent exposure to *Plasmodium falciparum*, the most deadly species causing malaria, using samples from 8 cohort studies representing diverse populations. Using an innovative approach that combined detailed individual-level exposure data, high-throughput screening of hundreds of antibody responses, and robust statistical methods, we were unable to identify a set of antibody responses predictive of recent exposure that was generalizable across settings. Although universal seromarkers may not exist, we found that accurate prediction of recent exposure was possible in a cohort of return travelers, suggesting the potential for seromarkers to be developed for specific settings, particularly those with limited cumulative exposure.

Dedication

For my grandfather, Dr. Gilbert Dickson Vinden, from whom I inherited a curious mind.

For my parents, Dr. Leslie Alanna Scott and Dr. Christopher Martyn Vinden, who fostered my curiosity, sponsored my sense of adventure, and shone unwavering love and support on me through every step of my journey - no matter how far it took me away from them. I won the lottery of life to have you as my family.

And for my nephew, Maxfield Wayne Vinden, whose companionship has brought more joy to my life than words can describe.

Acknowledgments

I would like to express my deepest gratitude to my advisors, Dr. Bryan Greenhouse, Dr. Eva Harris, and Dr. Isabel Rodriguez-Barraquer. It is a privilege to have been under their supervision and guidance for the last 6 years. Their mentorship challenged me to advance my skills in statistical programming and machine learning and transformed me into the scientist I am today. I am especially grateful to Dr. Eva Harris for supporting me through the most difficult personal and academic challenges I have faced over the course of my PhD.

I am extremely fortunate to have been surrounded by the malaria community at UCSF and around the world. I am especially indebted to Dr Phil Rosenthal, Melissa Conrad, Leanne Robinson, Ivo Mueller, Chris Drakeley, Kevin Tetteh, Issac Sswenyana, Yaobao Liu, Jun Cao, Chris King and Lenor Carias. I thank them for supporting this project and giving thoughtful feedback.

I would also like to thank EPPIcenter Lab members Jessica Briggs, Isobel Routeledge, Marco Hamins-Puertolas, Max Murphy, Sofonias Tessema, and Yoon Lee for helping me to debug code, commiserating about trash data, and being goofy with me after long days in the lab. In particular I would like to thank Sophia Maxfield and Charles Le without whom I would never have been able to generate all the data necessary for this project.

I would like to express my appreciation for the support I received from Dr. Laurent Coscoy and Dr. John Marshall as members of my dissertation committee, as well as Teresa Liu without whom I would have never been able to figure out CalCentral.

Finally I would like to thank my partner, Dr. Stephen Michael Baum, for being more supportive than I ever knew was possible.

Introduction

Background

Malaria is a significant global health burden with >200 million cases per year worldwide causing roughly 400,00 deaths, primarily in African children less than five years old.[1] Malaria is a blood-borne parasitic disease caused by protozoa of the Plasmodium genus. There are five species of Plasmodium that infect humans, of which P. vivax and P. falciparum are the most common. Whereas P. vivax causes the majority of cases in Southeast Asia, the Eastern Mediterranean, and the Americas, P. falciparum – which is the focus of this work–is the most fatal and accounts for >99% of malaria cases on the African Continent.[2]

Plasmodium species have complex life cycles involving several morphological stages in humans and mosquitoes. Infection in humans occurs through the bite of an infected female Anopheline mosquito. Upon infection, the sporozoite stage of the parasite enters the bloodstream and travels to the liver. Inside hepatocytes, the parasite reproduces asexually and matures into merozoites, which rupture the hepatocyte and enter the bloodstream to infect red blood cells. Infection of red blood cells marks the beginning of the asexual erythrocytic stage of infection, which is responsible for the symptoms of malaria. In the erythrocytic phase, merozoites cyclically infect red blood cells, undergo asexual reproduction, and rupture the red blood cell to release merozoites that infect more red blood cells. Eventually, some merozoites mature into gametocytes, which are taken up by a mosquito, where sexual reproduction of the parasite occurs.[3]

Studying malaria epidemiology is complicated by clinical immunity, multiplicity of infection, and heterogeneity of exposure

Infection with Pf leads to one of three clinical outcomes: severe malaria, uncomplicated malaria, or asymptomatic infection. Severe malaria is the most deadly and is defined as the presence of parasitemia with fever, anemia, and at least one severe outcome (i.e., coma, kidney failure, etc). Uncomplicated malaria is defined as the presence of parasitemia with fever and with or without mild anemia. Asymptomatic infection is defined as the presence of parasitemia with no fever. Outcome of infection is determined by age and ‘clinical immunity’—immunity against the symptoms of malaria (but not against infection).[4] Clinical immunity develops gradually with repeated infections such that, in a malaria-endemic setting, incidence of malarial disease peaks in childhood, while incidence of asymptomatic parasitemia continues throughout life.[5–7] Although clinical immunity does not offer protection from infection, it does result in infections with lower parasite density, making these infections more difficult to detect.[8]

Improved surveillance methods are critical for malaria control

Control and prevention of malaria depends on interventions such as insecticide-treated bed nets, indoor residual spraying, and the use of artemisinin combination therapy. Increased investment in access to malaria control measures between 2000-2015 led to a substantial decrease in malaria burden.[9] however, the decline has reached a plateau as we face new challenges in malaria elimination.[10]

As transmission declines, exposure becomes increasingly heterogeneous, with a small proportion of the population experiencing a majority of infections.[11–16] To address this challenge, WHO's Global Technical Strategy emphasized improved malaria surveillance as a necessary tool to detect and target resources towards transmission hotspots and evaluate effectiveness of interventions.[17] Particularly as global funding for malaria control is decreasing,[10] using surveillance data to allocate limited resources effectively will be paramount to the success of intervention strategies.

Existing methods of surveillance are limited by high cost or low accuracy

Clearly, affordable and accurate methods for measuring surveillance are essential. Unfortunately, existing methods force a trade-off between cost and accuracy. The most accurate methods to measure exposure, using cohort studies or entomological methods, are time-consuming and far too costly for routine surveillance. The most commonly used surveillance metric is the proportion of individuals in a population with parasitemia, termed the parasite prevalence, or parasite rate (PR). PR is estimated via cross-sectional studies, which are inexpensive and can be done rapidly. However, since they only capture active infections at a single point in time, they may be biased by seasonality or antimalarial treatment levels. At very high transmission intensities, the PR becomes saturated—where a large change in transmission yields a small change in PR—and at low transmission intensity a prohibitively large sample size is needed to maintain statistical power.[18] Further, the accuracy of PR is limited by the sensitivity of the diagnostic test used. Because acquired immunity decreases parasite density of infections, false negative tests caused by infections below the limit of detection will be more likely with increased age and exposure.

The use of routine passive surveillance at healthcare facilities is another inexpensive method of measuring exposure. In this case, detection may not be able to detect spatial heterogeneity if information on where transmission occurred is not collected. The value of these data also depends on the accuracy of the diagnostic test and the completeness of reporting, which can vary significantly between settings. Further, these data will fail to

capture asymptomatic cases and populations without access to formal healthcare systems.

Serological methods of measuring exposure offer a means of increasing accuracy of surveillance at low cost

Serological methods, which measure antibody responses against malarial antigens, offer a means of measuring exposure that is both accurate and affordable.[19–21] The genome of malaria encodes over 5000 genes, hundreds of which are known to elicit IgG antibody responses. The antibody response will have a distinct rate of boost and decay following exposure, offering a wealth of information on previous exposure—if we know how to interpret it. With serological data, samples from a cross-sectional study could be used to detect exposure over a period of time, rather than being limited to active infections as with PR. This would allow exposure to be estimated in low-transmission settings with an achievable sample size, and resolution for detection of spatial hotspots or changes in transmission over time. Although there has been increased interest in recent years as to the potential for using serology for routine surveillance, currently there is no consensus as to which antibody responses to measure in order to determine recency of exposure.

Chapter 1

Serological markers of recent *Plasmodium falciparum* exposure in a cohort of Chinese overseas laborers returning from sub-Saharan Africa

Abstract

Improving precision of malaria surveillance methods will be paramount to the effort to eliminate malaria, particularly in areas of declining transmission. Using serological responses to assess exposure at the individual level represents a potential method to improve the power of surveillance methods by detecting infection over a period of time, rather than at a single point in time. This would allow for accurate estimates using smaller sample sizes, which will be vital in declining transmission or elimination settings, where required sample sizes for detecting infection at a single time point become prohibitively large. Currently, few studies have investigated which combinations of antigen-specific antibody responses are predictive of recent exposure, and there is no consensus between studies on which antigens to use. In this study, we measured antibody responses against a panel of 19 malarial antigens (18 Pf antigens and one mosquito salivary protein) in plasma samples taken at varying times post infection from a cohort of Chinese nationals who tested positive for Pf malaria upon return from sub-Saharan Africa where they were overseas laborers. We find that several combinations of antigens can predict exposure within the previous 6 months with sens95 (sensitivity at 95% specificity) above 50%. Future studies will be required to assess the validity of these results in individuals with endemic exposure levels.

Introduction

The promising decline of the global malaria burden achieved in the early 2000s has plateaued since 2015 as countries with declining endemicity face new challenges in malaria control.[22] Among the main challenges is the increasing heterogeneity of malaria exposure that occurs with declining endemicity, making it increasingly relevant to target limited resources to the areas where they will have the most impact.[23–25] Improved surveillance systems to better identify hotspots and evaluate interventions will be important in overcoming this challenge, as outlined in the WHO Global Technical Strategy for Malaria.[26]

Currently, most malaria surveillance programs rely on clinical case counts and prevalence surveys.[27] Clinical case counts can help identify where malaria is occuring, but fail to capture asymptomatic cases and exclude subpopulations that do not engage in formal healthcare. Prevalence surveys in low transmission settings require sample sizes that are prohibitively large.

Using serology to detect recent exposure (i.e. identifying individuals with at least one infection in the previous 6 months) could improve the accuracy and power of surveillance methods over incidence and prevalence measurements alone, particularly in low transmission settings.[19,28] By capturing infections that occurred over a period of time rather than at a single time-point, testing for recent exposure requires smaller sample sizes than testing for active infection, or alternatively offer more spatial resolution for a given sample size. It could also capture recently cleared infections that have contributed to transmission. Additionally, being able to detect infections over a period of time could add flexibility to the timing of active surveillance.

Antibodies present an opportune method of measuring recent exposure as they remain detectable for months to years after infection. Further, antibodies can be measured with relatively simple assays, including inexpensive laboratory-based assays and point-of-contact tests, which makes them a feasible option for surveillance programs in resource-limited settings. For low transmission settings in particular, antibodies against AMA1 and MSP1 have been shown to be strongly linked with population-level cumulative exposure.[29–32] However, a serological tool that provides accurate information about the recency of infection has yet to be developed.

A small number of studies have identified individual level seromarkers of recent exposure to P. falciparum in different settings,[33,34] but there is no consensus between these studies on the optimal seromarkers to use. At an individual level, the magnitude of the antibody response is dependent not only on the time since the most recent exposure but may also depend on the also the individual's number and timing of prior exposure throughout life.[35] Therefore it is possible that the performance of putative seromarkers may vary based on factors including the transmission intensity of the setting and the age distribution of the population.

Studies in returned travelers to non-endemic areas offer a way of observing antibody responses to natural infection in individuals with limited pre-existing immunity and in complete absence of reinfection. A study by Yman et al investigated the kinetics of antibody responses against 111 Pf antigens in 65 return travelers followed longitudinally after a natural malaria infection. They identified panels of 5 serological markers that

could detect exposure within 3 months with sensitivity and specificity above 80%. These results are promising but have yet to be replicated in a larger population.

In order to investigate seromarkers of recent exposure in return travelers, we established a cohort of Chinese overseas laborers who tested positive for malaria upon return to China from sub-Saharan Africa. Samples were collected from a range of time-points since antimalarial treatment and were probed with a Luminex panel of 19 malarial antigens selected for their potential to provide information on Time Since Infection) based on previous studies[36]. We demonstrate that in this population, we can use a small set of antigens to predict recent exposure with reasonable accuracy.

Results

The study population consists of 99 Chinese adults, nearly all males aged between 22 and 64 years, who spent varying durations (ranging from 2 weeks to 3 years) working as overseas laborers in Sub-Saharan Africa (Table 1, Supplemental Table 1). Overseas labor placements occurred in 20 countries across West, Central and East Africa, with Nigeria and Angola as the most common countries visited. Most participants were sampled once (93), but 2 participants were sampled three times and 4 participants were sampled twice. For the purposes of our analyses, all samples were treated as independent.

Our study assessed antibody responses to a panel of 19 Plasmodium falciparum antigens . On average, the breadth of antibody responses, defined as the number of antigens for which antibody levels exceeded two standard deviations above the mean of naive controls was 8, but declined as a function of time since infection from a median (IQR) of 12 (8-16) in the first 3 months after infection to 6 (3-9) after one year (Figure 2b). MFIs of antibody responses against single antigens showed varying negative non-linear associations with TSI. On average, MFIs decreased by 50% (range 22-68%) during the first 6 months after infection (FIgure 1a and 2c). Notably, antibodies against GLURP and ETRAMP5 displayed the strongest correlations with Time Since Infection (TSI), with correlation coefficients of -0.42 and -0.33, respectively (Fig 2a).

We next assessed the ability of antibody responses to predict whether individuals had been recently infected (defined as an infection occurring in the 6 months prior to sample collection). We calculated area under the receiver operating characteristic curve (AUC) as a general measure of classification performance, and sensitivity at 95% specificity ($sens_{95}$) as a more relevant measure of the potential utility of antibody response(s) for surveillance. The average AUC and $sens_{95}$ of the single antigen predictors was 0.68 (range 0.56-0.82) and 27% (range 12-41%), respectively. In terms of overall classification, GLURP.R2 emerged as the best performing single antigen predictor, with

an AUC of 0.82 ($sens_{95}$ 35%). However, MSP2.Dd2 had the highest value of $sens_{95}$ at 41% (AUC 0.70) (Figure 3).

In order to identify a parsimonious combination of antigens that maximized predictive accuracy, we evaluated random forest models using all possible 2-5 way combinations of the 20 antigens. We evaluated cross-validated prediction performance using both AUC and $sens_{95}$. There was modest improvement in the best cvAUC with each increase in panel size from 1 to 5 antibody responses, with GLURP included in all of the best combinations. $sens_{95}$ also improved with each increase in panel size from 1 to 5 antibody responses. The overall best $sens_{95}$ of 68% was obtained for a panel of 5 antigens: CSP, Hyp2, MSP2.CH150, Rh2.2030, and Rh5.1. Out of 21,699 possible combinations, 220 had $sens_{95}$ of 50% or above. Of these, MSP2.CH150, Hyp2, GLURP.R2, CSP, and Rh2.2030 were present in 115, 106, 105, 93, 82, respectively (Table 2). All combinations of the 6 antigens most frequently included in the best performing combination models showed robust performance, with AUC ranging from 0.86-0.91 and sens95 ranging from 0.32 to 0.62 (Figure 4).

We also assessed the predictive accuracy of combinations of antibody responses using a random forest classifier that incorporated all 20 antigens. Combining all responses provided a modest improvement over single antigen performance (cross-validated AUC (cvAUC) of 0.86 (95% CI: 0.79 to 0.93) vs. 0.82 for the best single antigen (Figure 5a). Improvements in sensitivity were even more modest, with $sens_{95}$ improving from 27% from a single antigen to 41% (95% CI: 21 to 62%) with all 20. GLURP demonstrated the highest variable importance within this model, followed by Rh2.2030 and MSP2.Dd2 (Figure 5b).

Discussion

Our study assessed the ability to use antibody responses to predict whether individuals had been infected with Plasmodium falciparum in the previous 6 months. We screened plasma samples from 99 returned travelers for IgG antibody responses against 18 Pf antigens and one mosquito salivary antigen. While responses to single antigens failed to achieve sensitivities (at 95% specificity, sens95) above 40%, we found that several combinations of antibody responses accurately predicted recent exposure with a sens95 above 50%. Ultimately, a useful serological tool for routine surveillance would require higher specificity and benefit from higher sensitivity than the combination of antigens and assay evaluated here, particularly in low transmission settings where detection methods with 95% specificity will yield more false positives than true positives. However, the performance achieved by combination models in this study from a limited panel of 19 antigens evaluated using a generic bead-based platform demonstrate the

potential for serological tools to be developed with further development of antigen panels.

Many combinations of antibody responses were able to predict recent exposure with sens95 above 50%. In particular, combinations from a subset of 6 of the 19 antigens used in this study (GLURP.R2, Hyp2, CSP, Rh5.1, MSP2.CH150, Rh2.2030) had better performance than combinations including other antigens. Interestingly, although MSP2.Dd2 had the highest sens95 when evaluated independently, it did not improve performance of models in combination with other antigens. These results indicate that while all combinations are not equivalent, there are several interchangeable combinations of antigens that may provide useful information on recent exposure.

To date, few other studies have attempted to identify seromarkers of exposure using cohorts of individuals with known exposure levels. Two studies in particular, by Helb et al and Yman et al have used machine learning methods to identify seromarkers predictive of recent infection.[33,34] Yman et al identified candidate seromarkers from a cohort of 65 adult returned travelers followed longitudinally after a natural malaria infection. They identified combinations of 5 antibody responses that were able to predict exposure in the previous three months with >80% sensitivity and specificity. Eight of the 111 Pf antigens assessed in their study were included here, including three of the 28 identified as informative in the prior study. Of these three, only one (MSP2) was found to be reasonably informative here.

Helb et al identified candidate seromarkers using a cohort of children in two malaria endemic regions of Uganda. Whereas in cohorts of return travelers there is no possibility of reinfection, in endemic areas TSI is measured using data from monthly active follow-up visits where children are monitored for parasitemia via microscopy. They identified combinations of 3 antigens that accurately classified exposure in the previous 6 months with an AUC of 0.86, which is comparable to the AUCs generated from the combination models here. Of the 655 Pf antigens assessed by protein microarray in Helb et al, 12 were included in our study's Luminex panel (GLURP, Hyp2, Rh5.1, CSP, Etramp5, Rh2.2030, PfAMA1, Etramp4, EBA140.RIII.V, Gexp18, PfMSP19, EBA175.RIII.V). Of the 31 antigens selected by Helb et al as informative of recent exposure, 7 were included in our Luminex panel (Hyp2, CSP, Etramp5, PfAMA1, Etramp4, Gexp18, PfMSP119), of which Hyp2, Etramp5, and CSP were also found to be informative in our study. Surprisingly, between this study, Yman et al, and Helb et al, no marker was selected as informative across all 3 studies. The lack of overlap between the antigens identified in these studies underscores the difficulty in identifying serological responses that can provide accurate estimates of exposure across populations with different Plasmodium falciparum exposure histories, including the

intensity, timing, and possibly genetic makeup of parasites, presence of other Plasmodium species, and genetic and environmental factors influencing the host response. Further, differences in the quality of antigen expression and antibody detection may affect results significantly.

