

Protective targets of PfSPZ vaccines identified from whole-genome sieve analysis of isolates from malaria vaccine efficacy trials in West Africa

Ankit Dwivedi^{1,2*}, Ryan J. Scalsky^{1*}, David G. Harris³, Thomas C. Stabler⁴, Biraj Shrestha⁵, Sudhaunshu Joshi⁵, Chakshu Gandhi¹, James B. Munro¹, Olukemi O. Ifeonu¹, Alphonse Ouedraogo⁶, Alfred B. Tiono⁶, Drissa Coulibaly⁷, Amed Ouattara⁵, Thomas L. Richie⁸, B. Kim Lee Sim⁸, Christopher V. Plowe⁵, Kirsten E. Lyke⁵, Shannon Takala-Harrison⁵, Stephen L. Hoffman⁸, Mahamadou A. Thera⁷, Sodiomon B. Sirima⁶, Matthew B. Laurens⁵, Joana C. Silva^{1,2,9 §}

¹Institute for Genome Sciences, University of Maryland School of Medicine; Baltimore, MD 21201, USA.

²Department of Microbiology and Immunology, University of Maryland School of Medicine; Baltimore, MD 21201, USA.

³Department of Computer Science, University of Maryland College Park; College Park, MD 20742, USA

⁴Swiss Tropical Public Health Institute; 4123 Allschwil, Switzerland

⁵Center for Vaccine Development and Global Health, University of Maryland School of Medicine; Baltimore, MD 21201, USA.

⁶Groupe de Recherche Action en Santé; Ouagadougou, Burkina Faso.

⁷Malaria Research and Training Center, University of Sciences, Techniques and Technologies, Bamako; Bamako, Mali

⁸Sanaria Inc.; Rockville, MD 20850, USA.

⁹Global Health and Tropical Medicine (GHTM), Instituto de Higiene e Medicina Tropical (IHMT), Universidade NOVA de Lisboa (NOVA); 1349-008 Lisboa, Portugal

*Authors contributed equally

§ Corresponding author. Email: jcsilva@som.umaryland.edu

Summary: Identification of antigens targeted by a protective response is a central quest in malaria vaccinology. Whole-genome sieve analysis (SA_{WG}) in samples collected from placebo-controlled field trials of *Plasmodium falciparum* (Pf) sporozoite (SPZ) vaccines may enable identification of Pf pre-erythrocytic antigens. We applied SA_{WG} to genomic data generated from Pf isolates collected during two field trials measuring the efficacy, in malaria-exposed African adults, of two PfSPZ vaccines. These randomized, double-blind, placebo-controlled trials were conducted in regions of Mali and Burkina Faso characterized by high seasonal transmission, where parasite genetic diversity is high. Genomic sites in which the vaccine allelic state was significantly underrepresented among breakthrough infections in vaccinees relative to placebo recipients were termed “target sites”. Protein-coding loci containing target sites that changed amino acids were termed “target loci”. The SA_{WG} conducted on clinical trial samples from the Burkina Faso and Mali trials identified 138 and 80 single-copy protein-coding target loci in the Burkinabe and Malian data sets, respectively, with twelve common to both, a number significantly higher than expected ($E = 3.9$; 99%CI = [0, 9]). Among these was the thrombospondin-related anonymous protein locus, which encodes PfSSP2|TRAP, one of the most abundant and well-characterized pre-erythrocytic stage antigen as well as other genes encoding membrane-associated proteins of unknown function. These results identify SA_{WG} as a potentially powerful tool for identifying protective vaccine antigens in recombining pathogens with large genome size and reveals potential new protective Pf antigens.

INTRODUCTION

The World Health Organization (WHO) recommended the very first malaria vaccine, RTS,S|AS01, in October 2021, reflecting decades of research. The formulation of RTS,S|AS01 is based on the variant of *Plasmodium falciparum* (Pf) circumsporozoite protein (CSP) encoded by the 3D7 strain. In October 2023, a second malaria vaccine, R21|Matrix-M, based on the same variant of CSP, was endorsed by the WHO.¹ While this momentous advance will reduce malaria morbidity and mortality, it is clear that neither vaccine will meet the current WHO strategic goal of achieving >90% efficacy against Pf infection.² Furthermore, RTS,S|AS01 has been shown to induce allele-specific protection³ and is, therefore, susceptible to vaccine escape, highlighting the need for additional vaccines to protect against breakthrough infections.

Whole organism vaccines based on Pf sporozoites (PfSPZ) are in late stage development,⁴ and are highly promising as pre-erythrocytic candidate vaccines that can prevent malaria disease and interrupt malaria transmission by preventing parasites from ever leaving the liver. They possess highly desirable features,⁵ including (i) the stimulation of human immunity to multiple Pf antigens present during the pre-erythrocytic phase of parasite development; ii) up to 100% protection against controlled human malaria infection (CHMI) with parasites homologous^{6,7} and heterologous⁸ to the vaccine parasite strain; and (iii) durable efficacy without boosting in malaria-experienced adults in highly endemic settings in sub-Saharan Africa.⁹⁻¹¹ These features represent an extremely high achievement for a malaria vaccine that must overcome a lifetime of immune exhaustion resulting from repeat malaria infection¹²⁻¹⁴ to stimulate a protective antimalarial human immune response. Efficacy in adults living in a high transmission setting with limited ability to respond strongly to a malaria vaccine suggests that PfSPZ vaccines may achieve higher efficacy in pediatric populations with less preexisting immunity to malaria. Nevertheless, the overall lower protection against heterologous relative to homologous CHMI with the same dose of PfSPZ¹⁵ is also consistent with genotype- (or allele)-specific efficacy.

Knowledge of Pf antigens that induce a protective response is required to advance malaria subunit vaccine development and could be used to enhance PfSPZ vaccines. New multivalent subunit vaccines or multi-strain PfSPZ vaccines, where the multiple components

complement each other, may reduce or eliminate vaccine escape and provide broad, strain-transcending protection against malaria. As we recently outlined, allele-specific efficacy, while detrimental for the vaccine efficacy, can potentially be leveraged to identify new protective antigens.¹⁶ Briefly, allele-specific efficacy, by preferentially preventing infection by pathogen strains with a vaccine-like genotype (which renders those parasites immunologically identical to the vaccine), provides a selective advantage to strains that escape vaccine-induced protection. Under this scenario, in placebo-controlled vaccine efficacy field trials, breakthrough infections in vaccinees should differ in allele frequency distribution relative to infections in placebo recipients, with the vaccine allele significantly under-represented among vaccinee infections, which can be formally tested with a sieve analysis.¹⁶ By extension, in the context of a whole organism-based vaccine, a sieve analysis across the whole genome (SA_{WG}) performed on parasite isolates collected from participants in randomized, placebo-controlled, field efficacy trials, by revealing protein-coding loci in which the vaccine allele is under-represented in vaccinees relative to controls, may enable the identification of protective targets. These field trials need to meet several conditions; in particular, i) they need to be conducted in a geographic region of high disease transmission, such that genetic diversity in the parasite population is high; ii) the pathogen should have a high rate of effective recombination, such that genetic linkage is limited to only a few hundred base pairs, and consecutive loci are essentially independent; and (iii) vaccine efficacy (VE) is genotype-specific.

