Supporting Information

Crompton et al. 10.1073/pnas.1001323107

SI Text

Validity and Reproducibility of the Protein Microarray Assay. To assess the validity of Ab reactivity against proteins expressed in the high-throughput translation system, we included three well-characterized *Pf* proteins being developed as malaria vaccine candidates on the protein microarray: circumsporozoite protein (CSP), apical membrane Ag 1-combination 1 (AMA1-C1), and merozoite surface protein 2 (MSP2). Ab reactivity against these proteins and the corresponding proteins expressed in the high-throughput system correlated [CSP ($r = 0.77, P < 0.001$), AMA-C1 ($r = 0.78, P < 0.001$), and MSP-2 ($r = 0.96$, $P < 0.001$); Fig. S1]. We assessed the reproducibility of the assay by probing two microarray chips printed on separate occasions with Ab against the 3′ hemagglutinin tag. Reactivity against hemagglutinin on all spots for the two chips was highly correlated ($r = 0.92, P < 0.001$; Fig. S1). We further assessed the reproducibility of the assay by constructing a second smaller microarray containing the 49 proteins to which Ab reactivity was associated with protection from malaria. A random subset of plasma samples ($n = 149$) was probed against this smaller array. Ab reactivity against the 49 proteins for the same plasma samples tested on the smaller and larger microarrays was highly correlated ($r = 0.91, P < 0.001$; Fig. S1).

Protein Microarray Chip Fabrication, Ab Profiling, and Data Normalization Procedure. Protein microarray construction is a four-step process that includes (*i*) PCR amplification of each complete or partial *Pf* ORF, (*ii*) in vivo recombination cloning, (*iii*) in vitro transcription/translation, and (*iv*) microarray chip printing. ORFs were derived from the *Pf* genomic sequence database (www.plasmodb.org) and selected based on stage-specific transcription or protein expression (1), subcellular localization in the parasite-infected red blood cell, secondary protein structure, or documented immunogenicity in humans or animal models. Custom PCR primers comprising 20 base pairs of gene-specific sequences with 33 base pairs of "adapter" sequences are used in PCRs with *Pf* clone 3D7 genomic DNA as a template. The 1,204 unique proteins on the array are represented by 2,320 whole or partial proteins, because ORFs >3,000 base pairs were cloned as overlapping segments. We optimized the PCR conditions to increase the efficiency of target sequence amplification using low-temperature annealing and elongation protocols (2). The adapter sequences, which become incorporated into the termini flanking the amplified gene, are homologous to the cloning site of the linearized T7 expression vector pXT7 (3) and allow the PCR products to be cloned by in vivo homologous recombination in competent DH5α cells. We have recently introduced a more efficient method allowing the recombination to occur in vitro with much lower amounts of recombination product to be transformed into *E. coli* (4). The resulting fusion protein also incorporates a 5′ polyhistidine epitope, an ATG translation start codon, and a 3′ hemagglutinin epitope and T7 terminator. Sequence-confirmed plasmids are expressed in 5-h in vitro transcription/translation reactions [rapid translation system (RTS) 100 *Escherichia coli* HY kits; Roche] according to the manufacturer's instructions. Protein expression is monitored either by dot blot or microarray using monoclonal antipolyhistidine (clone His-1; Sigma) and antihemagglutinin (clone 3F10; Roche). Hemagglutinin expression efficiency for this array was >93%. Microarrays are printed onto nitrocellulose-coated glass FAST slides (Whatman) using an Omni Grid 100 microarray printer (Genomic Solutions).

Each microarray chip contains the following controls: (*i*) a "no DNA" negative control in which an empty plasmid vector is placed in the RTS reaction, (*ii*) serially diluted human IgG (a positive control and standard curve to normalize data from arrays probed at different times), and (*iii*) serially diluted EBNA-1 (a positive control given the high prevalence of latent EBV infection in Africa). Actual examples of the microarray chips showing the various control spots are illustrated in Fig. S2.