One of the main motivations for identifying robust markers of recent exposure is to improve malaria surveillance, particularly in regions of low or declining transmission. Because the magnitude of an individual's antibody response is dependent not only on the time since the most recent infection, but also the prior history of infection[35], the responses measured in our cohort of return travelers in this study is unlikely to be representative of individuals living in some endemic areas who will have higher levels of historic exposure. It is possible that elimination settings could be comparable, but further studies will need to be conducted. Additionally, as the cohort only included adults and was 94% male, further studies will need to be done to compare results in children and women.

Our results were generated using a magnetic bead assay of 18 Pf antigens and 1 An. gambiae antigen that were selected based on results from previous studies of seromarkers of recent exposure.[37] This study has demonstrated that many of the 19 antigens are moderately informative of recent exposure when considered in combination, however there may be antigens not included in our assay that could further improve performance. Further, antigen quality may have a large impact on results. A study by Kobayashi et al assessed antibody responses to antigens produced as both IVTT-expressed exon products as well as E. coli-derived purified proteins on the same protein microarray and found that in many cases, reactivity to purified protein was significantly higher.[38] Although the antigens used in this study were purified, it is possible that further optimization of the expression system could better mimic protein folding in biological settings which may enhance reactivity. Future studies will be needed to assess more potential seromarkers and antigen expression systems.

Methods

Sample collection

Study participants were selected from a retrospective study of imported malaria cases in Jiangsu, China as described previously.[39] Briefly, imported malaria cases were identified using routine surveillance data from China Information System for Disease Control and Prevention (CISDCP), the national disease reporting system in China. Serum samples, as well as clinical information and travel history were collected at follow-up investigations as part of the study. Time Since Infection (TSI) was calculated as the number of days between malaria treatment and sample acquisition.

P. falciparum antigens

Magnitude of antibody responses to a panel of 20 recombinant antigens (18 P. falciparum antigens, 1 mosquito salivary antigen, and tetanus toxoid) was measured using a Luminex platform as described previously.[40] All antigens were expressed via E-coli GST-tagged fusion proteins except for AMA-1, which was expressed in Pichia pastoris as a histidine-tagged construct.[41] Antigens included in the panel were selected based on association with exposure or protection in previous studies. Namely, PfAMA1,[41] PfMSP1-19,[42] GLURP R2,[43] MSP2 CH150,[44] MSP2 Dd2[44] were included as markers of historic exposure; Etramp4 Ag2,[45] Etramp5 Ag1,[45] GEXP18,[33] HSP40 Ag1,[33] Hyp2,[46] SBP1,[47] PfSEA-1,[48] gSG6[49] as markers of recent exposure and Rh2 2030,[50] Rh4.2,[51] Rh5.1,[52,53] EBA140 RIII-V,[54] EBA175 RIII-V,[54] CSP[55] as markers of protection. Tetanus toxoid and glutathione S-transferase (GST) were included as internal controls for vaccine reactivity and GST-tag cross-reactivity, respectively. However, many of the study subjects were not vaccinated against tetanus,[56] and thus did not show reactivity.

Multiplex bead assay to measure IgG response

Antigen-bead coupling concentrations were optimised and coupled to MagPlex microspheres, or “beads”, as described previously.[20] Patient plasma samples were diluted 1:400 in Luminex assay buffer B (1xPBS, 0.05% Tween, 0.5% BSA, 0.02% sodium azide, 0.1% casein, 0.5% polyvinyl alcohol, 0.5% polyvinyl pyrrolidone, 15.25 g/mL E. coli lysate) the day before screening and stored overnight at +4°C.

Total IgG responses to the antigen panel were measured using a multiplex bead array assay described previously.[20] Briefly, 50 µl of a pooled antigen-coupled bead suspension was added to each well of a 96-well plate, washed with PBS/Tween-20 and incubated with 50 µl of 1/400 sample plasma. for 1.5 hours at 700RPM on a shaking platform at room temperature. A 6-point, 5-fold curve of pooled hyperimmune serum from Tanzania (CP3) from 1/10 was included on each plate to control for inter-plate variability, and two wells containing buffer B served as blanks. In addition to the 111 samples from returned overseas laborers, serum from 10 malaria-naive UK residents (Public Health England, 2016) were run as negative controls. Plates were then washed (PBS/Tween-20) and incubated for 1.5 hours with 50ul of fluorescent secondary antibody (Jackson Immuno 109-116-098: Goat anti-human Fcy-fragment specific IgG conjugated to R-Phycoerythrin (R-PE) at 700 RPM at room temperature. After washing, plates were incubated with 50ul of Luminex assay buffer A (1xPBS, 0.05% Tween, 0.5% BSA, 0.02% sodium azide) for 30 minutes at 700 RPM at room temperature. After a final wash, 100ul of PBS was added to each well and the plates were read on a MagPix bioanalyser and xPONENT 4.2 software (Luminex Corp, Austin, Texas). The background-adjusted median fluorescence intensity (MFI) of antibody responses achieving a bead count >30 were measured using a MAGPIX© bioanalyser. The CP3

curve on each plate was then used to standardise for inter-plate variability using loess normalization.

Data Analysis

R (R: A language and environment for statistical computing, v4.1.1) was used for data processing and analyses.

Recent exposure was defined as the infection having occurred within 6 months of sample collection. All samples were categorized as obtained from either a “recently infected” or “not recently infected” individual depending on whether or not they were collected within this specified time frame.

To evaluate if the antibody response to any single P. falciparum antigen was informative of recent exposure, binary classification using a threshold antibody level was applied to the data for each of the 20 antigens individually using ROC analysis. Performance of the threshold models was assessed by estimating the 5-fold cross-validated AUC, and 95% confidence intervals were estimated using the method described by LeDell et al.[57]

To assess the ability of combinations of antibody responses to predict recent exposure, random forest classifiers were fitted to the dataset containing all 20 antigens, as well as all possible two- to five-way combinations. Classifier performance and confidence intervals were determined by 5-fold cross-validated AUC, as for the single antigen threshold models. We assessed variable importance within these models using Mean Decrease in Accuracy, a permutation test-based metric.[58]

Table 1: Descriptive statistics of the study participants

Number of participants			99
Number of samples			107
Number female (%)			6 (6%)
Median age, years (range)			46.5 (22-64)
Median cumulative time of residency in endemic area, days (range)			220 (16-1222)
Median time since infection (TSI), days (range)			431 (2-2027)
Number of participants returning from endemic country	Country of travel	Angola	18
		Nigeria	18
		Equatorial Guinea	9
		Republic of Congo	8
		Guinea	7
		Other	40

Figure 1a: A heatmap of the MFI of the antibody response to each of the 20 antigens included in the Luminex panel

Rows correspond to individual antigens while columns correspond to individual samples. Samples are sorted from left to right by increasing TSI, except the 10 rightmost samples which come from naive individuals from the UK. Antigens are sorted from top to bottom by decreasing average MFI across all samples. Antibody responses against tetanus toxoid were also measured as a reactivity control, however vaccination of children against tetanus did not begin in China until 1978, and thus many of the adults in this cohort were unreactive.

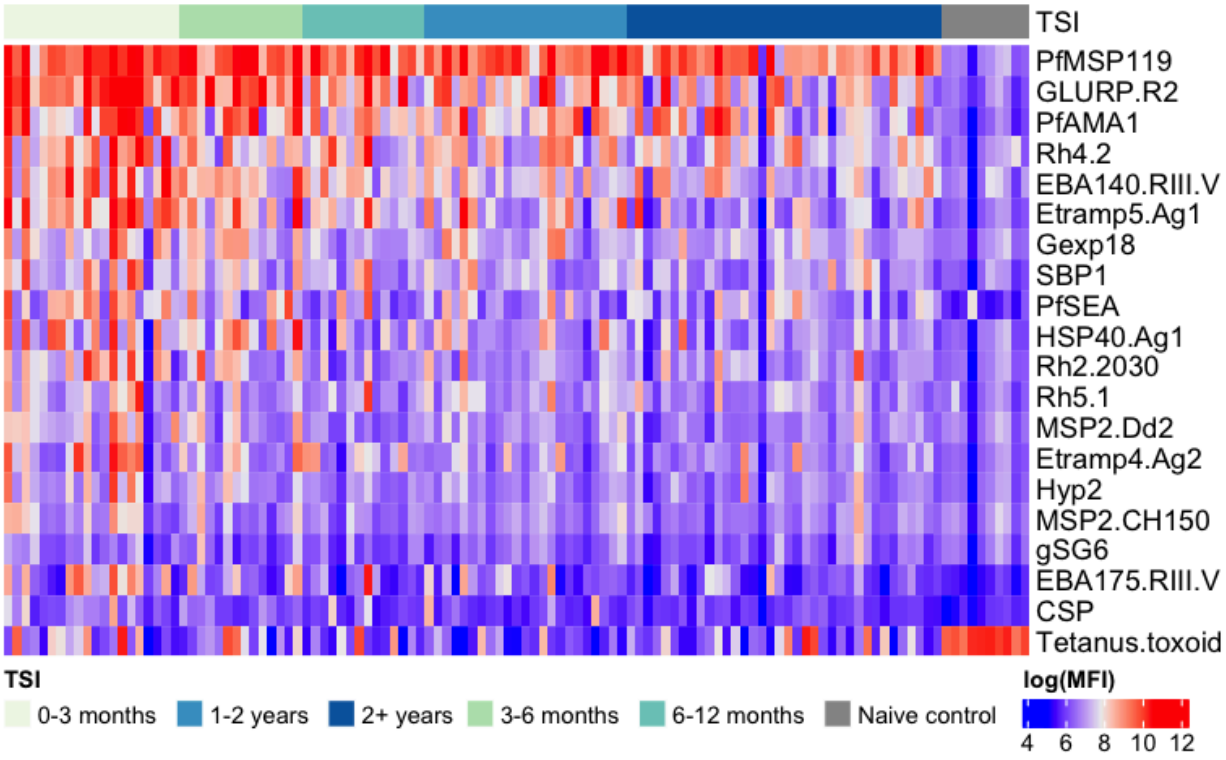


Figure 2: Antibody responses to a panel of 20 antigens show distinct relationships with TSI

Figure 2a

Points are colored by country of travel category as in Table 1. The black line represents the mean plus 2 standard deviations of the MFI across all naive control samples.

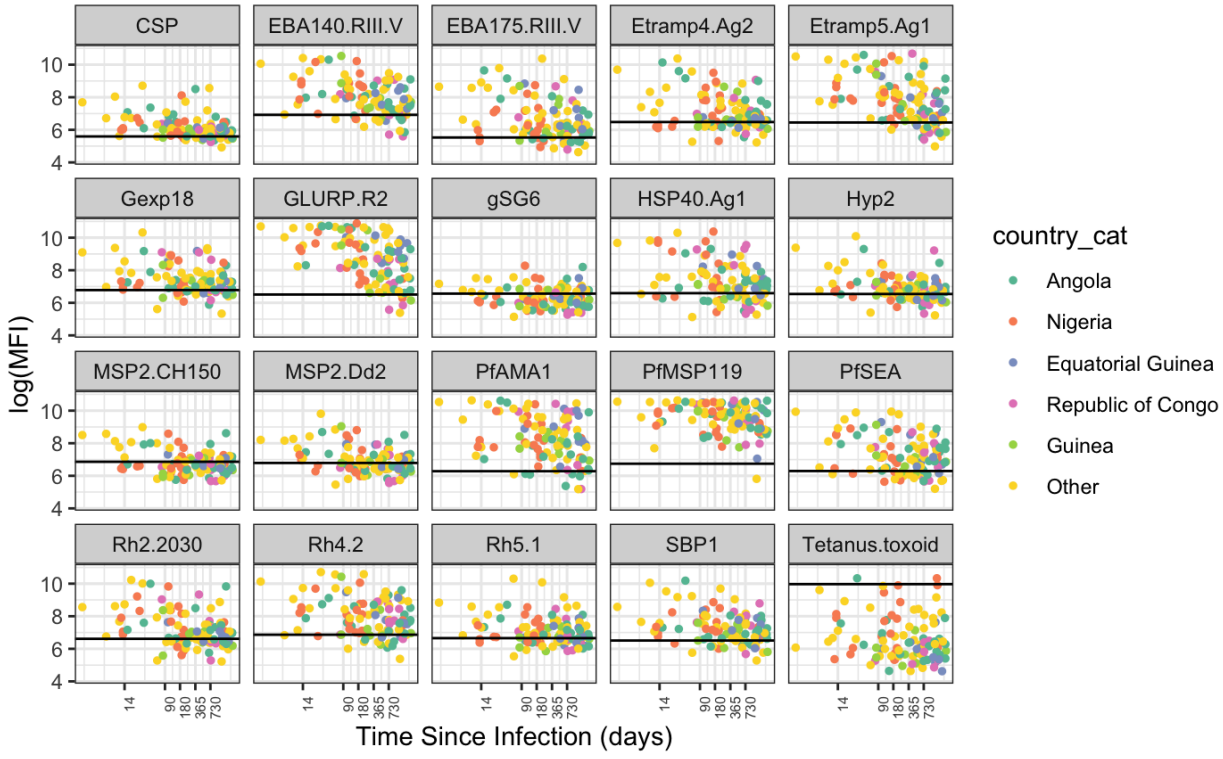


Figure 2b

Beeswarm plot overlaid with boxplot of the breadth of individuals' antibody responses over time. Breadth is expressed as the total number of antigens to which the individual has an MFI above that of the mean plus two standard deviations of the naive control samples. The lower and upper hinges of the boxes correspond to the first and third quartiles of the data and the center lines of the boxes represent the median.

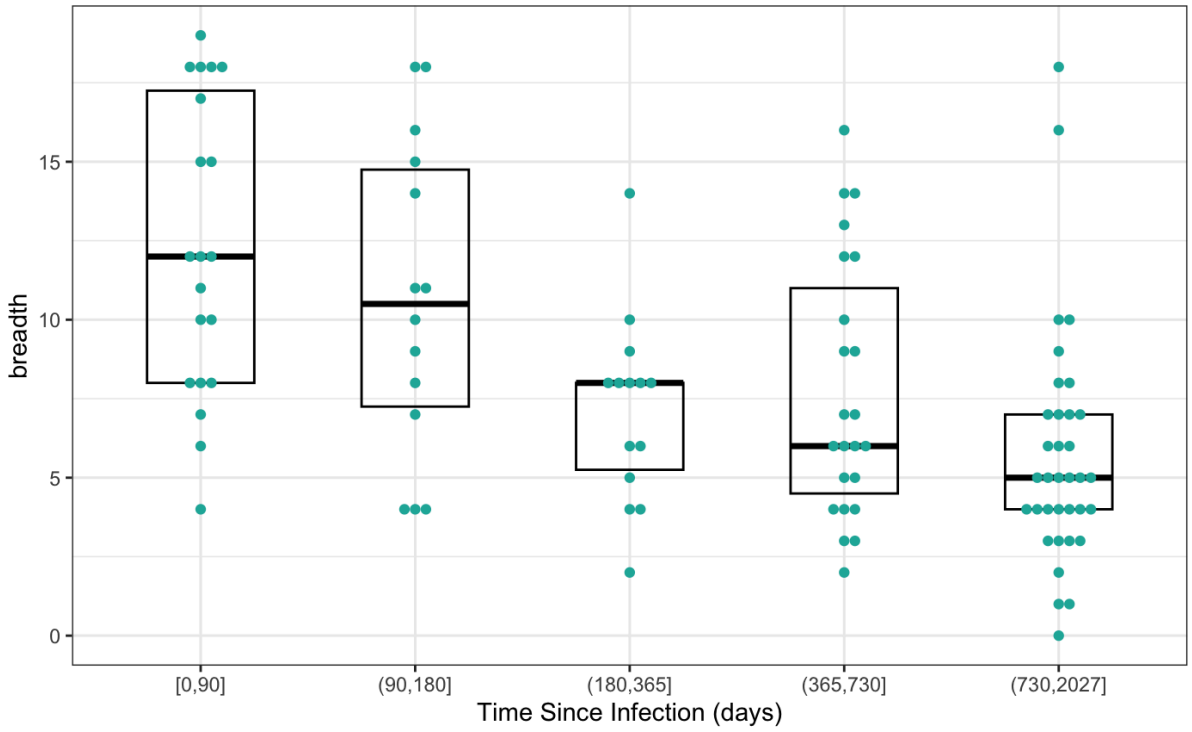


Figure 2c

Beeswarm plot overlaid with boxplot of the mean MFI of individuals' antibody responses across all 20 antigens over time. The lower and upper hinges of the boxes correspond to the first and third quartiles of the data and the center lines of the boxes represent the median.