Two such clinical trials were completed in regions with intense, seasonal malaria transmission of sub-Saharan Africa. One assessed the efficacy of PfSPZ Vaccine, a radiation-attenuated whole sporozoite vaccine *versus* saline placebo in 80 adults living in Burkina Faso, West Africa, who were cleared of parasites pre-vaccination.⁹ Efficacy against infection was determined by thick blood smear microscopy during illness and every 4 weeks for two malaria seasons (72 weeks total). VE (1 – hazard ratio) was 48% and 46% after 6 and 18 months of follow-up, respectively (p -value = 0.061 and 0.018). A second study tested PfSPZ-CVac (CQ),¹⁷ a chemoattenuated PfSPZ vaccine in which live, infectious PfSPZ were administered to participants receiving chloroquine to eliminate blood stage infection. Participants were not cleared of parasites before vaccination. VE in this study was assessed by thick blood smear during illness and every 4 weeks for a single malaria transmission season (24 weeks). VE

compared to saline placebo in 62 participants in this study was 34% (p -value = 0.21). Both studies randomized participants 1:1 to receive either the malaria vaccine or placebo. The PfSPZ in both vaccines were from the PfNF54 isolate, the parent stock from which the reference 3D7 was cloned.¹⁸ In protein-coding regions, PfNF54 and 3D7 differ in fewer than 100 non-synonymous sites, mostly in members of multigene families.¹⁹

Here, we report the outcome of SA_{WG} performed on Pf isolates collected from vaccinees and control subjects (placebo recipients) from these two vaccine clinical trials when they first became positive, two or more weeks post vaccination. We identified, in each trial, a set of parasite genes in which vaccine genotypes were significantly under-represented in infections from vaccinees relative to controls and prioritize genes identified in both studies. We describe these genes and show enrichment for features expected from protective antigens, consistent with identification of true protective candidates by SA_{WG}.

RESULTS

Whole genome sequence data and complexity of infection for Pf isolates

Thirty-three participants (13 and 20 in the vaccine and placebo study arms, respectively) in the Burkinabe PfSPZ Vaccine trial were infected at least once with Pf during the follow up period starting two weeks after the last vaccination dose. Total DNA was isolated for each participant's first infection. An average of 43 million Illumina reads were generated per sample (**table S1**). On average, 18 million reads per sample mapped to the parasite genome (range: 65,600 – 30,890,811). Two samples, one each in the vaccine and placebo arms did not reach the desired coverage threshold and were not used further. Complexity of infection, measured by number of clones per infection or by F_{WS} , did not differ significantly between study arms (**fig. S1**). Most samples were polyclonal in both study arms, as expected during the malaria transmission season in a high transmission area; average number of clones per infection was 2.6 and 2.7 in vaccine and placebo recipients, respectively. In the Malian PfSPZ-CVac (CQ) trial, 40 participants (17 vaccinees and 23 who received a placebo) experienced one or more Pf infections during follow up. An average of 60.5 million sequence reads were obtained per sample, and 23 million mapped to the parasite genome (range: 85,469 – 43,912,950) (**table S1**). Four samples

from vaccinees and five from placebo recipients did not pass the coverage threshold; 13 and 18 isolates from vaccinees and placebo recipients, respectively, were used in downstream analyses. Once more, complexity of infection did not differ significantly between study arms (**fig. S2**), and the average number of clones per infection was 3.5 and 3.1 in vaccinees and controls, respectively. A total of 31,122 and 31,659 SNPs in the Pf nuclear genome passed quality filters, respectively, for the Burkinabe and Malian samples (**table S2**) and were used for downstream analyses. As expected for endemic regions of high malaria transmission, the rate of effective recombination in the Burkinabe and Malian Pf populations was high, with linkage disequilibrium between variable sites decaying rapidly to background levels after ~800 bp in both data sets (**fig. S3**).

Identification of target sites and target loci with SA_{WG}

Since PfSPZ Vaccine efficacy is known to be genotype-specific under certain protocols,¹⁵ we conducted a whole-genome sieve analysis (SA_{WG}) on Pf isolates from each trial, to identify parasite proteins that are putative targets of a protective immune response. A SA_{WG}, by revealing genomic sites in which the vaccine allele (here, the allele in PfNF54) is significantly under-represented among vaccinee infections, identifies these “putative protective sites”, or “target sites”.^{16,20} This relies on the rationale that parasites with a vaccine-like genotype at these sites were unable to break through the vaccine-induced protection. Also, we term “target loci” the protein-coding genes that contain non-synonymous target sites, since they encode proteins that are, potentially, the target of a protective response, in a variant-specific manner.

Among the samples from the PfSPZ Vaccine trial in Burkina Faso, 684 significantly differentiated SNPs were identified between vaccinee and placebo recipient Pf isolates, 508 of which were target sites (**table 1**). Of those, 202 were non-synonymous SNPs, which mapped to 165 protein-coding loci or “target loci” (**table 1; table S3**). These loci included several that encode well-established sporozoite antigens, most notably PF3D7_0304600, which encodes CSP, the protein used in the formulation of the RTS,S|AS01 and R21|Matrix-M vaccines, and PF3D7_1335900, which encodes the thrombospondin-related anonymous protein (PfSSP2|TRAP). These are two of the most abundant proteins on the sporozoite surface.²¹ In Mali, 314 target sites were identified. Among those were 110 non-synonymous target sites, mapping to 93 target loci (**table 1; table S4**), including sporozoite antigens PfSSP2|TRAP and

E140 (encoded by PF3D7_0104100), among others. The target sites identified in each SA_{WG} were distributed across all 14 chromosomes (**fig. 1**). Some are unique to each study, while other target sites are identified in both studies (**fig. S4**).

Table 1. Classification of significantly differentiated sites between samples from vaccinees and placebo-recipients. The country and vaccine used in each case are mentioned in the header, for each study. Significantly differentiated sites in which the vaccine allele was underrepresented among vaccinee samples are termed “target sites” (dark blue columns); the loci to which non-synonymous (NSYN) target sites (tan) map are termed “target loci” (brown).

	Burkina Faso, PfSPZ Vaccine trial		Mali, PfSPZ CVac trial	
	Vaccine allele depleted in vaccinee samples (target sites and target loci)	PfNF54 allele depleted in placebo recipient samples	Vaccine allele depleted in vaccinee samples (target sites and target loci)	PfNF54 allele depleted in placebo recipient samples
Significantly differentiated sites	508	176	314	196
Synonymous (SYN)	94	34	47	47
Non-Synonymous (NSYN)	202	60	110	69
Other (intergenic, etc.)	212	82	157	80
Protein-coding loci with differentiated NSYN sites	165	48	93	58

Several factors can lead to potential false positive signals in SA_{WG} (i.e., polymorphic sites identified as “target sites” but that are not the target of a protective response), including differential genotype assortment into study arms by chance (a common challenge for studies with relatively small sample size, such as the two studies here) and genomic sites in linkage disequilibrium with protective sites.^{16,20} These factors can also explain the existence of significantly underrepresented vaccine alleles in placebo recipients (**table 1**, blue columns; **fig. 1**, open blue dots).