Venous blood samples were collected from study participants before and after the malaria season, and plasma was stored at −80 °C for analysis by protein microarray. Before incubating with the microarray chip, the plasma samples are diluted to 1/200 in Protein Array Blocking Buffer (Whatman) containing *E. coli* lysate at a final concentration of 30% and incubated at room temperature for 30 min on a rotating platform. The arrays are rehydrated in blocking buffer for 30 min and probed with the pretreated plasma overnight at 4 °C with constant agitation. The slides are then washed five times in Tris (hydroxymethyl)-aminomethane (Tris) buffer, pH 7.6, containing 0.05% (vol/vol) Tween 20 and incubated in biotin-conjugated goat anti-human Ig (anti-IgGfcγ; Jackson Immuno Research) diluted 1/200 in array blocking buffer. After washing the slides three times in Tris buffer containing 0.05% (vol/vol) Tween 20, bound Abs are detected by incubation with streptavidin-conjugated SureLight P-3 (Columbia Biosciences). The slides are then washed three times in Tris buffer containing 0.05% (vol/vol) Tween 20 and three times in Tris buffer without Tween 20, followed by a final water wash. The slides are air-dried after brief centrifugation at 1,000 × *g* for 4 min and analyzed using a Perkin-Elmer ScanArray Express HT microarray scanner. Intensities are quantified using QuantArray software (Packard BioChip Technologies).

It has been noted in the literature that data derived from microarray platforms are heteroskedatic (5–7). This mean-variance dependence has been observed in the arrays presented in this report (8, 9). To stabilize the variance, the vsn method (10) implemented as part of the Bioconductor suite (www.bioconductor.org) in the R statistical environment (www.r-project.org) is applied to the quantified array intensities. In addition to removing heteroskedacity, this procedure corrects for nonspecific noise effects by finding maximum likelihood shifting and scaling parameters for each array, such that the variances of each negative "No DNA" and positive human IgG probes are minimized. This calibration has been shown to be effective on a number of platforms (11–13).

Data Analysis. The probability of an individual experiencing malaria over the 8-month period was estimated by the Kaplan–Meier method, and the time-to-event curves of different age groups were compared by the log-rank test. The strength of association between two continuous variables was determined by calculating the Pearson correlation coefficient (*r*). The characteristics of malaria-protected and malaria-susceptible subjects were compared with Fisher's exact tests and Wilcoxon rank-sum tests for binary and continuous variables, respectively. The Wilcoxon matched-pairs signed-rank test was used to compare measurements of the same parameter at two time points for the same individual. The Kruskal–Wallis test was used to compare three or more unpaired groups. The relation between the combined Ab reactivity to immunogenic *Pf* proteins measure before the malaria season and the risk for experiencing at least one malaria episode was analyzed by the use of a logistic regression model that included age as a covariate. The regression coefficient yielded a point estimate and a 95% confidence interval expressed as the OR for a log increase in Ab reactivity. Individuals with sickle cell trait ($n = 22$) or asymptomatic *Pf* infection before the malaria season ($n = 16$) were excluded from this model, because these factors were associated with decreased malaria risk in this cohort (14). A detailed description of the micro-

array statistical analysis has been published elsewhere (3, 9, 15). The microarray data were normalized and calibrated as described above, and Bayes-regularized *t* tests were then used to identify significant differential Ab reactivity (16). The Benjamini–Hochberg method was used to correct for the false discovery rate. To determine the proportion of immunogenic proteins expressed during the sporozoite and blood stages of the *Pf* life cycle, we retrieved mass spectrometry data (1) from PlasmoDB (www.plasmodb.org). Because many proteins are expressed during more than one stage, we assigned each protein to the stage at which it is maximally expressed according to the Sequest algorithm. For gene ontology analysis, we electronically annotated the entire *Pf* genome by using the default settings of Blast2GO (http://blast2go.bioinfo.cipf.es/home;). From a total of 5,679 genes, we assigned gene ontologies to 3,416. To determine which gene ontology terms were overrepresented among the immunogenic *Pf* proteins relative to the 2,320 proteins on the *Pf* microarray, we used a one-tailed Fisher's exact test with a false discovery rate filter set at $P < 0.05$. With the exception of gene ontology analysis, for all tests, two-tailed *P* values were considered significant if they were <0.05. Statistical analyses were performed with R 2.0 (www.r-project.org) and STATA (version 10.0; StataCorp). Blast2GO (www.blast2go.org) was used for gene ontology annotation and enrichment analysis.