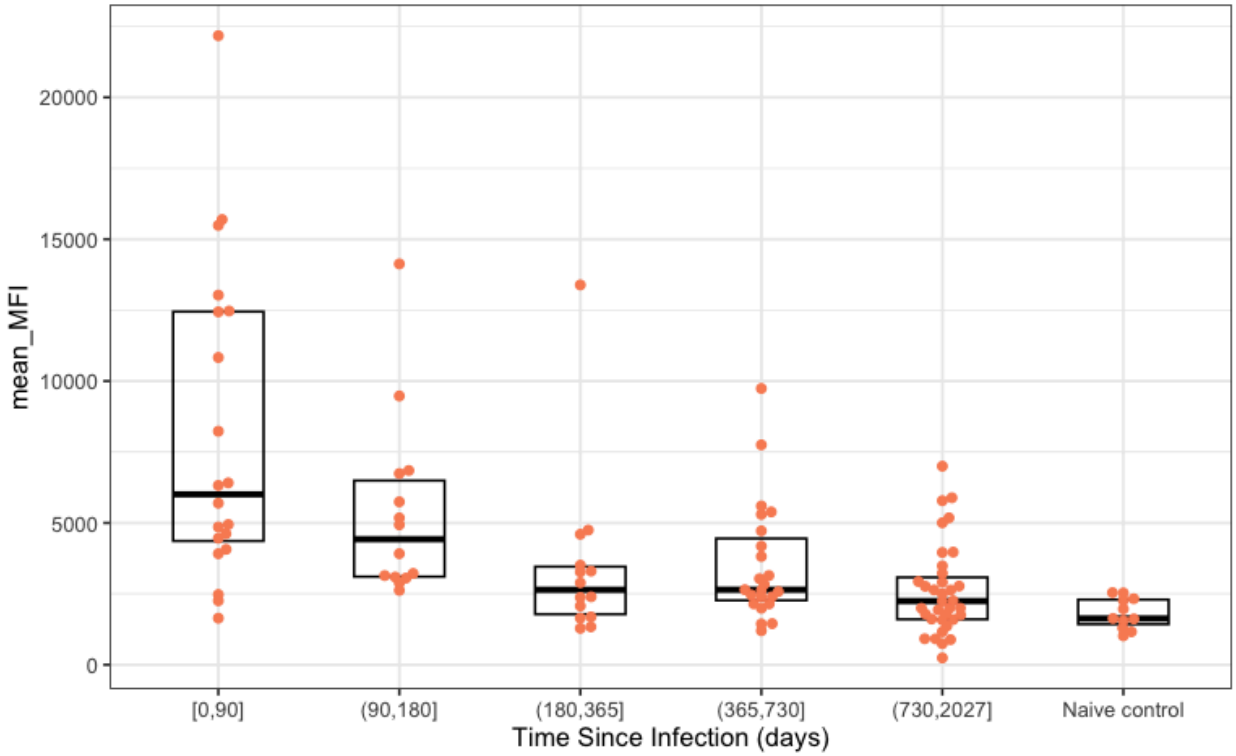


Figure 3: ROC curves for classifying infection in the previous 6 months using threshold antibody levels to a single antigen

Colored curves correspond to the top 5 antibody responses that were most accurate in detecting infection in the previous 6 months, with AUC 0.81, 0.77, 0.74, 0.73, and 0.72 for GLURP.R2, Etramp5.Ag1, CSP, Rh2.2030, and HSP40.Ag1, respectively.

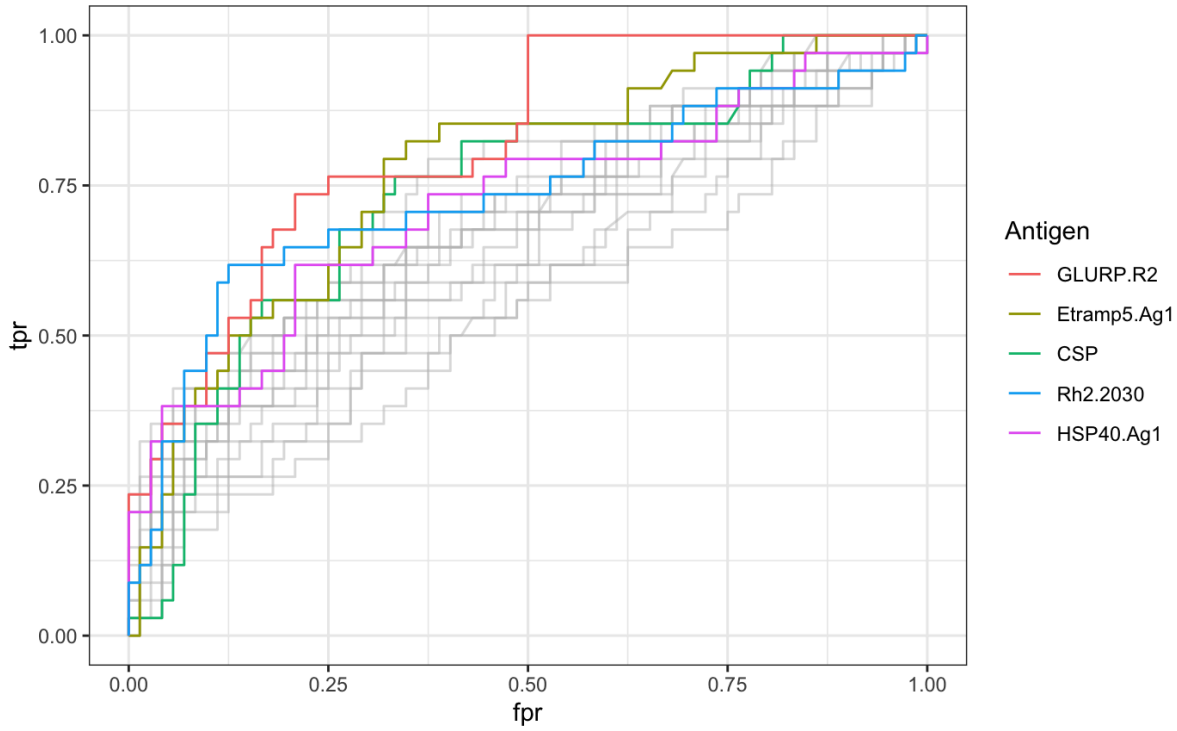


Table 2: Comparing variable importance in combination models

Ag	N models with $sens_{95}$ >50%	Ag selected as informative marker by	
		Helb, et al	Yman, et al
Total number of models	220	0	NA
GLURP.R2	105	0	NA
Hyp2	106	1	NA
CSP	93	1	NA
Rh5.1	67	0	0
Etramp5.Ag1	50	1	NA
MSP2.CH150	115	NA	1
Rh2.2030	82	0	NA
PfAMA1	53	1	0
MSP2.Dd2	37	NA	0
gSG6	58	NA	NA
HSP40.Ag1	56	NA	NA
Etramp4.Ag2	31	1	NA
PfSEA	36	NA	1
EBA140.RIII.V	44	0	0
Gexp18	14	1	NA
PfMSP119	33	1	1
Rh4.2	39	NA	NA
SBP1	9	NA	NA
EBA175.RIII.V	11	0	0
Tetanus Toxoid	16	NA	NA

Figure 4: Combinations of informative markers show robust performance

ROC curves for random forest classifier models of all possible 5-way combinations of the top 6 most informative antigens, MSP2.CH150, Hyp2, GLURP.R2, CSP, Rh5.1 and Rh2.2030

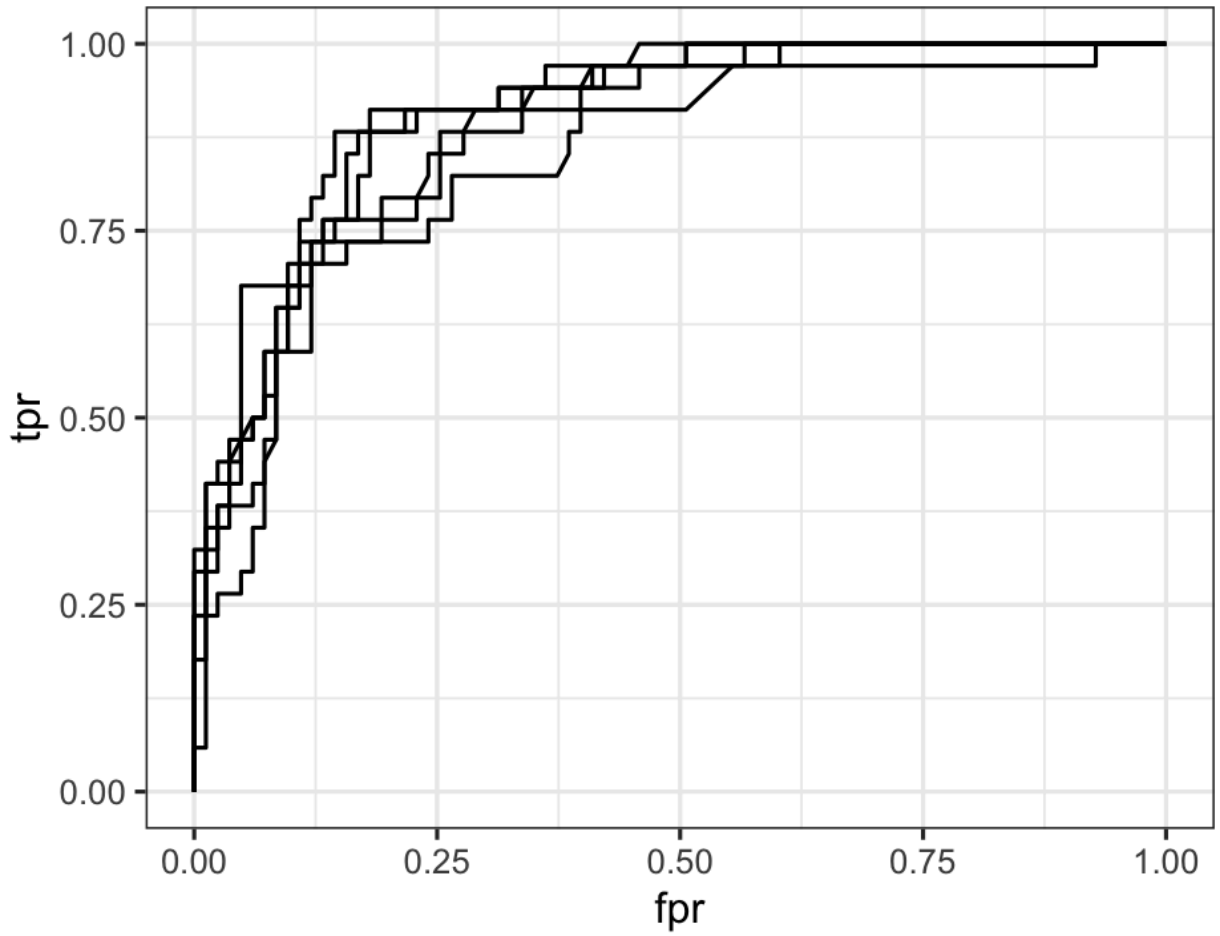


Figure 5: cvAUC and variable importance from random forest model fitted to data from all 20 antigens

Figure 5a

Gray curves represent individual cross-validated ROC curves from each fold of the 5-fold cross-validation of the random forest model. The purple curve represents the overall ROC curve from all folds together. The cross-validated AUC (cvAUC) is 0.86 with a 95% confidence interval of (0.79,0.93).

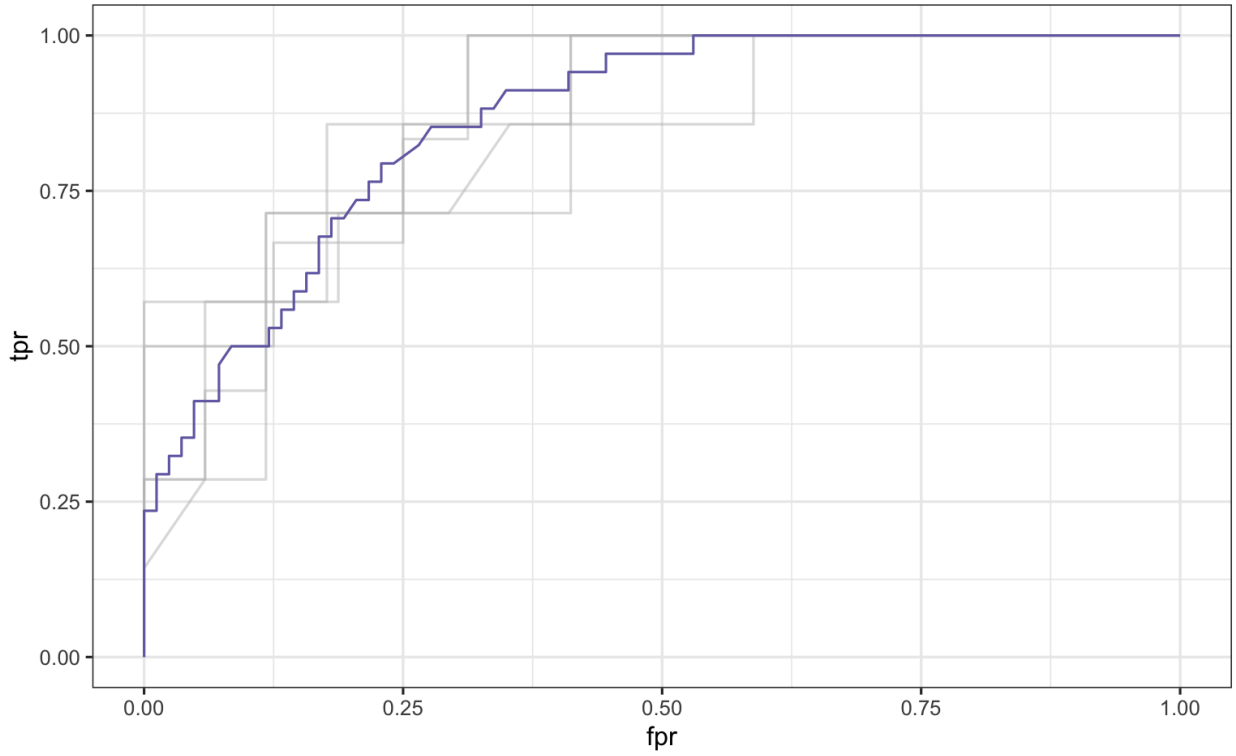
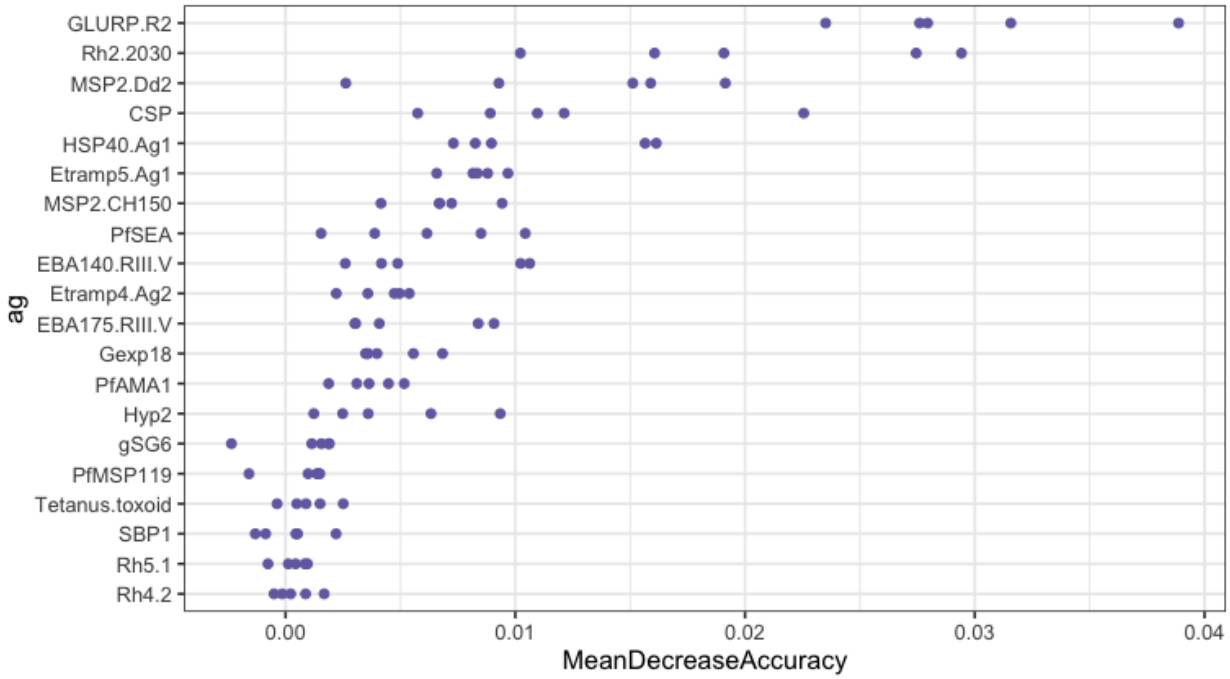


Figure 5b

Dots represent the MeanDecreaseAccuracy value for a given antigen for each of the 5 folds of the cross validated random forest model.

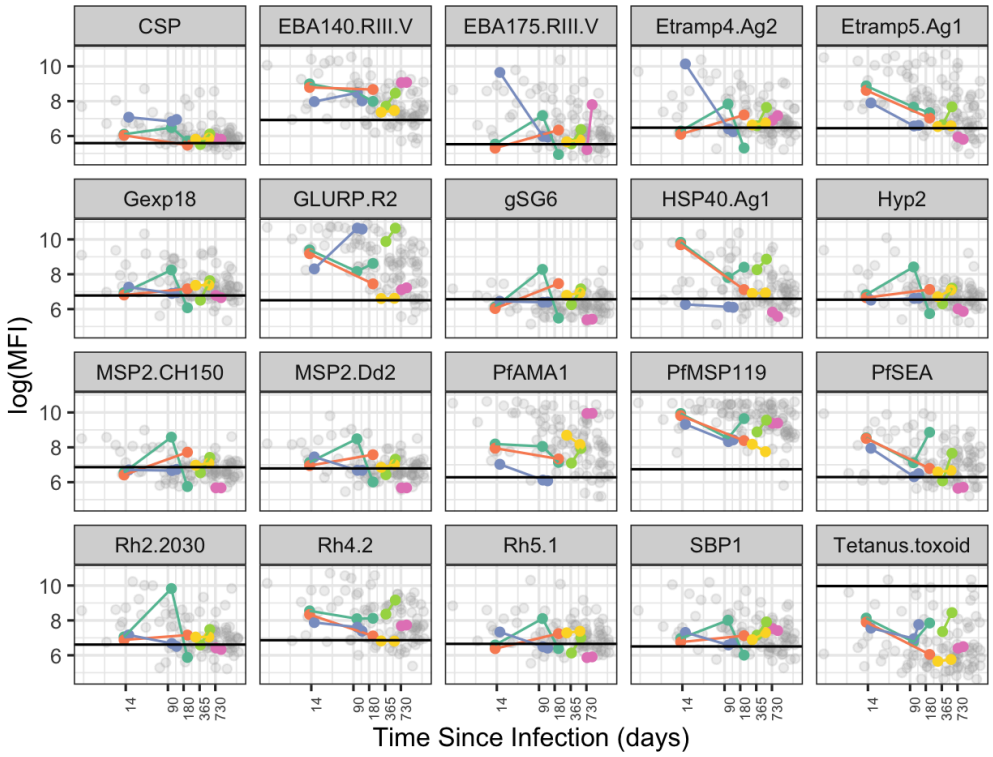


Supplemental Table 1: Countries of travel for all participants

Country	Number of participants
Angola	18
Nigeria	18
Equatorial Guinea	9
Republic of Congo	8
Guinea	7
Cameroon	5
Mozambique	5
Cote d'Ivoire	3
Democratic Republic of Congo	3
Sierra Leone	3
Togo	3
Zambia	3
Benin	2
Gabon	2
Ghana	2
Liberia	2
Uganda	2
Central African Republic	1
Malawi	1
Namibia	1
South Sudan	1