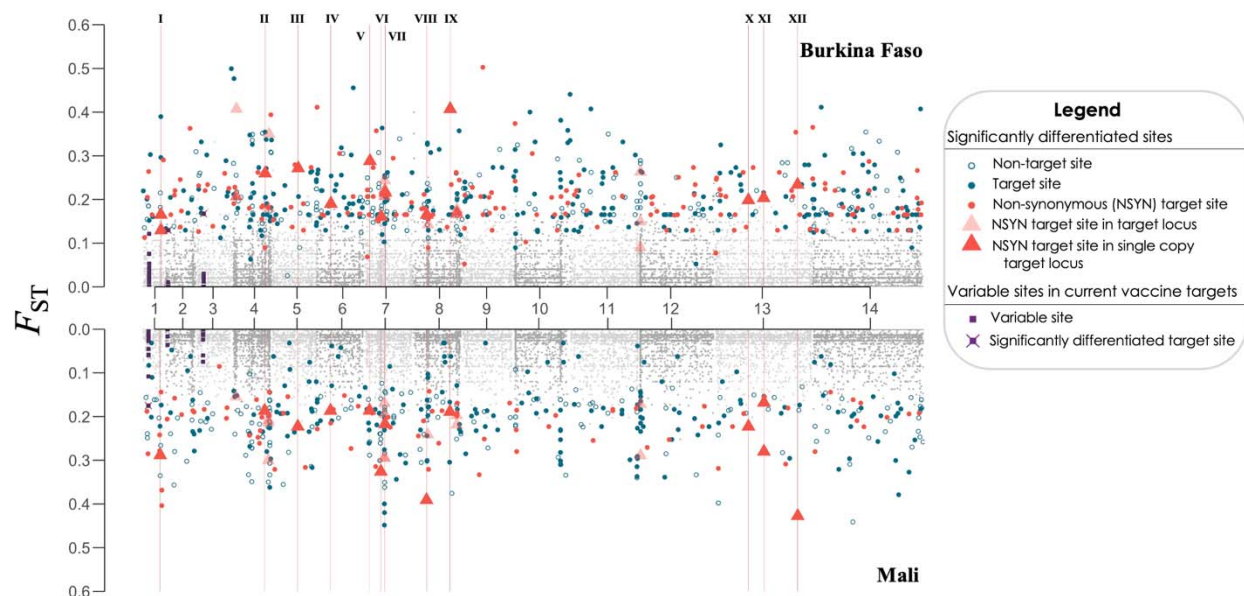


Figure 1. Whole-genome sieve analysis of *P. falciparum* isolates from field trials of two PfSPZ vaccines, in Burkina Faso and in Mali. Site-wise F_{ST} values between placebo and vaccinee parasites are shown for Burkina Faso and Mali vaccine trials, separately. Significantly differentiated genomic sites (permutation method; $k=5000$; $p\text{-value}<0.05$) are shown for each trial (non-grey), including sites where the vaccine allele was underrepresented in isolates in controls (open circles) and those where the vaccine allele was underrepresented among vaccinee isolates (target sites; filled circles and triangles). Non-synonymous target sites (red colors) are further highlighted when they were present in the same target locus in both studies (triangles), and the locus is a member of a multigene family (light red triangle); and when they are observed in same single-copy target locus in both studies (dark red triangle). The latter ($n=12$; shown in roman numerals) are shown across the genome (red, vertical line): PF3D7_0113800 (I), PF3D7_0421700 (II), PF3D7_0518700 (III), PF3D7_0609600 (IV), PF3D7_0703900 (V), PF3D7_0711200 (VI), PF3D7_0713600 (VII), PF3D7_0808100 (VIII), PF3D7_0826000 (IX), PF3D7_1324300 (X), PF3D7_1335900 (XI), PF3D7_1361800 (XII). Variable sites in E140 (in chromosome 1), LSAP2 (in chromosome 2) and CSP (in chromosome 3) are shown (purple). In each of these three loci, one non-synonymous target site is present in one study but not the other.

To determine if these SA_{WG} were likely to have revealed functionally protective proteins, we sought to determine whether the target sites and associated target loci displayed patterns expected from antigens. Several lines of evidence suggest this to be the case, as follows. Analyses of the 262 non-synonymous sites significantly differentiated between vaccinee and control infections in the SA_{WG} of Burkinabe samples revealed that the vaccine allele was significantly depleted primarily among samples from vaccine recipients ($n = 202$; 77.1%) versus placebo recipients ($n = 60$; 22.9%) (table 1; table S5; $p\text{-value} < 0.001$), as expected from the

genotype-specific sieve effect of a vaccine-primed immune system. In the Malian study, the vaccine allele was again significantly depleted more frequently among samples from vaccine (61.5%) *versus* placebo recipients (38.5%) (**table 1; table S5**; p -value < 0.001). Additionally, a GO term enrichment analysis with the 165 target loci identified with the SA_{WG} from the Burkinabe clinical trial samples revealed a significant enrichment of membrane-associated proteins and proteins involved in host-parasite interactions, traits common among antigens (**fig. S5**). A similar analysis of the 93 target loci identified in the analysis of the Malian samples (a study in which VE was not significant) did not reveal GO term enrichment. Enrichment analyses on the loci with significantly differentiated non-synonymous SNPs in which the vaccine allele was depleted in controls did not show significant enrichment of GO categories in either study (**fig. S5**), consistent with the presence of many false positives among these gene sets.

Possibly the most compelling lines of evidence that SA_{WG} identifies *bona fide* protective targets rest with the number and nature of the target loci that were identified in both studies, and in the fact that target sites are immunologically relevant. These aspects are detailed in the next two sections.

Twelve single-copy target loci were common to both studies

Among all target loci identified in each study, 18 were common to both studies (**table 2**), a set larger than expected by chance ($E = 5.0$; p -value $\sim 3.7 \times 10^{-7}$), given the number of loci with non-synonymous variable SNPs across studies ($n = 3326$; **table S2**). In fact, the observed value of 18 shared target loci was found outside of the 99% confidence interval of the expectation (99%CI = [0, 10]). Among these 18 target loci, twelve are single copy genes and the remaining six are members of multigene families. Since SA_{WG} false positives could arise from read mapping artifacts among members of multigene families, we focused further on single copy genes. Given 138 and 80 single copy target loci in the Burkina Faso and Mali studies, respectively, and a total of 2848 single copy genes with non-synonymous variable sites, the probability of finding twelve single copy target loci in common between studies is also significantly higher than expected ($E = 3.9$, 99%CI = [0, 9]; p -value $\sim 3.6 \times 10^{-4}$). Finally, as expected for protective targets of pre-erythrocytic vaccines, transcripts for all twelve single-copy target loci are present in sporozoites, and again during the liver stage, although one is only detected by day 6 (**table 2**).^{22,23} Encouragingly, one of the targets common to both studies is

PfSSP2|TRAP, a well-established pre-erythrocytic antigen, as well as several conserved, membrane-associated Pf genes encoding proteins of unknown function, possibly novel candidate vaccine targets.