1. Florens L, et al. (2002) A proteomic view of the Plasmodium falciparum life cycle. *Nature* 419:520–526.
2. Su X, Wellems TE (1996) Toward a high-resolution Plasmodium falciparum linkage map: Polymorphic markers from hundreds of simple sequence repeats. *Genomics* 33:430–444.
3. Davies DH, et al. (2005) Profiling the humoral immune response to infection by using proteome microarrays: High-throughput vaccine and diagnostic antigen discovery. *Proc Natl Acad Sci USA* 102:547–552.
4. Beare PA, et al. (2008) Candidate antigens for Q fever serodiagnosis revealed by immunoscreening of a Coxiella burnetii protein microarray. *Clinical and Vaccine Immunology* 15:1771–1779.
5. Ideker T, Thorsson V, Siegel AF, Hood LE (2000) Testing for differentially-expressed genes by maximum-likelihood analysis of microarray data. *J Comput Biol* 7:805–817.
6. Baldi P, Hatfield GW (2002) *DNA Microarrays and Gene Expression: From Experiments and Data Analysis to Modeling* (Cambridge Univ Press, Cambridge, U.K.).
7. Durbin BP, Hardin JS, Hawkins DM, Rocke DM (2002) A variance-stabilizing transformation for gene-expression microarray data. *Bioinformatics* 18 (Suppl 1): S105–S110.
8. Sundaresh S, et al. (2006) Identification of humoral immune responses in protein microarrays using DNA microarray data analysis techniques. *Bioinformatics* 22: 1760–1766.
9. Sundaresh S, et al. (2007) From protein microarrays to diagnostic antigen discovery: A study of the pathogen Francisella tularensis. *Bioinformatics* 23:i508–i518.
10. Huber W, von Heydebreck A, Sültmann H, Poustka A, Vingron M (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* 18 (Suppl 1):S96–S104.
11. Kreil DP, Karp NA, Lilley KS (2004) DNA microarray normalization methods can remove bias from differential protein expression analysis of 2D difference gel electrophoresis results. *Bioinformatics* 20:2026–2034.
12. Barbacioru CC, et al. (2006) Effect of various normalization methods on Applied Biosystems expression array system data. *BMC Bioinformatics* 7:533.
13. Sarkar D, et al. (2009) Quality assessment and data analysis for microRNA expression arrays. *Nucleic Acids Res* 37:e17.
14. Crompton PD, et al. (2008) Sickle cell trait is associated with a delayed onset of malaria: Implications for time-to-event analysis in clinical studies of malaria. *J Infect Dis* 198:1265–1275.
15. Doolan DL, et al. (2008) Profiling humoral immune responses to P. falciparum infection with protein microarrays. *Proteomics* 8:4680–4694.
16. Baldi P, Long AD (2001) A Bayesian framework for the analysis of microarray expression data: Regularized t-test and statistical inferences of gene changes. *Bioinformatics* 17:509–519.