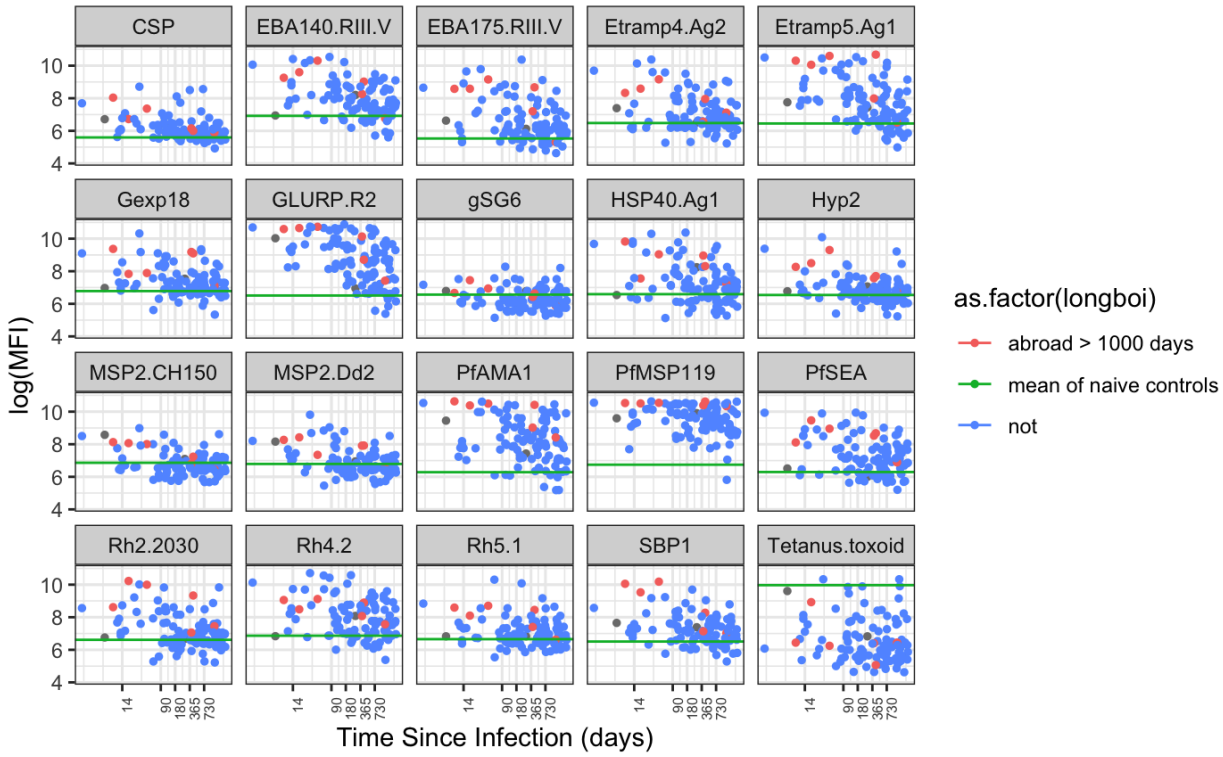
Supplemental Figure 1: Longitudinally sampled participants mirror the trends seen in the cohort as a whole

Colored lines represent individuals from whom multiple samples were collected. Light gray points represent all other samples.



Supplemental Figure 2: Participants who spent more than 1000 days abroad do not show systematically increased antibody titers

Red dots represent samples from individuals who spent more than 1000 days in sub-Saharan Africa. Blue dots represent all other samples.



Chapter 2

Lack of accurate seromarkers of recent *P. falciparum* infection that are generalizable across 8 cohort studies

Introduction

Surveillance plays a key role in malaria control and elimination efforts by informing effective allocation of limited resources and providing data to evaluate effectiveness of interventions. Ideally, surveillance programs would rely on accurate estimates of the risk of infection that can provide the spatial resolution necessary to identify hotspots of transmission as well as the temporal resolution required to detect changes in transmission over time.[16,59–64] Unfortunately, existing methods force a trade-off between cost and accuracy.[18] The most accurate methods to measure exposure, using cohort studies or entomological methods, are time-consuming and far too costly for routine surveillance. More affordable metrics such as clinical case counts and parasite rate (PR) are the most commonly used by surveillance programs. Clinical case counts can help identify populations experiencing malaria at low cost, but fail to capture asymptomatic cases and can only provide geographic data on where treatment is occurring and not where transmission is. PR, defined as the proportion of individuals in a population with parasitemia, only captures active infections at a single point in time and can therefore be biased by seasonality. In high transmission settings, the PR becomes saturated—where a large change in transmission yields a small change in PR.[19,20] In low transmission settings, a prohibitively large sample size is needed to maintain statistical power.[18]

Using serology to detect recent exposure could improve the accuracy and power of surveillance metrics at low cost. As antibody responses remain detectable for months to years after infection has passed, samples from a cross-sectional survey could be used to detect exposure over a period of time. Such surveys would be more statistically powerful than detection of active infections at a single time-point and therefore require smaller sample size than prevalence surveys or provide greater spatial resolution at a fixed sample size.

Serological data have been used previously to accurately estimate levels of cumulative exposure at the population level.[30–32,65–68] However, population level estimates rely on long-lived antibody responses, mainly against AMA1 and MSP1, and are not sensitive

to small (less than log-fold) or recent changes in exposure.[18] In order to leverage the full potential of serosurveillance, we need to determine the relationship between antibody responses and exposure and identify a set of antibody responses that can be used to accurately predict recent infection (e.g. at least one infection in the previous 6 months) at the individual level such that incidence can be estimated from cross sectional surveys.
One way we and others have attempted to identify candidate seromarkers is to quantify antibody response profiles against a large number of malarial antigens in individuals with known Time Since Infection (TSI). Studies in returned travelers offer a way of observing antibody responses to natural infection in individuals with limited pre-existing immunity and in complete absence of reinfection. While useful, results from these studies may not be generalizable to endemic settings. Assessing markers of recent infection in endemic settings requires longitudinal cohort studies with frequent active surveillance to detect asymptomatic infections.

Previous studies have identified candidate serological markers of recent exposure in either return travelers or endemic settings, however to date there is no consensus between studies on the optimal markers to use across settings. Two studies in particular, by Helb et al and Yman et al have used machine learning methods to identify seromarkers predictive of recent infection. Yman et al used samples from a cohort of 65 adult returned travelers followed longitudinally after a natural malaria infection to identify combinations of five serological markers that were able to detect exposure within the previous three months with >80% sensitivity and specificity.
Helb et al used samples from cohort of children in two malaria endemic regions of Uganda to identify panels of three antigens that accurately classified whether an individual had been infected in the previous six months with an AUC of 0.86. These studies demonstrate that accurate estimates of recent exposure can be obtained from a small set of antibody responses. However, the markers identified in each study do not overlap, illustrating the need to determine whether serological markers of recent exposure can remain accurate across populations.

Identifying markers of recent exposure of malaria in endemic settings is complicated by the inherent relationship between recency and frequency of exposure. Antibody titers are dependent on both TSI and cumulative number of infections experienced throughout life. Individuals who are at higher exposure risk (i.e. living in high transmission settings) will be infected more frequently, leading to higher cumulative exposure and consequently higher antibody responses. At the same time, individuals infected more frequently will have shorter TSIs on average. This confounding relationship may lead to erroneously attributing relationships to TSI that are actually the result of higher

cumulative exposure. To date, no existing studies have accounted for cumulative exposure as a confounding variable.

In this study, we aimed to identify a parsimonious set of antibody responses that accurately predict recent exposure (defined as at least one infection in the previous 6 months) across different populations. We used samples from 8 existing cohort studies that span a range of age groups, transmission intensities, and geographic settings. We used a novel method of sample selection designed to account for the confounding effects of cumulative exposure.

Results

Studies and sample selection

Participants consisted of 1196 individuals from 8 parent cohort studies representing varied geographic regions, transmission intensity settings, and demographics (Table 1). The Ilaita, Mugil, Kampala, Jinja, Kanungu, Tororo and Kilifi cohorts were longitudinal studies that took place in malaria endemic settings, where participants were followed by active and passive surveillance for at least 1 year before the collection of the plasma samples analyzed in this study. The Jiangsu study consists of returned travelers living in China, where there is no risk of malaria exposure, most of whom were surveyed only once post infection. Detection of parasitemia was done using molecular amplification methods (LAMP or qPCR) for all study sites except Kampala, where only microscopy was used, meaning subpatent infections could not be detected in this population.

Samples were selected for antibody profiling from the parent studies using a stratified selection procedure to minimize confounding due to the inherent relationship between time since infection (TSI) and cumulative exposure. Although most parent studies were longitudinal, only one sample per individual was selected for this study.

The proportion of samples selected from individuals with recent infection (defined as at least one infection in the previous 6 months) was highest for Mugil (90%) and lowest for Kilifi (37%) (Table 2). The proportion of samples where the most recent infection was symptomatic was highest in Jiangsu (100%) -- where all participants in the parent study were identified by presenting with symptomatic malaria upon returning to China from sub-Saharan Africa – and Kampala (84%), where subpatent infections were not assessed (Table 2).

Antibody responses against 888 P. falciparum antigens show

Antibody profiles for all 1196 samples were measured using protein microarrays containing 888 *P. falciparum* antigens (Figure 1). A subset of samples from the Jiangsu

cohort were run in parallel on a Luminex panel containing 16 of the same proteins, albeit not covering identical protein regions (Figure 2). Signals generated from the 2 assays were comparable for most antigens, with a mean correlation coefficient of 0.52 (IQR 0.42-0.68).

Overall, normalized signal from each of the 888 antigens decreased with increasing TSI (Figure 3, Figure 4). Recent infection was associated with higher breadth, defined as the number of antigen-specific antibody responses per sample with normalized signal greater than 2 standard deviations above the mean of the naive controls (Spearman's correlation r = -0.25, P<0.001), as well as the mean signal intensity per sample (r = -0.21, P<0.001) (Figure 4b).

Classifier models using combinations of antibody responses show limited ability to predict TSI

To assess the ability to predict recent exposure status based on antibody responses, we built classifiers to predict the binary outcome of whether an individual had been exposed within 6 months prior to sample collection (Figure 5). We calculated area under the receiver operating curve (AUC) as a measure of classifier performance. The average AUC of single antigen threshold modelsl was (mean, range). PTP5 was the best performing single antigen predictor, with an AUC of 0.66 (95% CI 0.62 to 0.69) .

We assessed the predictive accuracy of combinations of antibody responses using a random forest classifier that incorporated all 888 antigens (Figure 6a). Combining antigens did not provide substantial improvement over single antigen predictors (AUC 0.68, 95% CI: 0.64 to 0.72). PTP5 demonstrated the highest variable importance in this model (Figure 6b). Downselecting the antigens included in the model to only the top 20 antigens by mean variable importance values did not improve performance.

Though performance of classifiers predicting recent exposure across all ages and transmission intensity settings was low, we hypothesized performance might be better within subgroups of the data defined by age or transmission intensity. We stratified the samples into transmission intensity categories based on estimates of incident infections per year generated from the parent studies (Table 1). Samples from the Jiangsu cohort were placed in their own category because of the inherent differences in exposure histories between travelers and residents of malaria endemic regions. We generated random forest classifiers for each stratum using data from all 888 antigens (Figure 7a). Performance of each category of the transmission-specific classifiers was at or below that for all samples, except for the return traveler category, which had an AUC of 0.8 (95% CI 0.70 to 0.90) (Figure 7a). As the return traveler category was comprised only of

samples only from the Jiangsu cohort, the higher performance of this category may be due to other factors specific to the cohort, such as age of the study population.

Since immunity develops gradually with age and repeated infection, children are more likely to have high density infections which may stimulate responses that are distinct from those against low-density asymptomatic infections experienced by immune adults. We hypothesized that there may be different sets of antibody responses most predictive for different age groups. Again, most age-specific classifiers performed at or below that for all samples (Figure 7b).

As no models were able to accurately predict recent exposure across different sites, we hypothesized that there may have been site-specific relationships, due to either biological or technical differences between sites. Site-specific prediction accuracy was modestly higher than the combined data for two sites, Jiangsu and Kampala (Figure 7c).

Although they had similar performance, none of the antigens selected as most important for each site's classifier model overlapped. The Jiangsu and Kampala cohorts were comprised very different participants – Jiangsu is a cohort of adults with limited exposure history, and Kampala is a cohort of children living in a low to medium transmission setting. However, an important similarity between the samples obtained from these sites is that the majority of most recent infections were symptomatic.

To assess whether improved performance of Jiangsu- and Kampala-specific classifiers was driven by the fact that these participants' most recent documented infections (used to calculate TSI) were primarily symptomatic, we stratified the samples by the type of their most recent infection (symptomatic, patent, subpatent) (Figure 7d). Classifiers built from samples whose most recent infection was symptomatic (whether that infection occurred in the previous 6 months or earlier) outperformed other infection types, as well the as all-sample classifier (AUC 0.75, 95% CI: 0.71 to 0.78).

Whereas symptomatic infections were treated upon detection in each of the parent studies, asymptomatic infections were not. Clearance of untreated, asymptomatic P. falciparum infections can last weeks or months,[69] meaning endpoints of asymptomatic infections are further from the onset of infection compared to symptomatic infections. Because TSI was defined as the time since the end of infection, we hypothesized that improved performance of classifiers for samples whose most recent infection was symptomatic was driven by the shift in infection endpoint closer to the onset of infection. To assess this effect, we calculated 3 TSI subtypes for each sample: time since the end of the most recent infection (as before), time since the onset of the most recent infection, and time since the most recent symptomatic infection (even if not the most

recent infection). Classifiers built using each TSI subtype revealed no significant improvement in AUC (Figure 7e).

Discussion

Whereas previous studies have been able to identify sets of antibody responses that can accurately predict recent exposure within specific populations, it is yet to be seen whether seromarkers can be identified that accurately predict exposure across diverse populations. In this study, we generated a database of antibody responses against 888 P. falciparum antigens from 1196 plasma samples selected from 8 different cohort studies representing different age groups, transmission intensities, and geographic settings. Unlike previous studies, we used a novel, stratified sample selection algorithm to account for the confounding relationship between cumulative exposure and Time Since Infection (TSI). Combined antibody responses were only marginally informative in predicting exposure in the previous 6 months across all studies (AUC of 0.69, 95% CI: 0.63 to 0.73), and performance was not higher in any particular transmission intensity or age group. Responses were modestly more informative in predicting recent exposure among individuals whose most recent infection was symptomatic.

Of particular surprise was the lack of accuracy of classifiers even within a given site. A previous study evaluated seromarkers of exposure in Ugandan children from the same cohort in Kanungu as used here using a similar protein microarray. Although no participants from the prior study were included here, the lower performance seen in this study is notable (0.93 vs. 0.62, respectively). One key difference of our study compared to the previous study is the use of stratified sample selection to account for the inherent relationship between TSI and frequency of exposure, which will exist where exposure risk is heterogeneous (i.e. nearly everywhere).[6,70–75] Individuals who are at higher exposure risk will be infected more frequently, which leads to shorter average TSIs and higher levels of cumulative exposure. Without appropriately accounting for this confounding in study design and analysis, associations seen between higher antibody titers and short TSI may, at least in part, be reflective of higher cumulative exposure in these individuals. While such results may be informative within a specific population, they are less likely to be generalizable to populations with different or changing transmission intensity.

The accuracy of our results for participants living in malaria endemic settings (7 out of the 8 cohorts) is contingent upon accurate detection of each participant's infections. In individuals with partial immunity to malaria, this requires active surveillance for infection in addition to passive surveillance for symptomatic malaria. Frequency of active follow up in the 7 cohort studies in endemic areas ranged from every 4 weeks to every 12 months (Table 1). Any asymptomatic infections that occurred and were cleared between

active follow up visits were missed, which may lower the performance of classifiers by generating false positives. The Jiangsu study, for which a site-specific classifier had relatively higher performance, consisted of samples from return travelers living in China, where there is no risk of reinfection. Lack of undetected infections may contribute in part to the improved performance of classifiers for this site.

The ability of antibody responses to accurately predict recent exposure may also be affected by antigen quality. One limitation of the high throughput protein microarray approach is the potential for antigens to lack characteristics of natively expressed protein. Specifically, the *E. coli*-based cell free IVTT system used to create the proteins on the microarray used in this study lacks the proper redox environment for formation of disulfide bonds,[76,77] meaning that proteins may not be folded correctly. Although polyclonal responses may include antibodies against linear epitopes that are not affected by 3D protein structure, it is likely that conformational epitopes were missed.

In summary, these results suggest that antibody responses may not be able to consistently estimate the timing of recent P falciparum exposure across populations in different settings, at least in those with similar characteristics to those evaluated here. Serosurveillance may be more accurate in very low transmission settings due to several factors. Limited pre-existing immunity in very low transmission settings will cause less background due to cumulative exposure. Further, with less immunity, infections are more likely to be symptomatic, which may result in greater boosting and will also make infections easier to detect. Efforts to identify the potential utility of precision serosurveillance for *P. falciparum* may be most fruitful in these settings.

Methods

Study populations

Kanungu, Jinja, and Tororo, Uganda

The Kanungu, Jinja, and Tororo cohorts have been described in detail previously.[78,79] Briefly, Between 2011 and 2016, comprehensive surveillance studies were conducted in three sub-counties in Uganda: Jinja, Kanungu, and Tororo, under the Program for Resistance, Immunology, and Surveillance of Malaria Cohort Study (PRISM).[80] Jinja is a relatively low transmission, peri-urban area near Lake Victoria in the south-central part of the country. Kanungu is a rural area with moderate transmission intensity bordering a national park in the southwestern part of the country. Tororo is a rural area in the southeastern part of the country near to the border with Kenya which had high intensity transmission at the beginning of the study, followed by low transmission due to the success of an Indoor Residual Spraying (IRS) program – spraying households with long

lasting insecticides as a means of malaria control.[81,82] Transmission in all three areas is perennial, with two annual peaks following the rainy seasons. Estimates of the annual entomological inoculation rate at the beginning of the study were 2.71, 20.90, and 175.54 infectious bites per person-year in Jinja, Kanungu, and Tororo, respectively.[83] Participants were recruited from randomly selected households within the catchment area of the participating health facility at each of the three sites. All children aged 6 months to 10 years and a primary adult caretaker in each household were enrolled. Participants agreed to come to the study clinic for any febrile illness and to avoid anti-malarial medications administered outside the study. Study participants attended the clinic at enrollment and then every 4 weeks for routine visits. Every 12 weeks, plasma samples were obtained and subjects were tested for asymptomatic parasitemia via microscopy and Loop-mediated isothermal amplification (LAMP), which is similarly sensitive to PCR.[84–86]

Kampala, Uganda

The Kampala cohort has been described in detail previously.[87–89] Briefly, between November 2004 and April 2009, children aged 1–10 years were surveyed in a randomized trial of combination antimalarial therapies.[90] Subjects were recruited from the Mulago parish of Kampala, Uganda, a densely populated urban slum with an area of approximately one square kilometer. The parish has a large swamp area which serves as a breeding ground for mosquitoes. Malaria exposure in this area is mesoendemic and inversely related to distance of residence from the swamp.[91] Caretakers of study participants were asked to bring their children to a designated study clinic for all medical care. Children underwent routine assessment monthly for febrile illness, and every 12 weeks blood samples were examined via microscopy to identify asymptomatic parasitemia.