Table 2. Target loci identified in both studies. The 18 target loci are shown, each containing at least one non-synonymous (NSYN) target site (i.e., significantly differentiated NSYN SNPs between samples from vaccinees and controls, with the vaccine allele depleted in the former) in both the studies and associated gene expression. Single-copy genes (not shaded) and members of multigene families (shaded rows) are shown.

Gene ID	Gene product	Protein length (AA)	SPZ expression ^c	Liver expression ^d	No. NSYN target sites (BF M) ^a	No. NSYN target sites in CD8+ T cell epitopes (BF M) ^b
PF3D7_0113800	DBL-containing protein	2900	+	- - +	2 1	2 1
PF3D7_0421700	Conserved Plasmodium protein	1936	+	+ + +	1 1	1 1
PF3D7_0518700	mRNA-binding protein PUF1	1894	+	+ + +	1 1	1 1
PF3D7_0609600	Conserved Plasmodium protein	1257	+	- + +	1 1	1 1
PF3D7_0703900	Conserved Plasmodium membrane protein	4358	+	+ + +	2 1	2 1
PF3D7_0711200	Conserved Plasmodium protein	1000	+	- + +	1 1	0 1
PF3D7_0713600	Ribosomal protein S5, mitochondrial, putative	917	+	+ + +	1 1	0 1
PF3D7_0808100	AP-3 Complex subunit Delta	1609	+	- + +	1 1	1 1
PF3D7_0826000	Conserved Plasmodium protein	1172	+	- + +	1 1	1 1
PF3D7_1324300	Conserved Plasmodium membrane protein	5415	+	+ + +	1 1	1 0
PF3D7_1335900	Thrombospondin-related anonymous protein, TRAP ^e	574	+	+ + +	1 2	1 1
PF3D7_1361800	Glideosome-associated connector, GAC	2605	+	- + +	1 1	1 0
PF3D7_0402200	surface-associated interspersed protein 4.1 (SURFIN 4.1), pseudogene	2217	+	- - -	4 1	4 1
PF3D7_0424400	surface-associated interspersed protein 4.2 (SURFIN 4.2)	2380	+	- - +	2 3	1 2
PF3D7_0713000	rifin	346	+	NA	6 4	5 4
PF3D7_0808800	rifin	388	+	- - -	1 1	1 1
PF3D7_0830800	surface-associated interspersed protein 8.2 (SURFIN 8.2)	2049	-	- + +	6 2	4 1
PF3D7_1200600	erythrocyte membrane protein 1, PfEMP1	3056	+	- - +	3 2	2 2

^a Number of non-synonymous (NSYN) target sites in a locus, in each of the two studies; BF: Burkina Faso; M: Mali.

^b NSYN target sites in both studies, overlapping with predicted CD8+ T-cell epitopes corresponding to 22 most prevalent HLA alleles in Burkina Faso (BF) and in Mali (M). ^c Presence of transcripts in the sporozoite stage was classified dichotomously (presence|absence). ^d Liver stage expression is classified into five groups based on CPM values²² (ubiquitously expressed: +++; present by day 4: -++; present by day 6 ---; not expressed in liver: ---; and information not available: NA). ^e TRAP (but none of the other gene products) has a signal peptide.

Target sites are in immunologically relevant regions of target loci

CD8⁺ T cells play a critical role in protection following immunization with Pf sporozoites (SPZ).²⁴⁻²⁶ Therefore, evasion of PfSPZ vaccine-induced protection could be associated with polymorphisms in CD8⁺ T cell-restricted epitopes present in Pf SPZ or liver stage antigens that prevent parasite detection by T cells primed with the Pf NF54 (vaccine) variant, in protective antigen-encoding loci. Accordingly, to determine if the target sites identified with SA_{WG} were immunologically relevant, we predicted the CD8⁺ T cell immunopeptidome of Pf NF54. Since the HLA genotype of individual participants is not available, we defined the CD8 T cell immunopeptidome of Pf NF54 as the union of all non-synonymous sites in NF54 genomic sequences that encode predicted CD8⁺ T cell strong-binding epitopes restricted to the 22 most common HLA alleles in Mali and Burkina Faso. We observed that, in fact, most target sites fall within the CD8⁺ T immunopeptidome (**table 2**), even though enrichment is not significant (**table S6**). The need to use a large HLA allele set to define the immunopeptidome resulted in inclusion of roughly 3/4 of all amino acid residues in the immunopeptidome, possibly reducing statistical power (**table S6**).

We wanted to compare the distribution of target sites in current malaria vaccine pre-erythrocytic targets relative to the immunopeptidome. The vaccine development pipeline for malaria includes twelve pre-erythrocytic antigens, including CSP (PF3D7_0304600) and PfSSP2|TRAP (PF3D7_1335900).²⁷ We found 138 and 135 variable sites among these twelve antigens in Burkinabe and Malian sample sets, respectively, but only a small subset of those are target sites (**tables S3 and S4**). Among samples from Burkina Faso, three of the 202 non-synonymous target sites mapped to three of those twelve antigens, one each in CSP, PfSSP2|TRAP, and LSAP2 (PF3D7_0202100) (**fig. 1; table S3**). Importantly in the context of the current WHO-endorsed malaria vaccines, we note that the target site in CSP (polymorphism in Pf3D7 CSP nucleotide 970 C>A, leading to a Gln324Lys change) falls in the Th2R region, which contains several MHC Class I-restricted epitopes.²⁸ Among Malian samples, two target sites were identified among these twelve vaccine antigens, both of which map to PfSSP2|TRAP (**table S4**).

PfSSP2|TRAP presents as a particularly relevant case study, given its status as a pre-erythrocytic vaccine candidate, and its identification as a target locus in both SA_{WG} studies, since

the vaccine allelic state in three non-synonymous, polymorphic site were significantly under-represented in vaccinees relative to controls (**fig. S6**). Analysis of the distribution of target sites relative to predicted immunopeptidome reveals that the target site identified in the Burkinabe study (829 T>A, leading to Leu277Ile) falls within the predicted and/or validated CD8+ T cell immunopeptidome, as does one of the target sites identified in the Malian study (542 G>T, leading to Arg181Leu) (**table 2, fig. 2**). To investigate the possible contribution of CD8+ T cells to the reduction in frequency of parasites carrying the vaccine allele in vaccinee infections, we considered all peptide variants identified in each study (residue containing the target site plus the eleven residues up- and down-stream from it) and the distribution within them of predicted CD8 T cell epitopes. In the case of target sites in residues 66 and 181, the vaccine peptide variant was the most common variant across all infections, while the target site in residue 277 was the third most common in the Burkinabe study and was rare in Mali (**fig. S7**). In each case, the vaccine variant was more common in infections from controls than from vaccinees, but lack of power from small sample size (and the low frequency of the vaccine variant) precludes significance testing (**fig. S7**). Interestingly, in the Burkinabe study, where the vaccine variant in residue 277 is common to two of the seven variants detected in that study and only that variant is recognized by a common HLA allele, the vaccine variant is under-represented in vaccinee infections relative to those in controls (**fig. 2**), suggesting that recognition, by CD8⁺ T cells, of cells infected with Pf this genotype could in fact contribute to their elimination from the vaccine arm.