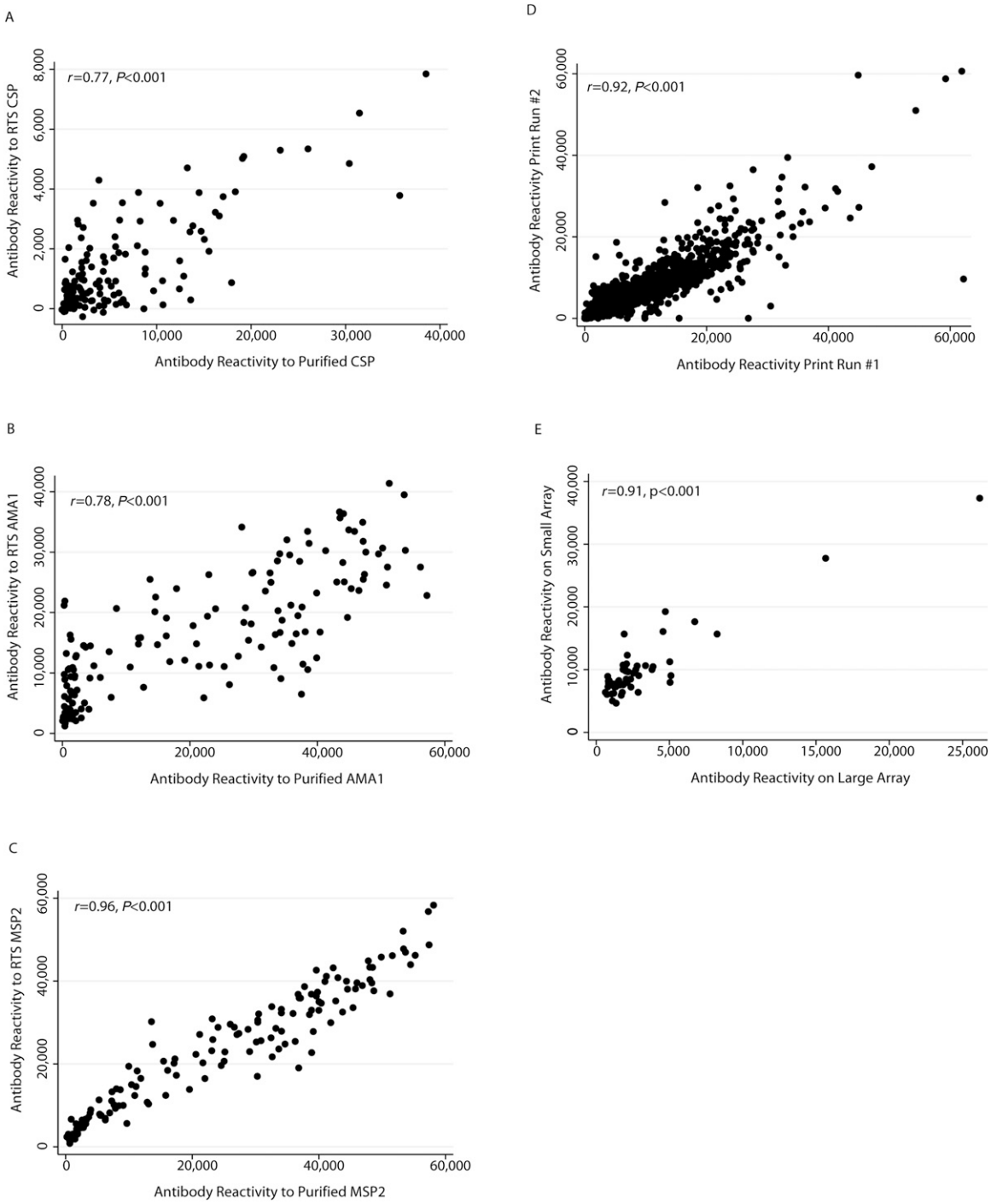

Fig. S1. Validity and reproducibility of the protein microarray assay. To test the validity of Ab reactivity against proteins expressed in the RTS, a subset of randomly selected plasma samples from the study cohort (n = 149) was probed with a microarray containing three well-characterized correctly folded *Pf* proteins and the corresponding proteins expressed in the RTS. Ab reactivity against corresponding protein pairs was correlated for CSP (*A*: $r = 0.77$, $P < 0.001$), AMA-1 (*B*: $r = 0.78$, $P < 0.001$), and MSP-2 (*C*: $r = 0.96$, $P < 0.001$). The reproducibility of the microarray assay was assessed by probing two microarray chips that had been printed on separate occasions with Ab against the 3′ hemagglutinin tag. Reactivity against hemagglutinin on all spots for the two microarray chips was highly correlated (*D*: $r = 0.92$, $P < 0.001$). To assess the reproducibility of the microarray assay further, we constructed a second smaller microarray containing the 49 signature proteins identified on the larger microarray. A random subset of the original plasma samples (n = 149) was probed against this second array and showed that Ab reactivity against the 49 proteins for the same plasma samples tested on the smaller and larger microarrays was highly correlated (*E*: $r = 0.91$, $P < 0.001$). Of note, the displacement in the intercept from zero in *E* is attributable to the difference in laser power.