Kilifi, Kenya

The Kilifi cohort has been described in detail previously.[92] Between 1998 and 2016, the Kilifi Malaria Longitudinal Cohort study surveilled subjects from Kilifi County at the Kenyan coast.[93–95] Specifically for this project, subjects were from an area within Kilifi called Ngenyera, where malaria transmission was initially moderate but fell to low transmission after 2002.[96,97] Transmission was higher during the rainy seasons, with the long rainy season occurring between May-July and short rainy season between October-December. Children and adult caretakers from randomly selected households were recruited into the study. Children born into the households were then subsequently recruited into the cohort over time. Weekly follow-up visits were performed, where axillary temperature was recorded, and a blood sample was taken for a malaria test by microscopy (before 2006) or PCR (after 2006) only if fever was detected. Additionally, cross-sectional blood samples were collected (whether febrile or not) to detect

asymptomatic parasitemia at various timepoints over the years of the study. Between 1998 and 2005, blood samples were collected at the beginning of the rainy seasons in May and October; in 2006, no samples were collected; in 2007, samples were collected every three months; and from 2008 to 2016, blood samples were collected yearly.[98]

Mugil and Ilaita, Papua New Guinea

The cohorts in Mugil and Ilaita have been described previously.[99–101] The study in Mugil was conducted between January and September 2000 at elementary schools in the Madang region, situated on the northern coast of Papua New Guinea where both Pf and Pv are endemic.[99] The study in Ilaita was conducted in children aged 1 to 4 in the Ilaita area of Maprik District of Papua New Guinea.[100] In both cohorts, children were followed-up for symptomatic infection every two weeks, where blood samples were collected and tested by PCR if fever was present. Every 4 weeks, children were surveyed for asymptomatic infection, where blood samples were collected and tested by PCR regardless of fever status.

Jiangsu, China

Samples from Jiangsu are from a unique cohort of overseas laborers returning to China from 20 countries in West, Central and Southern Africa.[39] Subjects were identified from routine surveillance by China Centres for Disease Control and Prevention. All subjects were diagnosed with falciparum malaria between January 2013 and June 2019, shortly after returning from a malaria endemic country for the purpose of overseas labor. Blood samples were collected between May 2018 and June 2019. Subjects were aged 22-65 at the time of sample collection and had spent between one month and four years working abroad. Since malaria is not endemic to Jiangsu, this cohort represents a unique opportunity to study serological markers that remain detectable several years after exposure.

Microarray generation and probing

Antibodies from serum of all selected samples were quantified using protein microarrays. Arrays contained 1024 printed protein spots, of which 888 were Plasmodium falciparum antigens (TableAllProts) (Figure 1). Antigens were expressed using methods developed by the Felgner lab.[102–104] Details of the procedure have been previously described.[104] Briefly, proteins were expressed in cell-free E. coli-based IVTT reactions from T7 plasmids with N-terminal poly-His and C-terminal hemagglutinin tags for quality control. Whole (unpurified) IVTT reaction printed directly onto microarray slides. In addition to the 888 antigens, each array also contained 28 IVTT control reactions, termed NoDNA spots, that lacked expression plasmids.

Probing the microarrays was done as previously described.[102,105–108] Plasma samples were diluted to 1/100 in protein array blocking buffer (GVS, Sanford, ME, USA) supplemented with E. coli lysate (GenScript, Piscataway, NJ, USA) and incubated at room temperature for 30 minutes. Arrays were rehydrated in blocking buffer for 30 minutes. After removing blocking buffer, diluted plasma was added to the arrays and incubated overnight at 4°C on a rocker. Arrays were then washed with Tris-buffered saline with 0.05% Tween-20 and incubated for 2 hours at room temperature with a solution of 1/2000 anti-human IgG () in blocking buffer. Arrays were then washed with TBST, rinsed with water, and air dried via centrifugation at 500g for 10 minutes.

Prepared slides were scanned at 532 nm using a GenePix® 4000B scanner (Molecular Devices) and results obtained using the GenePix® Pro 7 software (Molecular Devices). In addition to the selected samples, positive and negative control samples were run each day an experiment was done. Positive control samples consisted of pooled serum from highly immune Ugandan adults, and negative control samples consisted of serum from malaria naive American adults.

Data normalization

The median fluorescent intensities (MFI) of the local spot background surrounding each spot was subtracted from the MFI of each antigen spot. To account for variation between arrays, MFIs were log transformed, and the average log-transformed signal from the 28 NoDNA spots (for which IVTT reactions lacked an expression plasmid) on each array was subtracted.

Prediction of recent exposure and evaluation of classifier models

R (R: A language and environment for statistical computing, v4.1.1) was used for data processing and analyses.

Recent exposure was defined as the infection having occurred within 6 months of sample collection. All samples were categorized as obtained from either a “recently infected” or “not recently infected” individual depending on whether or not they were collected within this specified time frame.

To evaluate if the antibody response to any single P. falciparum antigen was informative of recent exposure, binary classification using a threshold antibody level was applied to the data for each of the 20 antigens individually using ROC analysis. Performance of the threshold models was assessed by estimating the 5-fold cross-validated AUC, and 95% confidence intervals were estimated using the method described by LeDell et al.[57]

Table 1: Overview of cohort studies from which samples were selected.

Site	Country	Estimated incidence rate (infections per person year)	Date range	Frequency of active follow up	Infection detection method
Ilaita	Papua New Guinea	1.32	2000	4 weeks	qPCR
Mugil	Papua New Guinea	0.59	2000	4 weeks	qPCR
Kampala	Uganda	1.23 near swamp 0.59 far from swamp	2004-2009	12 weeks	Microscopy
Jinja	Uganda	0.74	2011-2016	12 weeks	LAMP
Kanungu	Uganda	2.28	2011-2016	12 weeks	LAMP
Tororo	Uganda	3.58 before IRS 0.6 after IRS	2011-2016	12 weeks	LAMP
Kilifi	Kenya	2.74 before 2002 0.20 after 2002	1998-2016	Seasonal before 2007 Annual after 2007	qPCR
Jiangsu	China	0	2018-2019	NA	qPCR

Table 2: Overview of selected samples

Site	Country	N samples included	Age, mean (range)	Proportion of samples from individuals infected in previous 6 months	Proportion of samples where most recent infection is symptomatic
Ilaita	PNG	153	3.7 (1.1-6.4)	0.82	0.27
Mugil	PNG	81	9.9 (5.6-12.9)	0.90	0.43
Jinja	Uganda	189	7.4 (2.2-12.7)	0.51	0.84
Kampala	Uganda	206	12.8 (2.0-59.1)	0.51	0.41
Kanungu	Uganda	197	13.3 (2.0-71.7)	0.60	0.59
Tororo	Uganda	195	14.1 (2.1-66.6)	0.49	0.49
Kilifi	Kenya	89	9.1 (2.1-15.6)	0.37	0.55
Jiangsu	China	86	44.6 (22.0-44.5)	0.44	1.00

Figure 1: Microarray layout

Figure 1a

Antigens and control spots were printed in 4 16x16 blocks. Control spots were repeated at different locations throughout the array to account for spatial variation. Teal spots represent P. falciparum antigens and colored spots represent control spots.

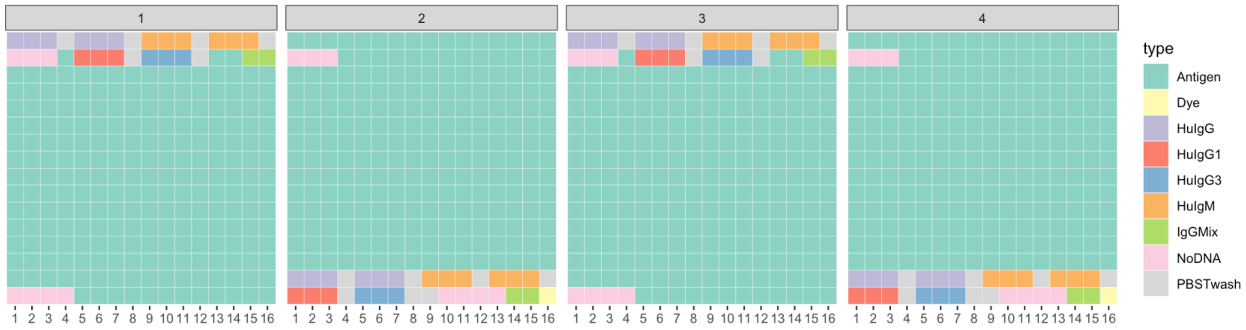


Figure 1B: Example image of scanned slide

Arrays were scanned using a Genepix 4000b microarray scanner at 532nm before probing with plasma to evaluate array printing quality.

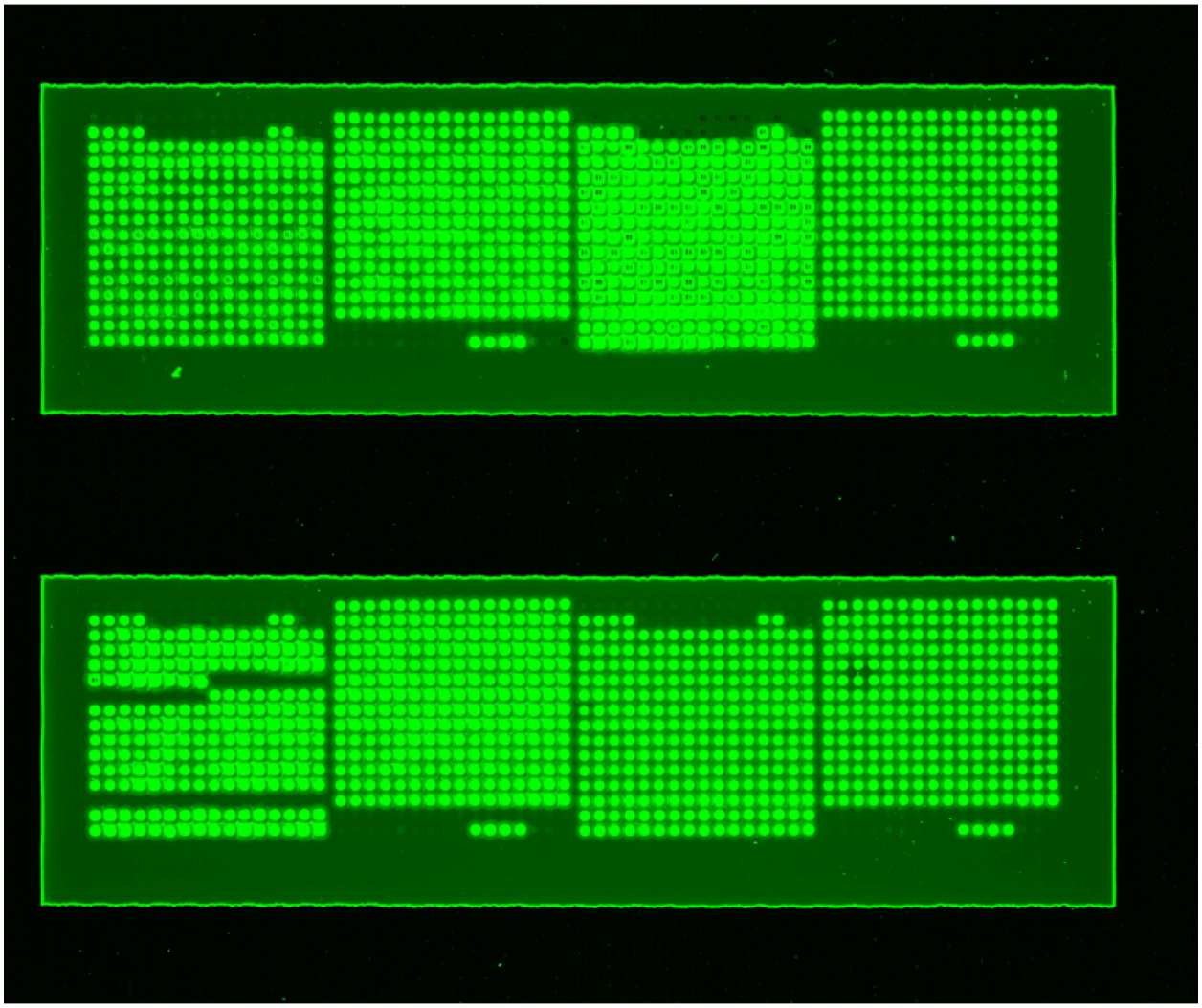


Figure 2: Results from Luminex panel run in parallel show comparable signal to microarray

Results from 89 samples run in parallel on both Luminex and microarray assay. Suffixes for protein names indicate location on the array, as some proteins were printed in duplicate.

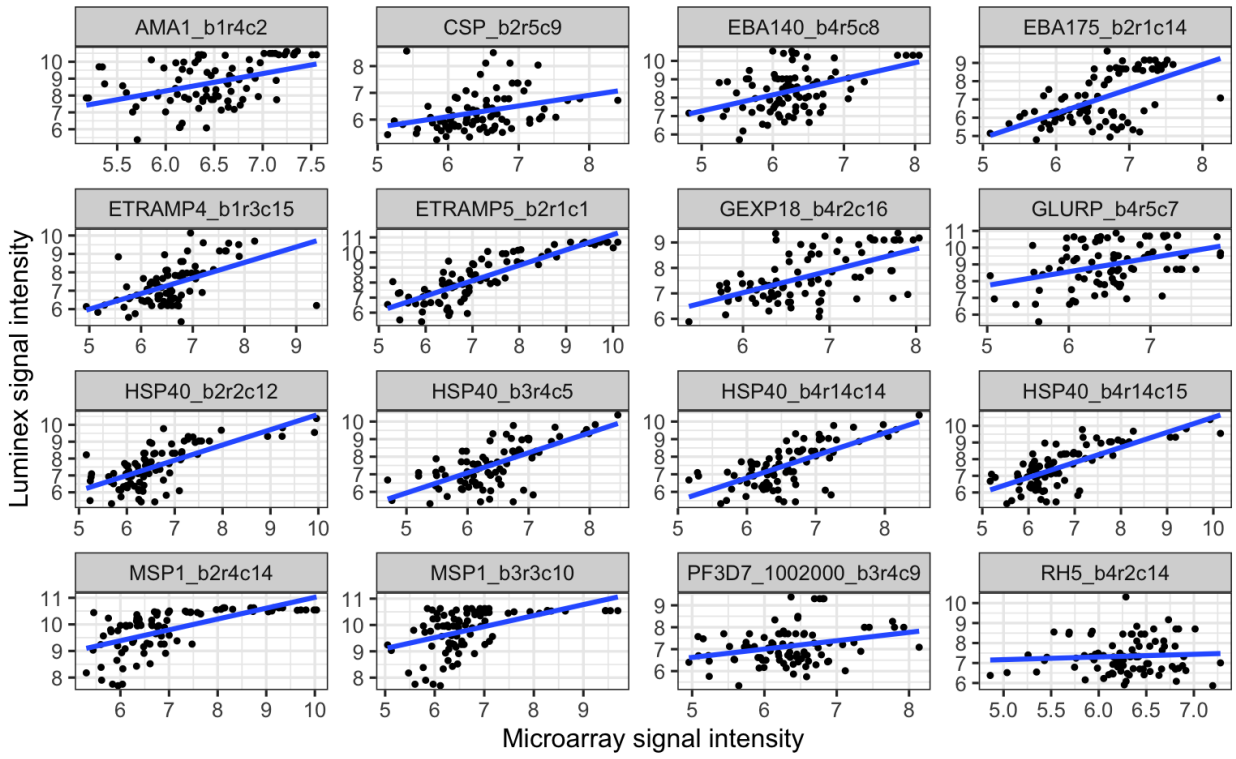


Figure 3: A heatmap of the antibody response to each of the 888 P. *falciparum* antigens included in the protein microarray

Rows correspond to individual antigens while columns correspond to individual samples. Samples are sorted from left to right by site, age category, and increasing TSI. Antigens are sorted from top to bottom by increasing average MFI across all samples.

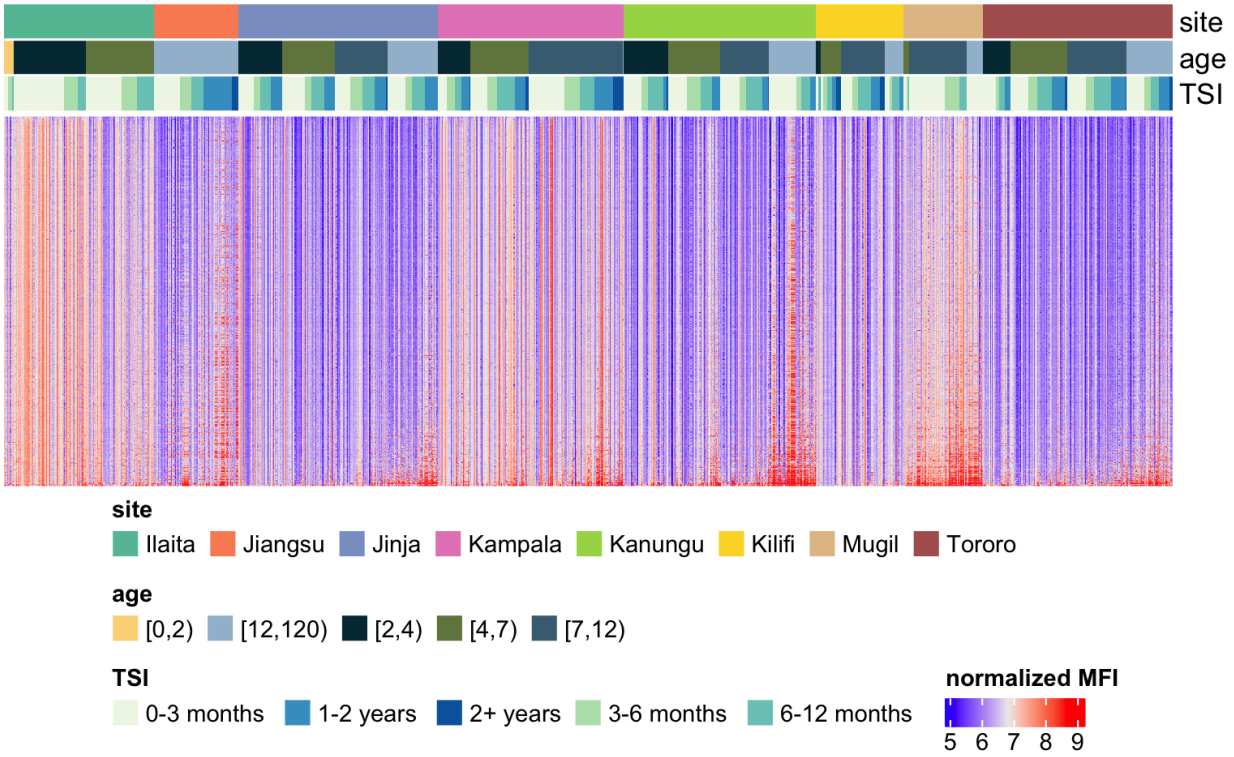


Figure 4: Antibody responses to show distinct relationships with TSI

Figure 4a

Time since infection versus normalized signal for the antigens with the highest (PTP5, REX1, PHISTc), middle (PF3D7_1030800, PF3D7_1466300, PF3D7_1244100) and lowest (RPS18, PF3D7_1355500, PF3D7_1343100) correlation coefficients. Correlation coefficients from left to right: -0.27, -0.24, -0.24 (top row), -0.10. -0.10. -0.10 (middle row), -0.03. -0.01. 0.00 (bottom row).