All three target sites found in PfSSP2/TRAP, including the site in the codon encoding the variant Asn66Lys (target site 198 C>G), fall within the two adhesive domains required for host cell adhesion (**fig. 2**).^{29,30}

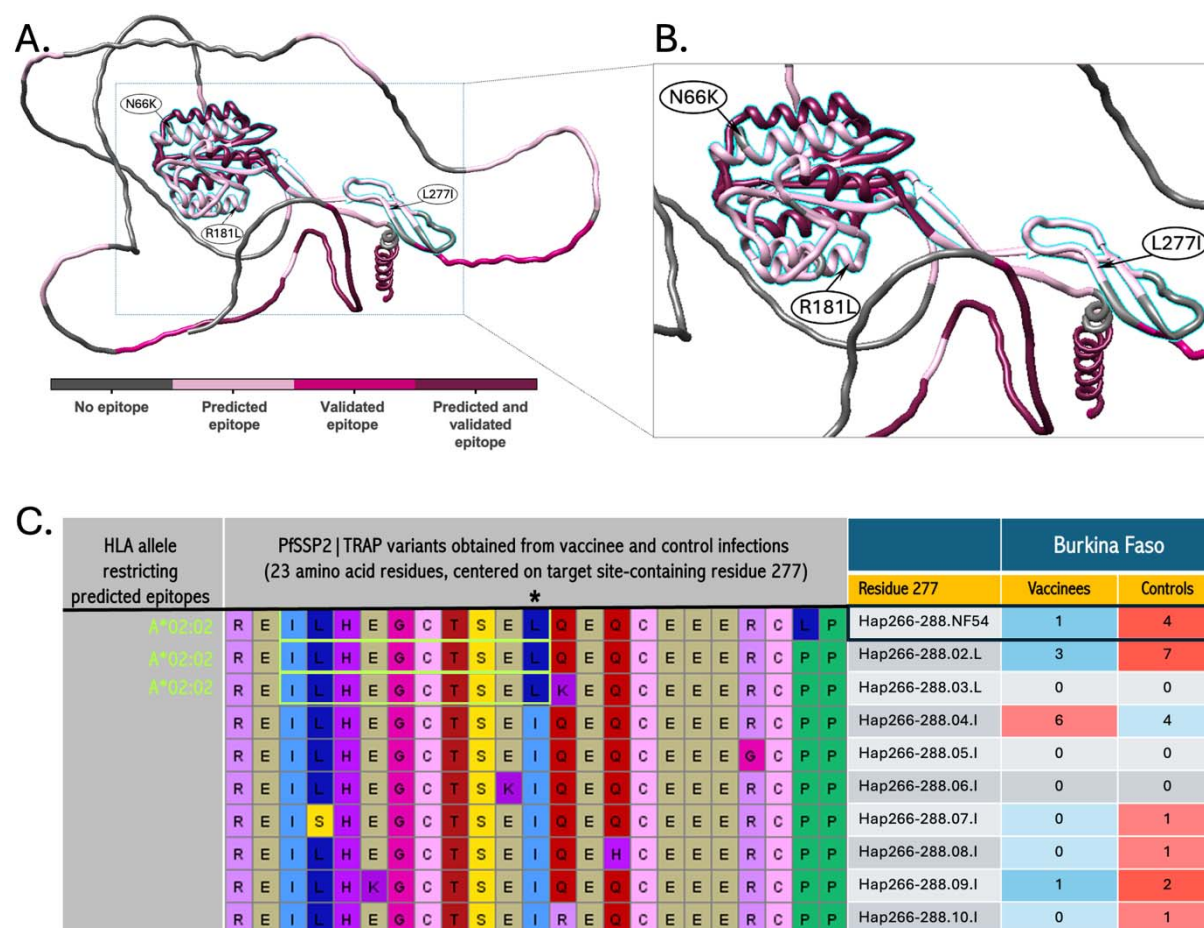


Figure 2. Genetic variation and distribution of predicted CD8⁺ T-cell epitopes at PF3D7_1335900, the locus encoding PfSSP2/TRAP. **A.** 3D structure of TRAP, obtained with AlphaFold.³¹ Amino acid residues containing target sites are noted, with residue coordinate (L66I; R181L; L277I). Magenta color spectrum shows protein regions with predicted strong binding CD8⁺ T-cell epitopes, experimentally validated epitopes or both. Display generated with UCSF Chimera v1.15.³² **B.** Magnification of protein regions containing target sites. Shading (in light blue) encapsulates structurally important motifs required for adhesion and invasion of host cells, including a von Willebrand factor A (vWA)-like fold (top left)²⁹ and a thrombospondin type 1 repeat (TSR) domain (center right).³⁰ **C.** Alignment of TRAP peptide sequences at and surrounding residue 277 (central panel; amino acid residues 266-288), with location of epitope restricted to HLA A*02:02 marked (green box). Frequency of each variant identified in Pf infections from vaccinees or controls is shown (right panel). Vaccine peptide in black frame. Color-coding: same color in both arms when a variant was found in equal number in vaccinees and controls; strong red/blue when frequency was double of more (red) in one arm than the other (blue); light blue/red otherwise.

DISCUSSION

In 2001, Gilbert and colleagues³³ proposed an approach they termed “sieve analysis” to determine if VE is allele(genotype)-specific, and to quantify the relative risk of infection, post-vaccination, associated with strains encoding different variants of a vaccine immunogen. Our group conducted the first comprehensive sieve analysis in malaria to identify residues and

epitopes associated with protection post-vaccination with the apical membrane antigen 1 (AMA1) variant encoded by 3D7, and in repeat infections after natural exposure to Pf malaria, and identified several key residues important for protection.³⁴⁻³⁶ More recently, we conducted a genome-wide comparative study of genetic variation between groups of individuals with differing levels of acquired immunity, to identify potential Pf protein targets of acquired immunity to clinical malaria.³⁷ Merging the concepts of sieve analysis and genome-wide differentiation studies between Pf genomes from phenotypically different groups of individuals, we recently proposed SA_{WG}, an extension of the sieve analysis concept to the whole genome of a genetically diverse, recombining parasite species, such as Pf, using samples from a placebo-controlled vaccine trial; by detecting genomic sites in which the vaccine allele is significantly under-represented in vaccinee Pf infections relative to those from placebo recipients, a SA_{WG} should thereby identify targets of vaccine-induced protection.¹⁶ Here, we conducted the first SA_{WG} on Pf isolates collected from VE trials of PfSPZ vaccines conducted in malaria-endemic areas, with the explicit purpose of identifying potential targets of PfSPZ vaccine-induced protection. The SA_{WG} conducted on the Burkinabe study isolates revealed 165 target loci (138 of which are single-copy genes) while the Malian SA_{WG} revealed 93 target loci (80 single-copy genes). A total of 18 target loci were identified in both analyses, including twelve single-copy genes, numbers significantly higher than expected by chance.