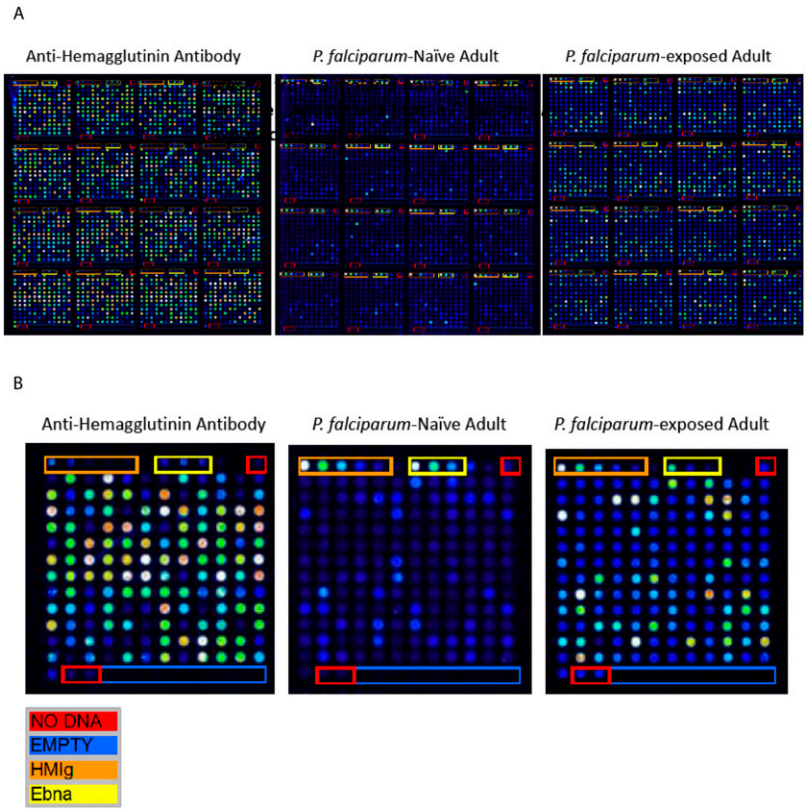

Fig. S2. Representative *Pf* protein microarray chips. (*A*) Representative microarrays containing *Pf* proteins probed with Ab against the 3′ hemagglutinin tag (*Left*), plasma from a *Pf*-naive US donor (*Center*), and plasma from a *Pf*-exposed Malian adult (*Right*). Hemagglutinin expression efficiency for this array was >93%. Ab reactivity to hemagglutinin was considered positive if it exceeded 2 SDs above the negative control (a rapid translation reaction into which an empty plasmid vector is placed). Biotin-conjugated anti-IgG secondary Ab followed by fluorescently labeled streptavidin was used to detect Abs in the samples that were reactive to proteins on the microarray. Plasma from *Pf*-exposed adults had a high degree of Ab reactivity to *Pf* proteins on the microarray (*Right*), whereas a low level of Ab reactivity was detected in *Pf*-naive adults (*Center*). (*B*) Close-up of single protein microarray quadrants showing a high percentage (>93%) of protein spots containing the C-terminal hemagglutinin tag (*Left*), low reactivity in a *Pf*-naive US adult (*Center*), and high reactivity against many antigens in a Malian adult (*Right*). Control spots on the microarray are indicated by colored boxes: red box, "no DNA" negative control; blue box, empty spot; orange box, serially diluted human IgG; yellow box, serially diluted EBNA1.