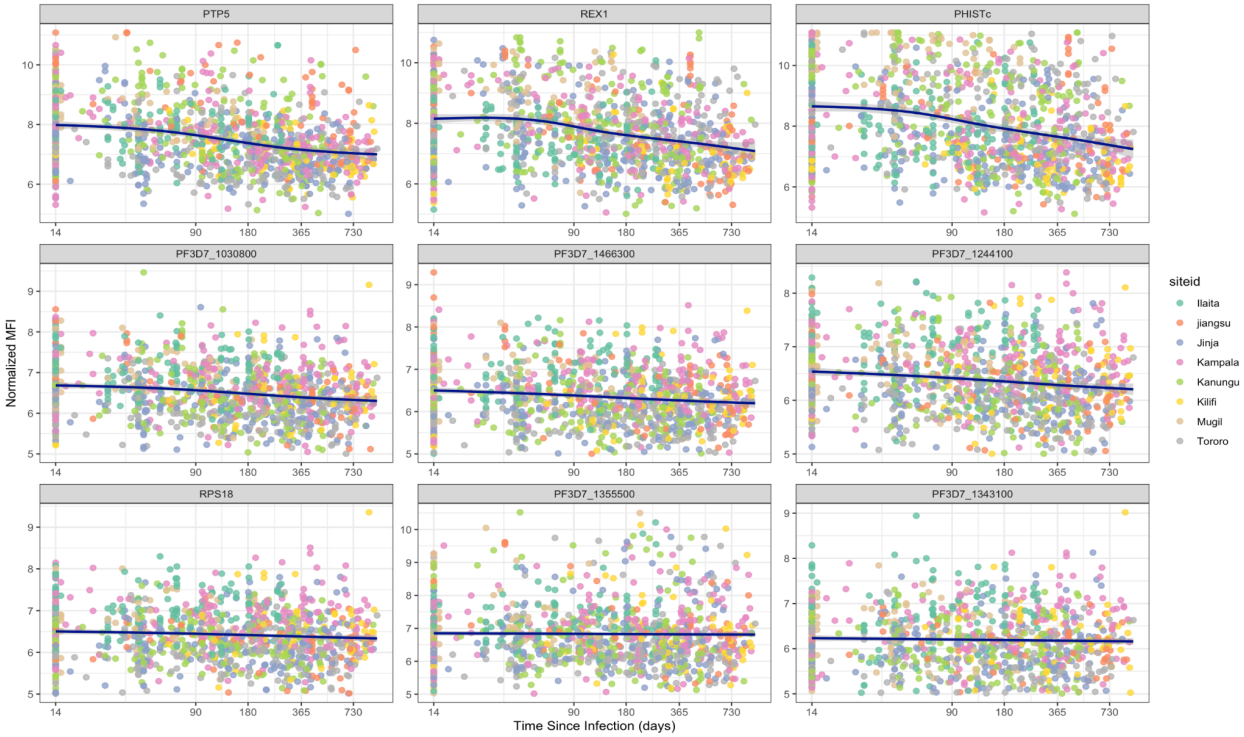


Figure 4b

Boxplot of the breadth of individuals' antibody responses over time. Breadth is expressed as the total number of antigens to which the individual has a normalized signal above that of the mean plus two standard deviations of the naive control samples. The lower and upper hinges of the boxes correspond to the first and third quartiles of the data and the center lines of the boxes represent the median.

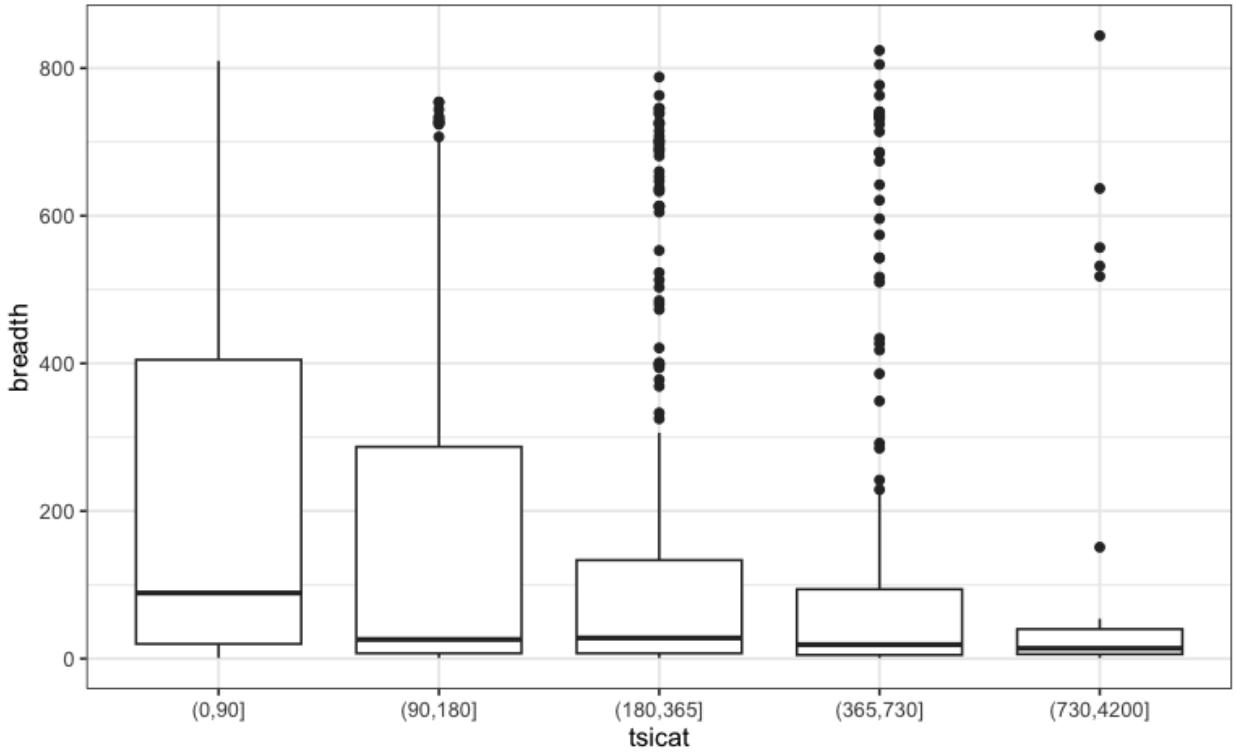


Figure 5: ROC from single antigen threshold models across all sites

Threshold classifiers were built for each of the 888 *P. falciparum* antigens on the protein microarray. Antigens with the 10 highest AUCs are represented in color. Suffixes on protein names indicate location on the array to distinguish duplicate proteins. Tpr stands for true positive rate, fpr stands for false positive rate.

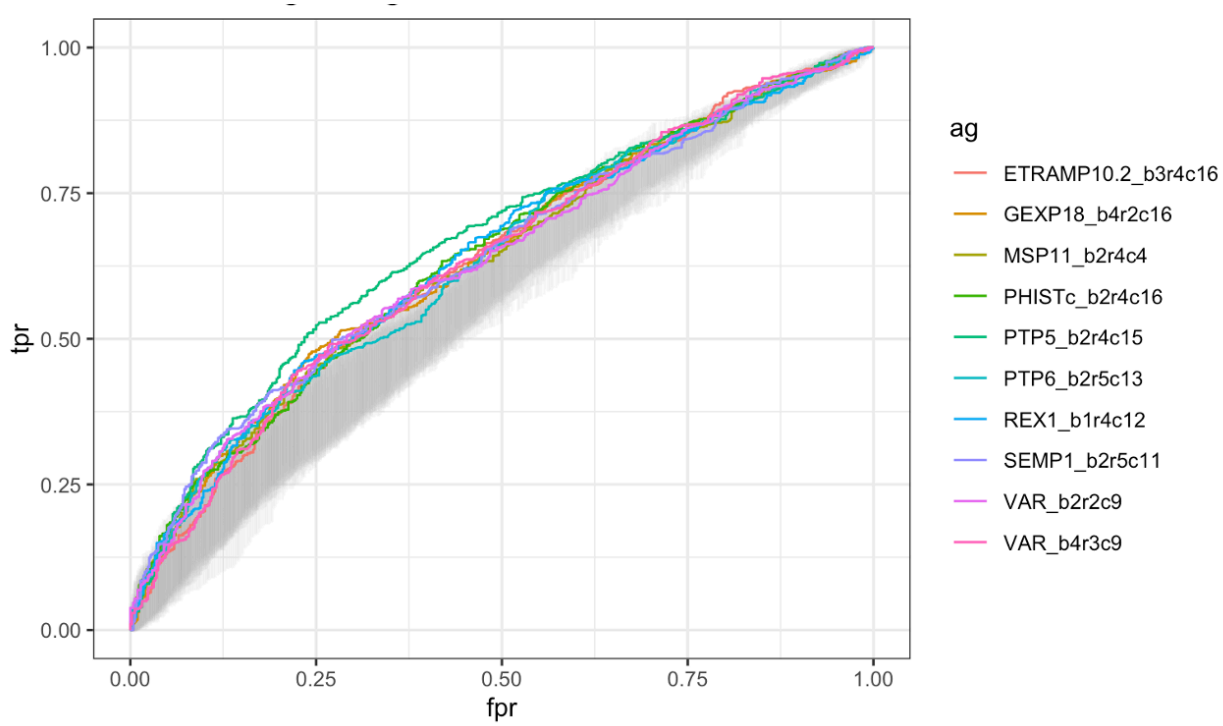


Figure 6: A random forest classifier incorporating all 888 *P. falciparum* antigens shows limited prediction accuracy of recent infection across all samples

Figure 6a

Gray curves represent individual cross-validated ROC curves from each fold of the 5-fold cross-validation of the random forest model. The purple curve represents the overall ROC curve from all folds together. The cross-validated AUC (cvAUC) is 0.68 with a 95% confidence interval of 0.66 to 0.71. Tpr stands for true positive rate, fpr stands for false positive rate.

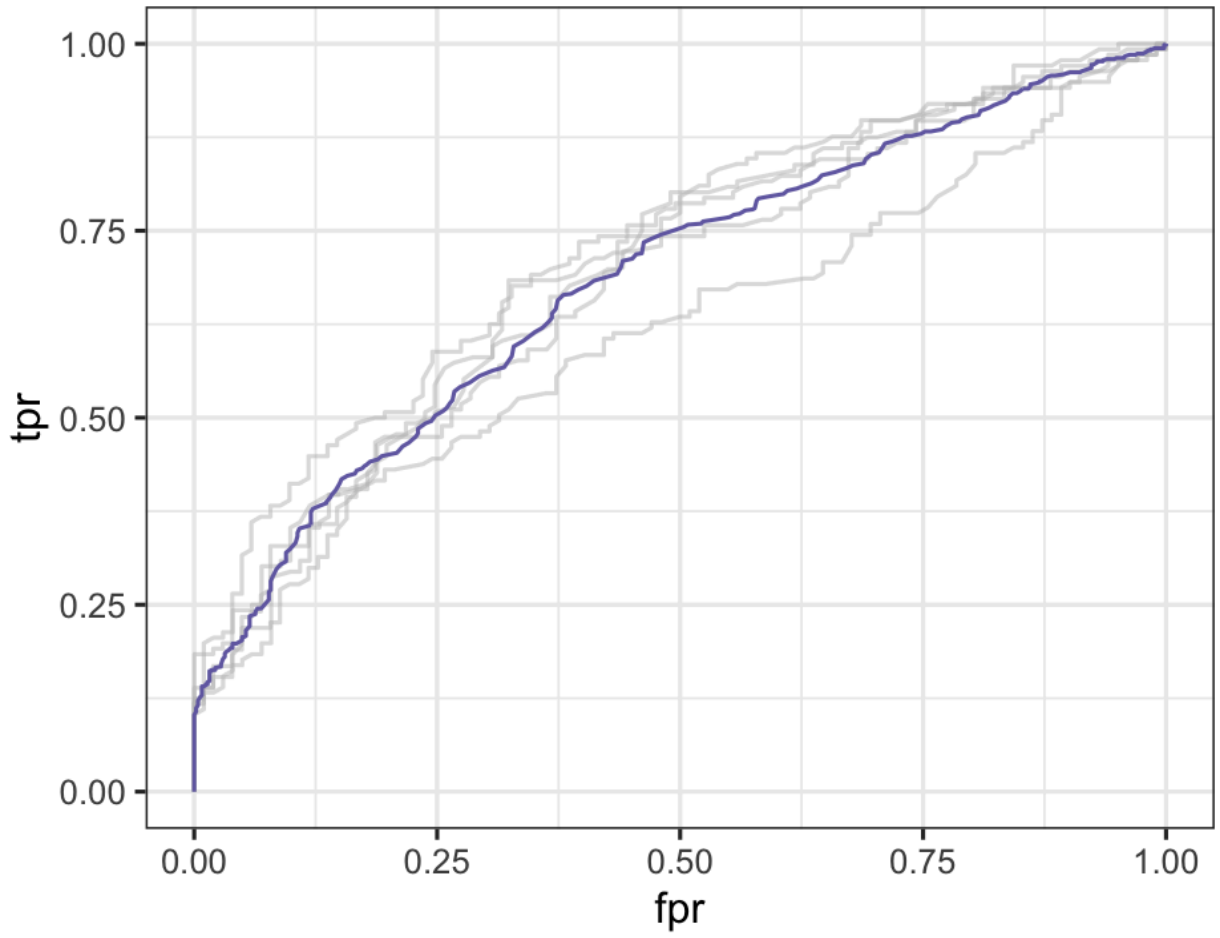


Figure 6b

Distribution of variable importance, as measured by MeanDecreaseAccuracy for the cross-validated ROC in the above plot.

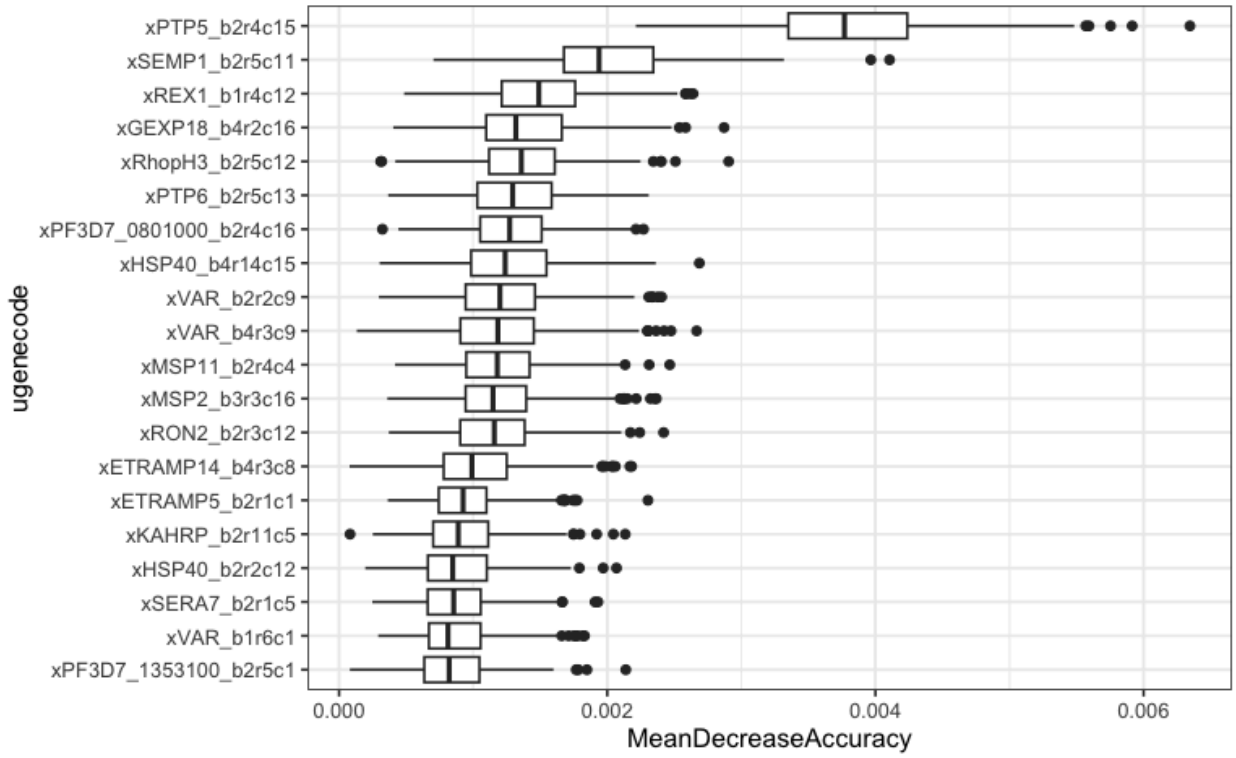


Figure 7: ROC curves for stratified random forest classifier models

Figure 7a

Random forest classifiers were constructed for data stratified by transmission intensity of the parent study site. AUCs are 0.68 (0.65-0.71) across all transmission categories and 0.68 (0.64-0.72), 0.68 (0.64-0.72), 0.62 (0.54-0.0.70), and 0.80 (0.70-0.90) for the [0 - 0.75), [0.75 - 1.5), [1.5-3.6], and traveler transmission categories, respectively. Transmission intensity categories reflect bounds of the estimated incidence rate, calculated as the number of incident infections per person-year follow up.