Other lines of evidence are consistent with the possibility that several of these twelve target loci may be targets of PfSPZ vaccine-induced protection. Most notably, one of twelve single copy targets, PF3D7_1335900, encodes one of the most well-established pre-erythrocytic sporozoite stage antigens, PfSSP2|TRAP.³⁸ Also, PF3D7_0113800 encodes a DBL domain known to activate Rho GTPases involved in hepatocyte invasion during Pf infection,^{39,40} and PF3D7_1361800 encodes the glideosome-associated connector (GAC), a crucial component of the gliding machinery that enables host tissue transversal, cell invasion and egress.⁴¹ GAC connects actin fibers to surface adhesins (including TRAP), playing a pivotal role in parasite forward movement. Interestingly, GAC becomes exposed and is shed as parasites migrate through host tissue,⁴¹ providing an opportunity for immune system priming and subsequent parasite detection. PF3D7_0703900 and PF3D7_1324300 encode conserved, membrane-associated proteins of unknown function, but may be accessible to either cell-mediated or humoral immunity if present on the sporozoite surface. PF3D7_0518700 encodes the mRNA-

binding protein PUF1, which may regulate stability and translation efficiency of specific mRNAs important for gametocyte development.⁴² PF3D7_0808100 encodes AP-3 complex subunit delta, with a role in endosomal protein trafficking and vesicle formation within the parasite.^{43,44} PF3D7_0713600 encodes a putative mitochondrial ribosomal protein S5. Even though the relationship between some genes identified by SA_{WG} and vaccine-induced protection is not immediately clear from the localization of the encoded proteins, surface exposure of targeted proteins is not a requirement for CD8⁺ T cell-effected protection. The SA_{WG} revealed four additional parasite genes encoding conserved proteins of uncharacterized function (**Table 2**). One of these, PF3D7_0421700 has been identified as a likely target of naturally acquired immunity to clinical malaria,³⁷ suggesting an overlap between protective targets of sporozoite-based vaccination and natural infection. Finally, within these 12 target loci, most non-synonymous target sites are located within predicted CD8⁺ T cell epitopes, consistent with cell-mediated immunity (**Table 2; table S4**).

The two SA_{WG} yielded different numbers of target loci, which partially overlapped between studies. Differences can result from several factors, including Pf treatment (or lack thereof) before vaccination, the vaccine tested, and VE outcome. The Burkinabe clinical trial⁹ tested three doses of PfSPZ Vaccine (2.7×10^6 PfSPZ/dose), which consists of radiation attenuated, non-replicating PfSPZ that arrest early in the liver stage. The Malian trial¹⁷ assessed three doses of PfSPZ-CVac (CQ) (2×10^5 PfSPZ/dose), which consists of non-attenuated, fully replicating PfSPZ that are eliminated by chloroquine after liver schizonts rupture and parasites reach the blood.⁵ Therefore, different antigen sets are presented by the two vaccines, and for different periods of time, likely inducing responses to different, and only partly overlapping, sets of targets. Patients were treated to clear blood stage parasitemia pre-vaccination in the Burkinabe study, but not in Mali. It is unclear if an ongoing infection with a genotype different from PfNF54 could have impacted protection outcome in a genotype-specific manner, hence contributing to an increase in noise to signal ratio, with a larger proportion of the differentially distributed genotypes being due to chance in the Mali study. In addition, the two studies differed in protection outcome, with the trial in Burkina Faso (but not the Malian trial) resulting in significant VE. Lack of significant VE likely also increases noise to signal ratio. This is consistent with the lower signal observed in site and gene enrichment analyses in Malian relative to Burkinabe samples, including a smaller proportion of differentially distributed non-

synonymous sites in which the vaccine allele was underrepresented among vaccinee samples relative to controls, and a lack of GO term enrichment among target loci. Therefore, the studies almost certainly differ in power to detect an association between Pf genetic variants and protection. Differences between host populations in the two studies, including in HLA allele frequencies, previous malaria exposure, and previous exposure to different Pf genotypes, could also have led to differences in the target sites detected in these two small trials.

This study has two key limitations: one was the small sample size of each SA_{WG}, which likely contributed to increase the number of false positives;¹⁶ the second was the unavailability of HLA genotype for each participant. The lack of this information resulted in reduced power to detect an association between target sites and individual immunopeptidomes, and precluded paired host HLA-Pf genotype analysis to determine the extent to which HLA genotype contributed to Pf genotype differentiation between vaccinees and controls. Finally, a limitation inherent to any SA_{WG} is the inability to identify sites that encode residues targeted by, and which contribute partially to, the protective immune response but that were invariable within the pathogen data set collected. Without genetic variation, it is impossible to detect pathogen genotype differentiation between study arms, the metric upon which SA_{WG} relies.

Overall, the encouraging results obtained support the use of SA_{WG} as an agnostic approach to identify novel targets of vaccine-induced protection. Analyses of additional PfSPZ-based placebo-controlled VE trials in malaria-endemic areas can be used to further test this approach, potentially strengthening identification and prioritization of target loci. Importantly, it is essential to validate these results for any targets intended for vaccine formulation, by demonstrating that they are both immunogenic and protective. Finally, the evidence presented is consistent with a scenario where multiple parasite proteins contribute to the protective immune response induced by PfSPZ vaccines and in which these protective immune responses are often allele-specific, as is the case, by definition, of the target loci identified here. Therefore, to achieve strain-transcendent, broad protection, PfSPZ vaccines will need to induce optimal protective immune responses to variable epitopes from multiple proteins, including subdominant epitopes, and could potentially be improved by including mixtures of immunologically complementary PfSPZ strains, genetic crosses of parasites from different geographic regions, and/or co-expression, in an individual transgenic Pf vaccine strain, of multiple, immunologically complementary variants of target antigens.

RESOURCE AVAILABILITY

Whole genome shotgun sequencing data are available under BioProject PRJNA1134691 (BioSamples SAMN43081578- SAMN43081613) for Burkinabe samples and BioProject PRJNA1134692 (BioSamples SAMN43082681 - SAMN43082748) for Malian samples. In-house scripts can be found at <https://github.com/jcsilva-lab>.

ACKNOWLEDGEMENTS

The authors thank Maryland Genomics for the preparation of all genomic libraries and generation of whole genome sequencing data. We thank Drs. Gloria Meng-Hsuan Lin and Dominique Soldati-Favre for discussions on the GAC protein presentation. We further thank the volunteers and their families in Mali and Burkina Faso for their participation in clinical trials of PfSPZ Vaccine and PfSPZ-CVac.

FUNDING

This study was supported in part by National Institutes of Health (NIH) awards R01AI141900 (JCS), U01AI112367 (CVP, MBL), and NIH contract number HHSN272201300022I (MBL).