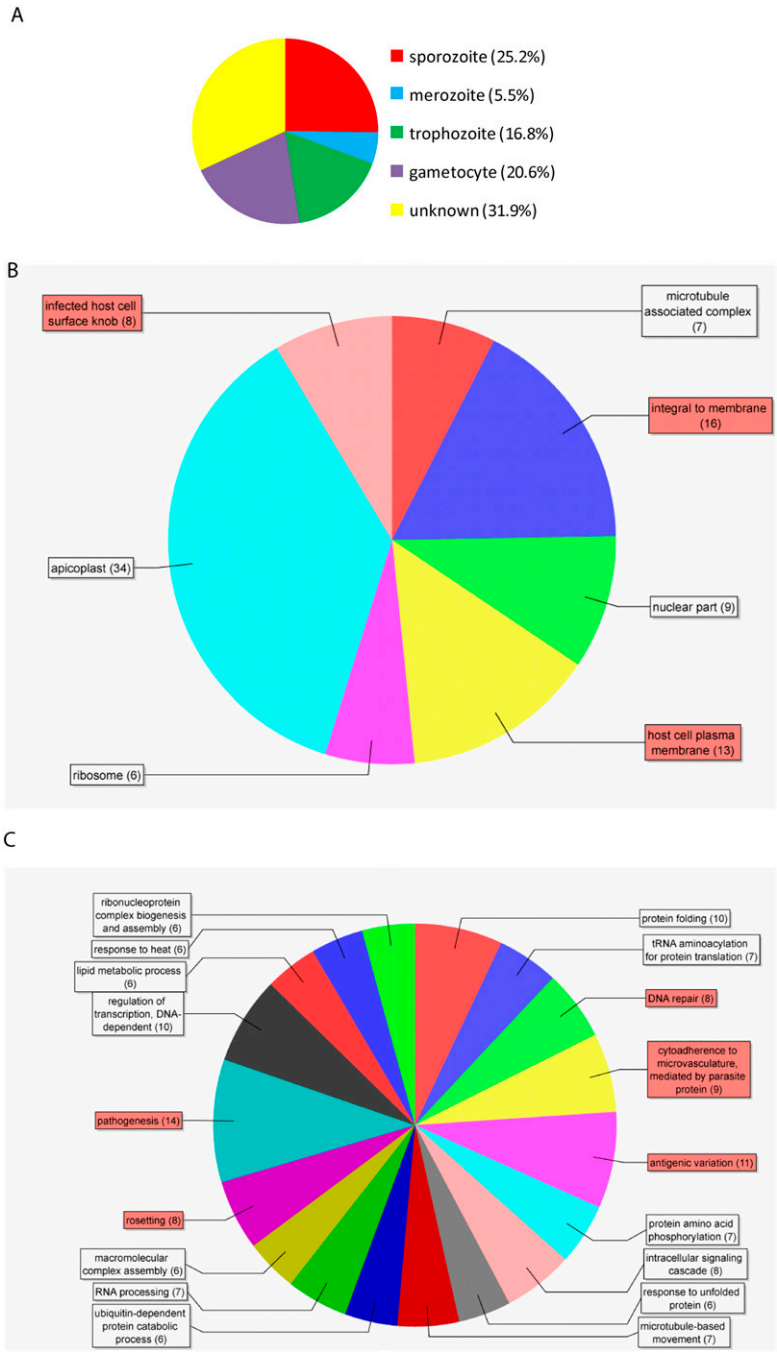

Fig. S3. Stage-specific expression and gene ontology classification of the 491 immunogenic *Pf* proteins. (*A*) Shown is the proportion of 491 immunogenic proteins expressed during the sporozoite and blood stages of the *Pf* life cycle based on mass spectrometry data (1) obtained from PlasmoDB (www.plasmodb.org). Because many proteins are expressed during more than one stage, each protein was assigned to the stage at which it is maximally expressed based on the Sequest algorithm. The cellular component (*B*) and biological process (*C*) gene ontology categories are shown. The size of each slice represents the number of proteins assigned to each category. Categories that are statistically significantly overrepresented relative to the entire microarray are indicated in red. Gene ontology categories with <5 assigned proteins were not included in the figure. The complete results of the gene ontology analysis for immunodominant proteins are given in Table S2.

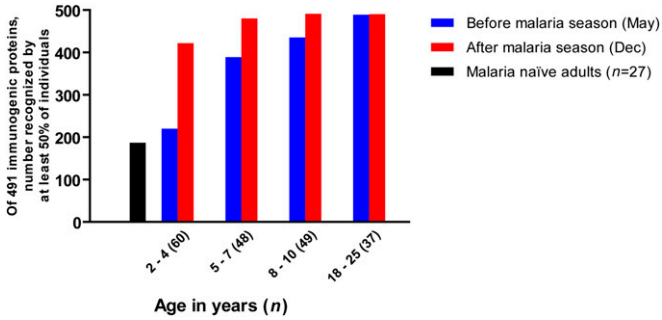

Fig. S4. Number of immunogenic proteins recognized by at least 50% of individuals. Of the 491 immunogenic proteins (2 SDs above the negative control), the number of proteins recognized by at least 50% of *Pf*-exposed individuals increased with age and from before to after the malaria season.

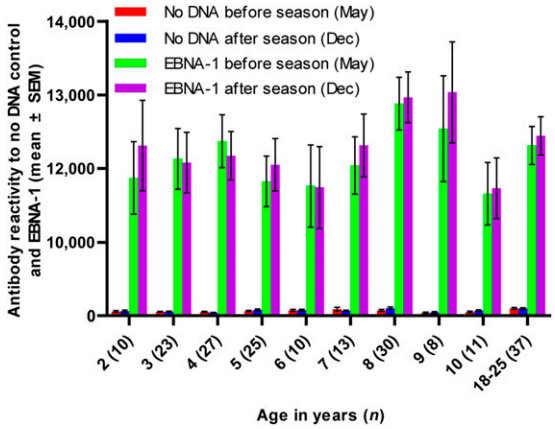

Fig. S5. Impact of age and malaria transmission on Ab reactivity to EBNA-1. Malaria transmission was not associated with consistent changes in Ab reactivity to the EBNA-1, and minimal Ab reactivity was detected in RTS reactions to which empty plasmid vectors were added (no DNA).

Other Supporting Information Files

Table S1 (DOC)
Table S2 (DOC)
Table S3 (DOC)
Table S4 (DOC)