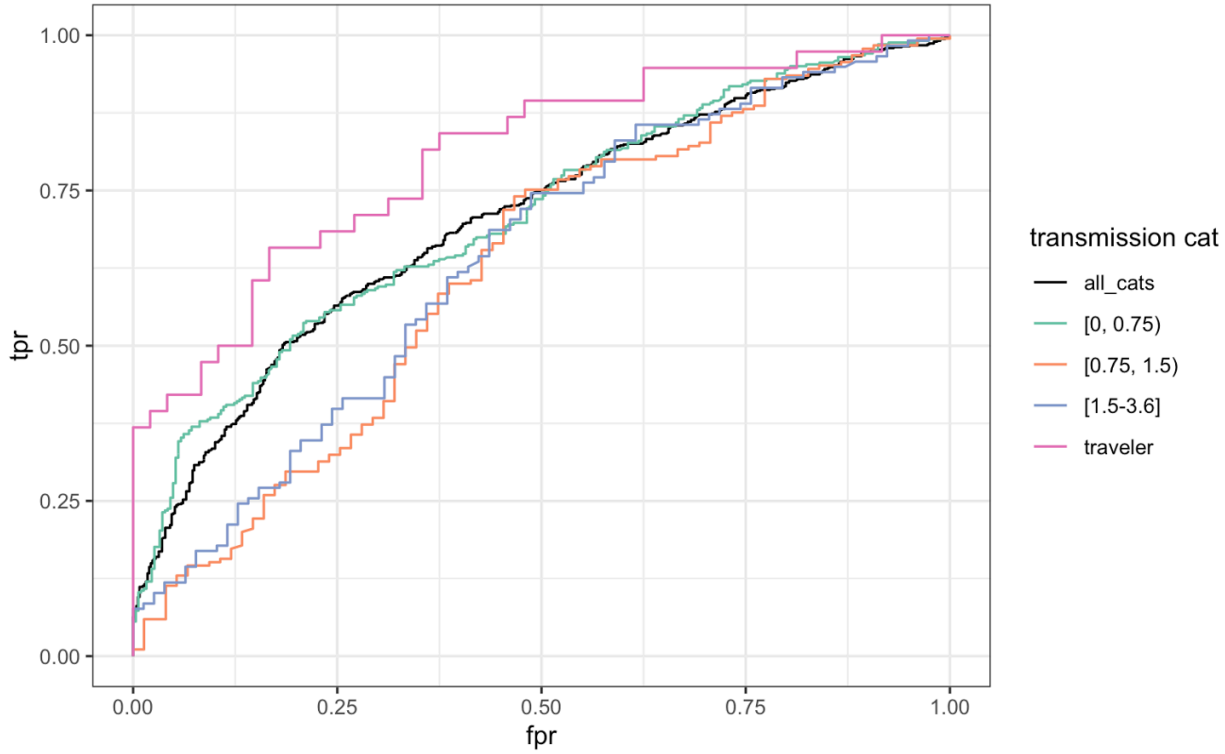


Figure 7b

Random forest classifiers were constructed for data stratified by age category of the participant. AUCs (95% CI) are 0.68 (0.65-0.71) across all age groups and 0.55 (0.48-0.61), 0.74 (0.70-0.78), 0.61 (0.46-0.75), and 0.75 (0.68-0.82) for the (2-5]), (5-15], (15-30], and (30-72) age categories, respectively.

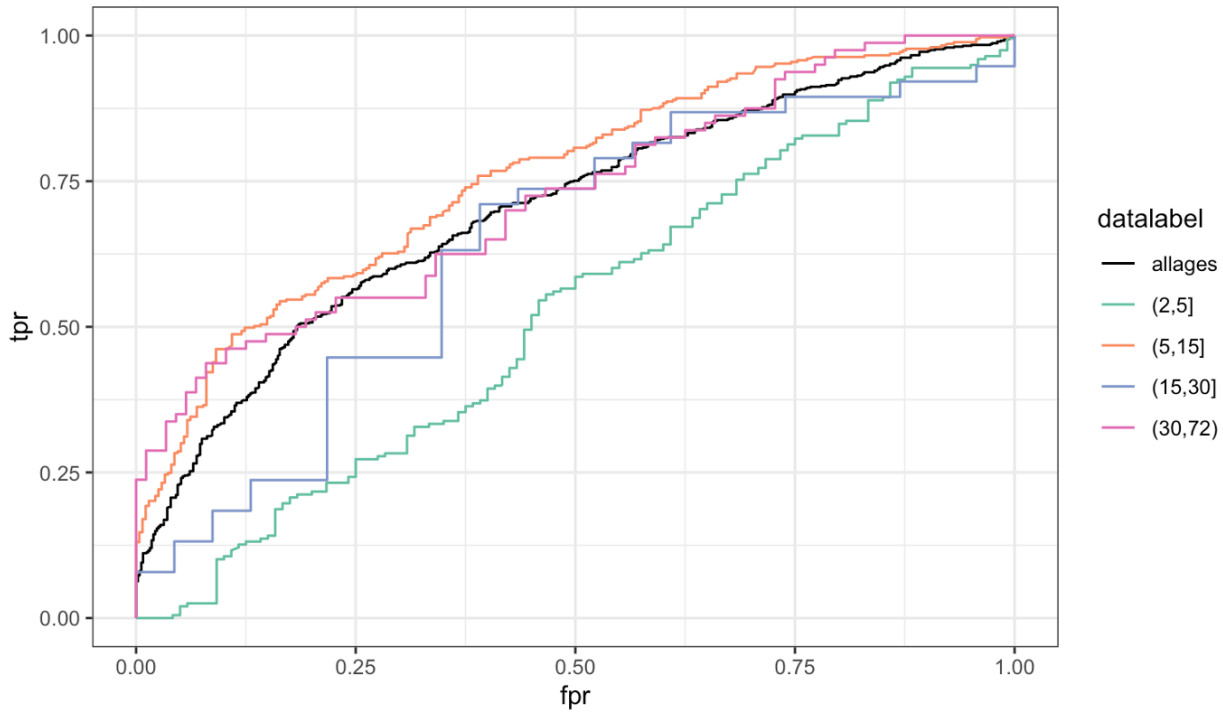


Figure 7c

Random forest classifiers were constructed for data stratified by site of the participant. AUCs (95% CI) are 0.68 (0.65-0.71) across all sites and 0.60 (0.48-0.72), 0.75 (0.58-0.91), 0.74 (0.67-0.81), 0.58 (0.50-0.66), 0.61 (0.54, 0.70), 0.69 (0.61-0.76), 0.55 (0.42, 0.67) and 0.80 (0.70-0.90) for Ilaita, Mugil, Kampala, Jinja, Kanunga, Tororo, Kilifi, and Jiangsu, respectively.

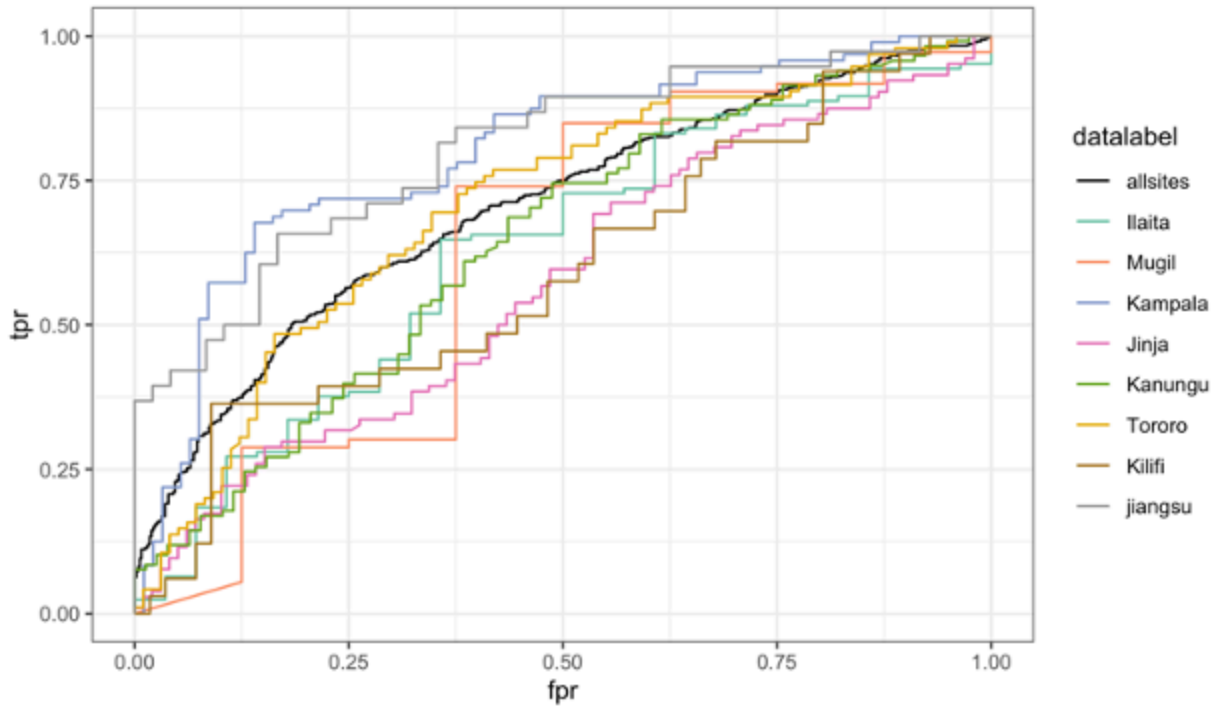


Figure 7d

Random forest classifiers were constructed for data stratified by infection type of the most recent infection experienced by the participant. AUCs (95% CI) are 0.68 (0.65-0.71) across all sites and 0.75 (0.72-0.78), 0.60 (0.50-0.71), and 0.56 (0.50-0.62) for symptomatic (malaria), patent, and subpatent infection types, respectively.

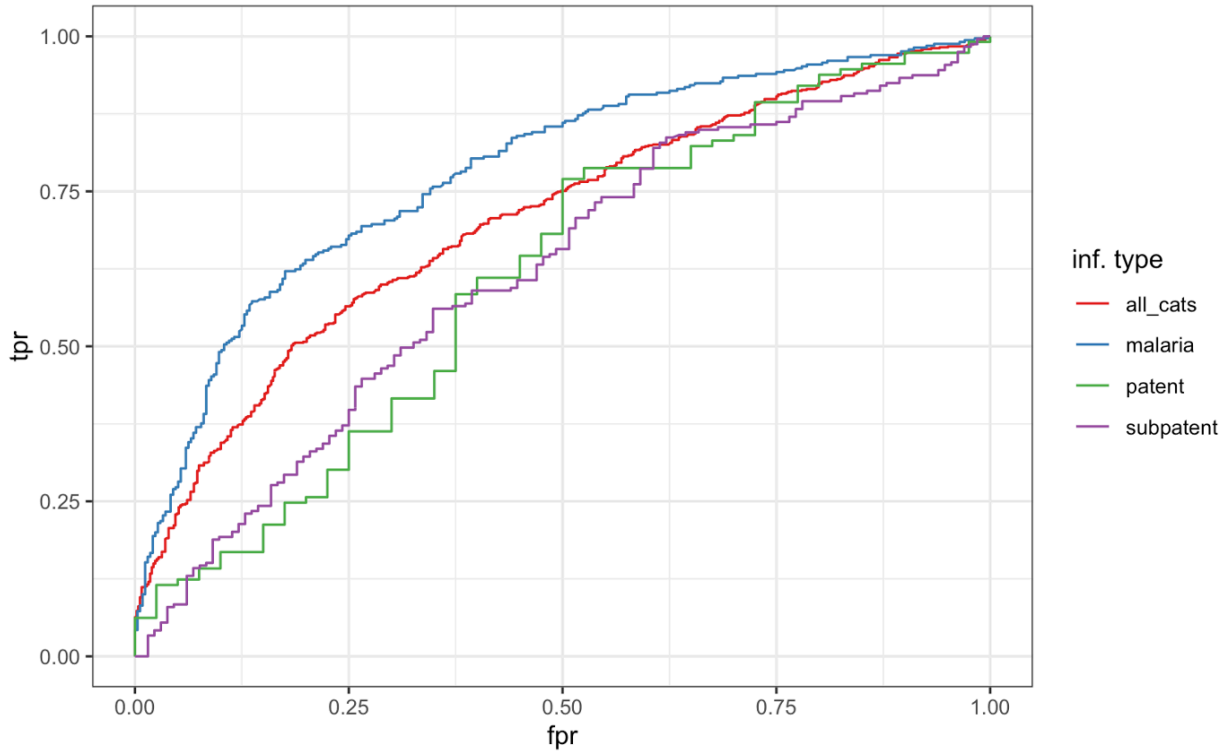
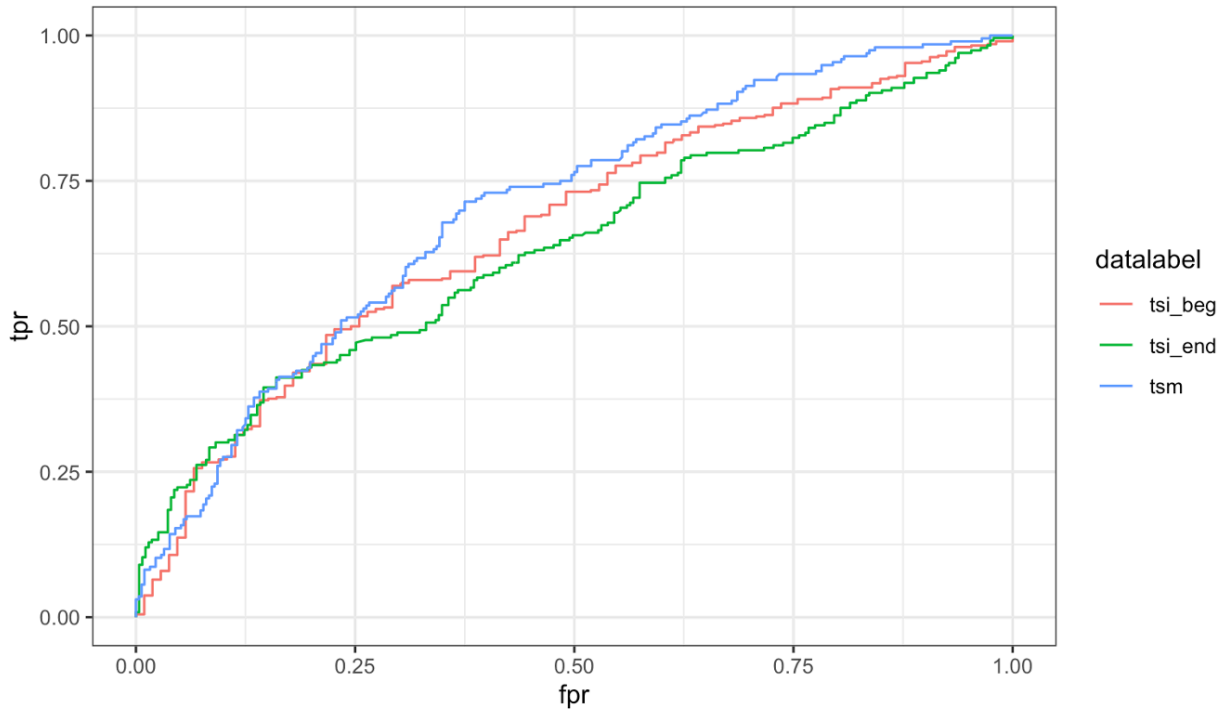


Figure 7e

Random forest classifiers were constructed for data where time since infection was recalculated in 3 types: time since end of infection, time since beginning of infection, and time since most recent symptomatic malaria infection. AUCs (95% CI) are 0.68 (0.65-0.71), 0.67 (0.61-0.72), 0.70 (0.63-0.73) for TSIs calculated from end, beginning and most recent symptomatic infection, respectively.



References

1. WHO | World malaria report 2018. *WHO* http://www.who.int/malaria/publications/world-malaria-report-2018/report/en/.

2. Fact sheet about Malaria. https://www.who.int/news-room/fact-sheets/detail/malaria.

3. CDC-Centers for Disease Control and Prevention. CDC - Malaria - About Malaria - Biology. https://www.cdc.gov/malaria/about/biology/index.html (2019).

4. Rodriguez-Barraquer, I. *et al.* Quantification of anti-parasite and anti-disease immunity to malaria as a function of age and exposure. *eLife* **7**, e35832 (2018).

5. Tran, T. M. *et al.* An Intensive Longitudinal Cohort Study of Malian Children and Adults Reveals No Evidence of Acquired Immunity to Plasmodium falciparum Infection. *Clin. Infect. Dis.* **57**, 40–47 (2013).

6. Rodriguez-Barraquer, I. *et al.* Quantifying Heterogeneous Malaria Exposure and Clinical Protection in a Cohort of Ugandan Children. *J. Infect. Dis.* **214**, 1072–1080 (2016).

7. Jagannathan, P. *et al.* Loss and dysfunction of Vδ2+ γδ T cells are associated with clinical tolerance to malaria. *Sci. Transl. Med.* **6**, 251ra117-251ra117 (2014).

8. Smith, D. L., Guerra, C. A., Snow, R. W. & Hay, S. I. Standardizing estimates of the Plasmodium falciparum parasite rate. *Malar. J.* **6**, 131 (2007).

9. Bhatt, S. *et al.* The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature* **526**, 207–211 (2015).

10. WHO | World malaria report 2019. *WHO* http://www.who.int/malaria/publications/world-malaria-report-2019/en/.

11. Mogeni, P. *et al.* Effect of transmission intensity on hotspots and micro-epidemiology of malaria in sub-Saharan Africa. *BMC Med.* **15**, 121 (2017).

12. Omedo, I. *et al.* Micro-epidemiological structuring of Plasmodium falciparum parasite populations in regions with varying transmission intensities in Africa. *Wellcome Open Res.* **2**, (2017).

13. Bejon, P. *et al.* A micro-epidemiological analysis of febrile malaria in Coastal Kenya showing hotspots within hotspots. *eLife* **3**, e02130 (2014).

14. Kreuels, B. *et al.* Spatial Variation of Malaria Incidence in Young Children from a Geographically Homogeneous Area with High Endemicity. *J. Infect. Dis.* **197**, 85–93 (2008).

15. Greenwood BM. The microepidemiology of malaria and its importance to malaria control. *Trans. R. Soc. Trop. Med. Hyg.* **83 Suppl**, 25–29 (1989).