AUTHOR CONTRIBUTIONS

Conceptualization: JCS, AD, STH, SLH

Sample acquisition: BS, SJ, AO, ABT, DC, MAT, SBS, MBL

Methodology: JCS, AD, DGH, RJS

Data Analysis: AD, RJS, DGH, TCS, OOI

Data Management: CG, JBM

Visualization: AD, RJS

Funding acquisition: JCS, STH, MBL, CVP

Project administration: JCS

Supervision: JCS

Writing – original draft: JCS, MBL, AD, RJS

Writing – review & editing: All co-authors

COMPETING INTERESTS

SL Hoffman TL Richie and BKL Sim are salaried employees of Sanaria Inc., the developer and owner of PfSPZ Vaccine and PfSPZ CVac. In addition, SL Hoffman and BKL Sim have a financial interest in Sanaria, Inc. All other authors declare that they have no competing interests.

MATERIALS AND METHODS

Study design

This study aimed to identify parasite proteins putatively targeted by the immune system in response to vaccination with whole *Pf* sporozoites (PfSPZ) using whole genome sieve analysis (SA_{WG}). To this end, for two placebo-controlled malaria vaccine efficacy field trials of PfSPZ-based vaccines, we compared allele frequency distribution, at each genomic site, between *Pf* infections collected from the unprotected vaccinees and from placebo recipients. In each study, we identified genomic sites in which the vaccine allele (allelic state in PfNF54) was significantly under-represented among vaccinee infections, relative to infections from control individuals.

Samples

Samples originated from two clinical trials assessing the efficacy of malaria vaccines against natural infection in West Africa, one in Burkina Faso (NCT02663700) and one in Mali (NCT02996695). Samples analyzed correspond to the timing of the first thick blood smear-positive malaria infection occurring at least two weeks after the last vaccination dose, corresponding to 33 and 40 participants, respectively, in the Burkinabe and Malian trials. Samples consisted of a 2 mL venous blood draw that was leukocyte-depleted, pelleted and stored at -20°C or -80°C. DNA was extracted from each sample, using the Qiagen Blood DNA Midi Kit (Valencia, CA, USA). Total DNA extracted varied between 8.0-1031.2 ng (median: 75.9 ng) for Burkinabe samples and between 0.0-78043.3 ng (median: 10.3 ng) for Mali samples. Nine out of 40 samples from Mali had unusually high DNA amounts, likely due to contamination. All samples with <100 ng of total DNA or those with likely contamination underwent selective whole genome amplification (sWGA) with *Pf*-specific primers to increase the relative amount of parasite DNA. Amplification was conducted in a 0.2 mL 96-well PCR plate. The reaction mixture comprised 1X bovine serum albumin (BSA), 1 mM deoxynucleotide triphosphates (dNTPs), 2.5 μM of each amplification primer, 1X Phi29 reaction buffer, and 30 units of Phi29 DNA polymerase. A total of 17 μL of template DNA was added to the reaction mixture, bringing the final reaction volume to 50 μL. The amplification was carried out in a thermocycler using a step-down protocol: 35 °C for 5 min, 34 °C for 10 min, 33 °C for 15 min, 32 °C for 20 min, 31 °C for 30 min, and 30 °C for 16 h. Enzyme inactivation was achieved by heating at 65 °C, followed by cooling at 4 °C.⁴⁵ Three out of 40 Malian samples had insufficient DNA for library construction (**table S1**).

DNA sequencing

Genomic DNA, extracted from leukocyte-depleted samples or generated through sWGA, was used to construct DNA libraries using the KAPA Library Preparation Kit (Kapa Biosystems, Woburn, MA). DNA was fragmented with the Covaris E220 to ~200-300 bp. Libraries were prepared using a modified version of manufacturer's protocol. The DNA was purified between enzymatic reactions and library insert size selection was performed with AMPure XT beads (Beckman Coulter Genomics, Danvers, MA). When necessary, a PCR amplification step was performed with primers containing an index sequence of six nucleotides in length. Libraries were assessed for concentration and fragment size using the DNA High Sensitivity Assay on the LabChip GX (Perkin Elmer, Waltham, MA). Library concentrations were also assessed by qPCR using the KAPA Library Quantification Kit (Complete, Universal) (Kapa Biosystems, Woburn, MA). The libraries were multiplex on a 150 bp paired-end Illumina HiSeq 4000 run (Illumina, San Diego, CA). For each sample, sequencing data was generated to a minimum of 75% breadth of coverage of the Pf3D7 genome with at least 5X depth of coverage. Samples for which that level of coverage was unfeasible, due to a high proportion of human DNA, were not used for downstream whole-genome sieve analysis.

Sequencing variant identification and classification

Reads were aligned to the 3D7 reference genome obtained from PlasmoDBv24, in EuPathDB,⁴⁶ using Bowtie2 v2.3.4.3.⁴⁷ All WGS data files with data from the same original specimen were combined, resulting in a merged BAM file that contained all reads regardless of library type (**table S1**). The resulting BAM files were processed according to GATK's Best Practices documentation.⁴⁸ Joint SNP calling was done using Haplotype Caller v1.7.0. In addition to 31 samples from Burkina Faso PfSPZ and 31 samples from Mali PfSPZ-CVac trial, short read (150 bp) whole genome shotgun sequencing (WGS) data for 1,289 *P. falciparum* isolates collected from 20 malaria-endemic countries in Africa, Southeast Asia, Oceania and S. America were downloaded from NCBI's Short Read Archive and were used in joint SNP calling (**table S7**). Diploid calls were allowed to account for heterozygous positions in polyclonal samples. Additional hard filtering was done to remove potential false positives using filter $DP < 5 \parallel QUAL < 50 \parallel FS > 14.5 \parallel MQ < 20$. In downstream analysis, only single nucleotide polymorphisms (SNPs) were used, and additional filtering was done on the SNP data. Sites with genotype information missing in >30% of samples were removed, as were sites with variants present in fewer than three samples, to avoid possible artifacts. Genomic sites were classified as

intergenic, intronic, synonymous and non-synonymous using snpEFF v4.3t.⁴⁹ For genotyping of loci of interest with potential sequence length variation, an in-house pipeline for *in silico* capture and assembly (ISCA v1.2.3; <https://github.com/jgs-jcsilva-lab/ISCA>) of short sequence reads that map to each of those loci was used.

Multiplicity of infection.

Multiplicity of infection was assessed with the statistic F_{WS} , estimated with the R package moimix.⁵⁰ Difference in F_{WS} difference between study arms was tested with a Wilcoxon Rank Sum test. The number of clones per infection was estimated with dEplod-IBD v0.5,⁵¹ with default parameters, and significance tested with a chi-square test.

Whole genome sieve analysis (SA_{WG}).

A SA_{WG} was conducted separately using WGS data generated from isolates collected in each clinical trial, and in each case was based on all polymorphic sites that passed filters, as described above. A SA_{WG} consists of two steps: *i*) identification of parasite genomic sites significantly differentiated between study arms (vaccinees and controls), and (*ii*) identification of sites from (*i*) where the vaccine allele is depleted among vaccinee infections. For each genomic site, difference in allele frequency distribution between study arms (step *i*) was quantified using Wright's Fixation Index, F_{ST} (Weir and Cockerham's implementation in vcftools v0.1015).⁵² For each analysis, F_{ST} value significance was determined by randomization of samples across study arms, with 5000 replicates. Sites found to be significantly differentiated between vaccinee and control infections, and in which the vaccine (PfNF54) allele was underrepresented in vaccinee infections, were termed "*target sites*". Deviation of expected frequency of target sites per study arm was tested with a chi-square test.