16. Woolhouse, M. E. J. *et al.* Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. *Proc. Natl. Acad. Sci.* **94**, 338–342 (1997).

17. WHO | Global Technical Strategy for Malaria 2016–2030. *WHO* http://www.who.int/malaria/publications/atoz/9789241564991/en/.

18. Tusting, L. S., Bousema, T., Smith, D. L. & Drakeley, C. Chapter Three - Measuring Changes in Plasmodium falciparum Transmission: Precision, Accuracy and Costs of Metrics. in *Advances in Parasitology* (ed. Rollinson, D.) vol. 84 151–208 (Academic Press, 2014).

19. Greenhouse, B., Smith, D. L., Rodríguez-Barraquer, I., Mueller, I. & Drakeley, C. J. Taking Sharper Pictures of Malaria with CAMERAs: Combined Antibodies to Measure Exposure Recency Assays. *Am. J. Trop. Med. Hyg.* **99**, 1120–1127 (2018).

20. Greenhouse, B. *et al.* Priority use cases for antibody-detecting assays of recent malaria exposure as tools to achieve and sustain malaria elimination. *Gates Open Res.* **3**, (2019).

21. Elliott, S. R. *et al.* Research priorities for the development and implementation of serological tools for malaria surveillance. *F1000Prime Rep.* **6**, (2014).

22. World malaria report 2023. https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2023.

23. Biggs, J. *et al.* Serology reveals heterogeneity of Plasmodium falciparum transmission in northeastern South Africa: implications for malaria elimination. *Malar. J.* **16**, 48 (2017).

24. Noor, A. M. *et al.* Establishing the extent of malaria transmission and challenges facing pre-elimination in the Republic of Djibouti. *BMC Infect. Dis.* **11**, 121 (2011).

25. Reiner, R. C., Jr *et al.* Mapping residual transmission for malaria elimination. *eLife* **4**, e09520 (2015).

26. Global technical strategy for malaria 2016-2030, 2021 update. https://www.who.int/publications-detail-redirect/9789240031357.

27. Tusting, L. S., Bousema, T., Smith, D. L. & Drakeley, C. Measuring Changes in *Plasmodium falciparum* Transmission. in *Advances in Parasitology* (ed. Rollinson, D.) vol. 84 151–208 (Academic Press, 2014).

28. Greenhouse, B. *et al.* Priority use cases for antibody-detecting assays of recent malaria exposure as tools to achieve and sustain malaria elimination. *Gates Open Res.* **3**, 131 (2019).

29. Drakeley, C. *et al.* Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. *Proc. Natl. Acad. Sci. U. S. A.*

102, 5108–13 (2005).

30. Cavanagh, D. R. *et al.* A longitudinal study of type-specific antibody responses to Plasmodium falciparum merozoite surface protein-1 in an area of unstable malaria in Sudan. *J. Immunol. Baltim. Md 1950* **161**, 347–359 (1998).

31. Polley, S. D. *et al.* Human antibodies to recombinant protein constructs of Plasmodium falciparum Apical Membrane Antigen 1 (AMA1) and their associations with protection from malaria. *Vaccine* **23**, 718–728 (2004).

32. Yman, V. *et al.* Antibody acquisition models: A new tool for serological surveillance of malaria transmission intensity. *Sci. Rep.* **6**, 19472 (2016).

33. Helb, D. A. *et al.* Novel serologic biomarkers provide accurate estimates of recent Plasmodium falciparum exposure for individuals and communities. *Proc. Natl. Acad. Sci.* **112**, E4438–E4447 (2015).

34. Yman, V. *et al.* Distinct kinetics of antibodies to 111 Plasmodium falciparum proteins identifies markers of recent malaria exposure. *Nat. Commun.* **13**, 331 (2022).

35. White, M. T. *et al.* Dynamics of the Antibody Response to Plasmodium falciparum Infection in African Children. *J. Infect. Dis.* **210**, 1115–1122 (2014).

36. Tetteh, K. K. A. *et al.* Analysis of Antibodies to Newly Described Plasmodium falciparum Merozoite Antigens Supports MSPDBL2 as a Predicted Target of Naturally Acquired Immunity. *Infect. Immun.* **81**, 3835–3842 (2013).

37. Achan, J. *et al.* Serologic Markers of Previous Malaria Exposure and Functional Antibodies Inhibiting Parasite Growth Are Associated With Parasite Kinetics Following a Plasmodium falciparum Controlled Human Infection. *Clin. Infect. Dis.* **70**, 2544–2552 (2020).

38. Kobayashi, T. *et al.* Distinct Antibody Signatures Associated with Different Malaria Transmission Intensities in Zambia and Zimbabwe. *mSphere* **4**, 10.1128/mspheredirect.00061-19 (2019).

39. Liu, Y. *et al.* Malaria in overseas labourers returning to China: an analysis of imported malaria in Jiangsu Province, 2001–2011. *Malar. J.* **13**, 29 (2014).

40. Wu, L. *et al.* Optimisation and standardisation of a multiplex immunoassay of diverse Plasmodium falciparum antigens to assess changes in malaria transmission using sero-epidemiology. *Wellcome Open Res.* **4**, (2020).

41. Collins, C. R. *et al.* Fine mapping of an epitope recognized by an invasion-inhibitory monoclonal antibody on the malaria vaccine candidate apical membrane antigen 1. *J. Biol. Chem.* **282**, 7431–7441 (2007).

42. Burghaus, P. A. & Holder, A. A. Expression of the 19-kilodalton carboxy-terminal fragment of the Plasmodium falciparum merozoite surface protein-1 in Escherichia coli as a correctly folded protein. *Mol. Biochem. Parasitol.* **64**, 165–169 (1994).

43. Theisen, M., Vuust, J., Gottschau, A., Jepsen, S. & Høgh, B. Antigenicity and immunogenicity of recombinant glutamate-rich protein of Plasmodium falciparum expressed in Escherichia coli. *Clin. Diagn. Lab. Immunol.* **2**, 30–34 (1995).

44. Polley, S. D. *et al.* High levels of serum antibodies to merozoite surface protein 2 of Plasmodium falciparum are associated with reduced risk of clinical malaria in coastal Kenya. *Vaccine* **24**, 4233–4246 (2006).

45. Spielmann, T., Fergusen, D. J. P. & Beck, H.-P. etramps, a new Plasmodium falciparum gene family coding for developmentally regulated and highly charged membrane proteins located at the parasite-host cell interface. *Mol. Biol. Cell* **14**, 1529–1544 (2003).

46. Helb, D. A. *et al.* Novel serologic biomarkers provide accurate estimates of recent Plasmodium falciparum exposure for individuals and communities. *Proc. Natl. Acad. Sci.* **112**, E4438–E4447 (2015).

47. Grüring, C. *et al.* Development and host cell modifications of Plasmodium falciparum blood stages in four dimensions. *Nat. Commun.* **2**, 165 (2011).

48. Raj, D. K. *et al.* Antibodies to PfSEA-1 block parasite egress from RBCs and protect against malaria infection. *Science* **344**, 871–877 (2014).

49. Stone, W. *et al.* IgG responses to Anopheles gambiae salivary antigen gSG6 detect variation in exposure to malaria vectors and disease risk. *PloS One* **7**, e40170 (2012).

50. Triglia, T. *et al.* Identification of proteins from Plasmodium falciparum that are homologous to reticulocyte binding proteins in Plasmodium vivax. *Infect. Immun.* **69**, 1084–1092 (2001).

51. Reiling, L. *et al.* The Plasmodium falciparum erythrocyte invasion ligand Pfrh4 as a target of functional and protective human antibodies against malaria. *PloS One* **7**, e45253 (2012).

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

53. Ragotte, R. J., Higgins, M. K. & Draper, S. J. The RH5-CyRPA-Ripr Complex as a Malaria Vaccine Target. *Trends Parasitol.* **36**, 545–559 (2020).

54. Richards, J. S. *et al.* Association between naturally acquired antibodies to erythrocyte-binding antigens of Plasmodium falciparum and protection from malaria and high-density parasitemia. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **51**, e50-60 (2010).

55. Kastenmüller, K. *et al.* Full-Length Plasmodium falciparum Circumsporozoite Protein Administered with Long-Chain Poly(I·C) or the Toll-Like Receptor 4 Agonist Glucopyranosyl Lipid Adjuvant-Stable Emulsion Elicits Potent Antibody and CD4+ T Cell Immunity and Protection in Mice. *Infect. Immun.* **81**, 789–800 (2013).

56. Yu, W. *et al.* Vaccine-preventable disease control in the People's Republic of China: 1949–2016. *Vaccine* **36**, 8131–8137 (2018).

57. LeDell, E., Petersen, M. & van der Laan, M. Computationally efficient confidence intervals for cross-validated area under the ROC curve estimates. *Electron. J. Stat.* **9**, 1583–1607 (2015).

58. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).

59. Mendis, K. *et al.* From malaria control to eradication: The WHO perspective. *Trop. Med. Int. Health* **14**, 802–809 (2009).

60. Bousema, T. *et al.* Hitting Hotspots: Spatial Targeting of Malaria for Control and Elimination. *PLOS Med.* **9**, e1001165 (2012).

61. Coleman, M. *et al.* Using the SaTScan method to detect local malaria clusters for guiding malaria control programmes. *Malar. J.* **8**, 68 (2009).

62. Moonen, B. *et al.* Operational strategies to achieve and maintain malaria elimination. *The Lancet* **376**, 1592–1603 (2010).

63. Bousema, T. *et al.* The impact of hotspot-targeted interventions on malaria transmission: study protocol for a cluster-randomized controlled trial. *Trials* **14**, 36 (2013).

64. Sturrock, H. J. W. *et al.* Reactive Case Detection for Malaria Elimination: Real-Life Experience from an Ongoing Program in Swaziland. *PLOS ONE* **8**, e63830 (2013).

65. Kusi, K. A. *et al.* Anti-sporozoite antibodies as alternative markers for malaria transmission intensity estimation. *Malar. J.* **13**, 103 (2014).

66. Cook, J. *et al.* Using serological measures to monitor changes in malaria transmission in Vanuatu. *Malar. J.* **9**, 169 (2010).

67. Drakeley, C. J. *et al.* Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. *Proc. Natl. Acad. Sci.* **102**, 5108–5113 (2005).

68. Ondigo, B. N. *et al.* Estimation of Recent and Long-Term Malaria Transmission in a Population by Antibody Testing to Multiple Plasmodium falciparum Antigens. *J. Infect. Dis.* **210**, 1123–1132 (2014).

69. Briggs, J. *et al.* Sex-based differences in clearance of chronic Plasmodium falciparum infection. *eLife* **9**, e59872 (2020).

70. Kang, S. Y. *et al.* Heterogeneous exposure and hotspots for malaria vectors at three study sites in Uganda. *Gates Open Res.* **2**, (2018).

71. Bousema, T., Kreuels, B. & Gosling, R. Adjusting for Heterogeneity of Malaria Transmission in Longitudinal Studies. *J. Infect. Dis.* **204**, 1–3 (2011).

72. Gul, D. *et al.* Investigating differences in village-level heterogeneity of malaria infection and household risk factors in Papua New Guinea. *Sci. Rep.* **11**, 16540 (2021).

73. Bousema, T. & Baidjoe, A. Heterogeneity in malaria transmission: underlying factors and implications for disease control. in *Ecology of parasite-vector interactions* 197–220 (Wageningen Academic, 2013). doi:10.3920/9789086867448_013.

74. Baidjoe, A. Y. *et al.* Factors associated with high heterogeneity of malaria at fine spatial scale in the Western Kenyan highlands. *Malar. J.* **15**, 307 (2016).

75. Bannister-Tyrrell, M. *et al.* Defining micro-epidemiology for malaria elimination: systematic review and meta-analysis. *Malar. J.* **16**, 164 (2017).

76. Yadava, A. & Ockenhouse, C. F. Effect of Codon Optimization on Expression Levels of a Functionally Folded Malaria Vaccine Candidate in Prokaryotic and Eukaryotic Expression Systems. *Infect. Immun.* **71**, 4961–4969 (2003).

77. Protein Expression Handbook - US. https://www.thermofisher.com/us/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/protein-expression-handbook.html.

78. Kamya, M. R. *et al.* Malaria Transmission, Infection, and Disease at Three Sites with Varied Transmission Intensity in Uganda: Implications for Malaria Control. *Am. J. Trop. Med. Hyg.* **92**, 903–912 (2015).

79. Nankabirwa, J. I. *et al.* East Africa International Center of Excellence for Malaria Research: Summary of Key Research Findings. *Am. J. Trop. Med. Hyg.* **107**, 21–32 (2022).

80. Kamya, M. R. *et al.* Malaria Transmission, Infection, and Disease at Three Sites with Varied Transmission Intensity in Uganda: Implications for Malaria Control. *Am. J. Trop. Med. Hyg.* **92**, 903–912 (2015).

81. Kigozi, S. P. *et al.* Rapid shifts in the age-specific burden of malaria following successful control interventions in four regions of Uganda. *Malar. J.* **19**, 128 (2020).

82. Namuganga, J. F. *et al.* The impact of stopping and starting indoor residual spraying on malaria burden in Uganda. *Nat. Commun.* **12**, 2635 (2021).

83. Katureebe, A. *et al.* Measures of Malaria Burden after Long-Lasting Insecticidal Net Distribution and Indoor Residual Spraying at Three Sites in Uganda: A Prospective Observational Study. *PLOS Med.* **13**, e1002167 (2016).

84. Notomi, T. *et al.* Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res.* **28**, e63–e63 (2000).

85. Paris, D. H. *et al.* Loop-Mediated Isothermal PCR (LAMP) for the Diagnosis of Falciparum Malaria. *Am. J. Trop. Med. Hyg.* **77**, 972–976 (2007).

86. Tambo, M., Mwinga, M. & Mumbengegwi, D. R. Loop-mediated isothermal amplification (LAMP) and Polymerase Chain Reaction (PCR) as quality assurance tools for Rapid Diagnostic Test (RDT) malaria diagnosis in Northern Namibia. *PLOS ONE* **13**, e0206848 (2018).

87. Clark, T. D. *et al.* Factors Determining the Heterogeneity of Malaria Incidence in Children in Kampala, Uganda. *J. Infect. Dis.* **198**, 393–400 (2008).

88. Dorsey, G. *et al.* Combination Therapy for Uncomplicated Falciparum Malaria in Ugandan ChildrenA Randomized Trial. *JAMA* **297**, 2210–2219 (2007).

89. Davis, J. C. *et al.* Longitudinal study of urban malaria in a cohort of Ugandan children: description of study site, census and recruitment. *Malar. J.* **5**, 18 (2006).

90. Davis, J. C. *et al.* Longitudinal study of urban malaria in a cohort of Ugandan children: description of study site, census and recruitment. *Malar. J.* **5**, 18 (2006).

91. Clark, T. D. *et al.* Factors Determining the Heterogeneity of Malaria Incidence in Children in Kampala, Uganda. *J. Infect. Dis.* **198**, 393–400 (2008).

92. Bejon, P. *et al.* Stable and Unstable Malaria Hotspots in Longitudinal Cohort Studies in Kenya. *PLOS Med.* **7**, e1000304 (2010).

93. Mwangi, T. W., Ross, A., Snow, R. W. & Marsh, K. Case Definitions of Clinical Malaria under Different Transmission Conditions in Kilifi District, Kenya. *J. Infect. Dis.* **191**, 1932–1939 (2005).

94. Mwangi, T. W. Clinical epidemiology of malaria under differing levels of transmission. (Open University, 2003).

95. Bejon, P. *et al.* A phase 2b randomised trial of the candidate malaria vaccines FP9 ME-TRAP and MVA ME-TRAP among children in Kenya. *PLoS Clin. Trials* **1**, e29 (2006).

96. O'Meara, W. P. *et al.* Relationship Between Exposure, Clinical Malaria, and Age in an Area of Changing Transmission Intensity. *Am. J. Trop. Med. Hyg.* **79**, 185–191 (2008).

97. Muthui, M. K. *et al.* Gametocyte carriage in an era of changing malaria epidemiology: A 19-year analysis of a malaria longitudinal cohort. *Wellcome Open Res.* **4**, (2019).

98. Muthui, M. K. *et al.* Kilifi Malaria Longitudinal Cohort cross-sectional survey and weekly-follow-up surveillance data for the estimation of parasite prevalence and factors associated with gametocyte carriage. Harvard Dataverse https://doi.org/10.7910/DVN/18QB3V (2019).

99. Michon, P. *et al.* THE RISK OF MALARIAL INFECTIONS AND DISEASE IN PAPUA NEW GUINEAN CHILDREN. *Am. J. Trop. Med. Hyg.* **76**, 997–1008 (2007).

100. Lin, E. *et al.* Differential Patterns of Infection and Disease with P. falciparum and P. vivax in Young Papua New Guinean Children. *PLOS ONE* **5**, e9047 (2010).

101. Stanisic, D. I. *et al.* Acquisition of Antibodies against Plasmodium falciparum Merozoites and Malaria Immunity in Young Children and the Influence of Age, Force of Infection, and Magnitude of Response. *Infect. Immun.* **83**, 646–660 (2015).

102. Felgner, P. L. *et al.* Pre-erythrocytic antibody profiles induced by controlled human malaria infections in healthy volunteers under chloroquine prophylaxis. *Sci. Rep.* **3**, 3549 (2013).

103. Davies, D. H. *et al.* Profiling the humoral immune response to infection by using proteome microarrays: High-throughput vaccine and diagnostic antigen discovery. *Proc. Natl. Acad. Sci.* **102**, 547–552 (2005).

104. Doolan, D. L. *et al.* Profiling humoral immune responses to P. falciparum infection with protein microarrays. *PROTEOMICS* **8**, 4680–4694 (2008).

105. Liang, L. *et al.* Identification of Potential Serodiagnostic and Subunit Vaccine Antigens by Antibody Profiling of Toxoplasmosis Cases in Turkey *. *Mol. Cell. Proteomics* **10**, (2011).

106. Nakajima, R. *et al.* Protein Microarray Analysis of the Specificity and Cross-Reactivity of Influenza Virus Hemagglutinin-Specific Antibodies. *mSphere* **3**, 10.1128/msphere.00592-18 (2018).

107. Assis, R. *et al.* Distinct SARS-CoV-2 antibody reactivity patterns in coronavirus convalescent plasma revealed by a coronavirus antigen microarray. *Sci. Rep.* **11**, 7554 (2021).

108. Jain, A. *et al.* Evaluation of quantum dot immunofluorescence and a digital CMOS imaging system as an alternative to conventional organic fluorescence dyes and laser scanning for quantifying protein microarrays. *PROTEOMICS* **16**, 1271–1279 (2016).