Loci were termed "*target loci*" if they contained non-synonymous target sites. The 99% confidence interval and p-value for the number of target loci found, by chance, in both SA_{WG} (given m and n targets genes, one set in each study), given the total number of genes g with variable non-synonymous sites across studies, were both determined by Fisher's exact conditional test for 2x2 independence (fixed row and column sums, m and n), which is a conservative version of a multinomial (unconditional) test. The multinomial test is more appropriate (fixed g but variable number of target loci, m and n) but it is computationally

expensive. The expected number of target loci that overlap between studies, E , is given by $E = m*n|g$.

Feature Identification and enrichment analyses

CD8+ T cell epitopes were predicted across target loci using NetMHCpan v4.1,⁵³ using default parameters. Only strong binding epitopes were considered in downstream analysis. The HLA alleles used for epitope prediction were the 22 most common among those identified in Mali and Burkina Faso.⁵⁴ Enrichment analyses of gene ontology (GO) terms among different gene sets were done using ShinyGO v0.741.⁵⁵ In target loci, enrichment of target sites in strong binding epitope-encoding regions was tested using a chi-square test.

References

1. Vogel, G. (2023). Second malaria vaccine gets WHO green light. *Science* 382, 16-17. 10.1126/science.adl1805.
2. WHO (2022). Malaria vaccines: preferred product characteristics and clinical development considerations. .
3. Neafsey, D.E., Juraska, M., Bedford, T., Benkeser, D., Valim, C., Griggs, A., Lievens, M., Abdulla, S., Adjei, S., Agbenyega, T., et al. (2015). Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. *N Engl J Med* 373, 2025-2037. 10.1056/NEJMoa1505819.
4. World Health Organization = Organisation mondiale de la, S. (2022). WHO Guidelines for malaria, 25 November 2022. World Health Organization.
5. Richie, T.L., Church, L.W.P., Murshedkar, T., Billingsley, P.F., James, E.R., Chen, M.C., Abebe, Y., Natasha, K., Chakravarty, S., Dolberg, D., et al. (2023). Sporozoite immunization: innovative translational science to support the fight against malaria. *Expert Rev Vaccines* 22, 964-1007. 10.1080/14760584.2023.2245890.
6. Seder, R.A., Chang, L.J., Enama, M.E., Zephir, K.L., Sarwar, U.N., Gordon, I.J., Holman, L.A., James, E.R., Billingsley, P.F., Gunasekera, A., et al. (2013). Protection against malaria by intravenous immunization with a nonreplicating sporozoite vaccine. *Science* 341, 1359-1365. 10.1126/science.1241800.
7. Mordmüller, B., Surat, G., Lagler, H., Chakravarty, S., Ishizuka, A.S., Lalremruata, A., Gmeiner, M., Campo, J.J., Esen, M., Ruben, A.J., et al. (2017). Sterile protection against human malaria by chemoattenuated PfSPZ vaccine. *Nature* 542, 445-449. 10.1038/nature21060.
8. Mwakingwe-Omari, A., Healy, S.A., Lane, J., Cook, D.M., Kalhori, S., Wyatt, C., Kolluri, A., Marte-Salcedo, O., Imeru, A., Nason, M., et al. (2021). Two chemoattenuated PfSPZ malaria vaccines induce sterile hepatic immunity. *Nature* 595, 289-294. 10.1038/s41586-021-03684-z.
9. Sirima, S.B., Ouedraogo, A., Tiono, A.B., Kabore, J.M., Bougouma, E.C., Ouattara, M.S., Kargougou, D., Diarra, A., Henry, N., Ouedraogo, I.N., et al. (2022). A randomized controlled trial showing safety and efficacy of a whole sporozoite vaccine against endemic malaria. *Sci Transl Med* 14, eabj3776. 10.1126/scitranslmed.abj3776.
10. Sissoko, M.S., Healy, S.A., Katile, A., Omaswa, F., Zaidi, I., Gabriel, E.E., Kamate, B., Samake, Y., Guindo, M.A., Dolo, A., et al. (2017). Safety and efficacy of PfSPZ Vaccine against *Plasmodium falciparum* via direct venous inoculation in healthy malaria-exposed adults in Mali: a randomised, double-blind phase 1 trial. *Lancet Infect Dis* 17, 498-509. 10.1016/S1473-3099(17)30104-4.
11. Sissoko, M.S., Healy, S.A., Katile, A., Zaidi, I., Hu, Z., Kamate, B., Samake, Y., Sissoko, K., Mwakingwe-Omari, A., Lane, J., et al. (2022). Safety and efficacy of a three-dose regimen of *Plasmodium falciparum* sporozoite vaccine in adults during an intense malaria transmission season in Mali: a randomised, controlled phase 1 trial. *Lancet Infect Dis* 22, 377-389. 10.1016/S1473-3099(21)00332-7.
12. Dooley, N.L., Chabikwa, T.G., Pava, Z., Loughland, J.R., Hamelink, J., Berry, K., Andrew, D., Soon, M.S.F., SheelaNair, A., Piera, K.A., et al. (2023). Single cell transcriptomics shows that malaria promotes unique regulatory responses across multiple immune cell subsets. *Nat Commun* 14, 7387. 10.1038/s41467-023-43181-7.
13. Portugal, S., Tipton, C.M., Sohn, H., Kone, Y., Wang, J., Li, S., Skinner, J., Virtaneva, K., Sturdevant, D.E., Porcella, S.F., et al. (2015). Malaria-associated atypical memory B cells exhibit markedly reduced B cell receptor signaling and effector function. *Elife* 4, 10.7554/eLife.07218.
14. Mitchell, R.A., Ubbilos, I., Requena, P., Campo, J.J., Ome-Kaius, M., Hanieh, S., Umbers, A., Samol, P., Barrios, D., Jimenez, A., et al. (2024). Chronic malaria exposure is associated with inhibitory markers on T cells that correlate with atypical memory and marginal zone-like B cells. *Clin Exp Immunol* 216, 172-191. 10.1093/cei/uxae015.
15. Epstein, J.E., Paolino, K.M., Richie, T.L., Sedegah, M., Singer, A., Ruben, A.J., Chakravarty, S., Stafford, A., Ruck, R.C., Eappen, A.G., et al. (2017). Protection against *Plasmodium falciparum* malaria by PfSPZ Vaccine. *JCI Insight* 2, e89154. 10.1172/jci.insight.89154.
16. Scalsky, R., Dwivedi, A., Stabler, T.C., Mbambo, G., Ouattara, A., Lyke, K.E., Takala-Harrison, S., and Silva, J.C. (2025). Whole-genome sieve analysis: Identification of protective malaria antigens by leveraging allele-specific vaccine efficacy. *Vaccine* 50, 126783. 10.1016/j.vaccine.2025.126783.
17. Coulibaly, D., Kone, A.K., Traore, K., Niangaly, A., Kouriba, B., Arama, C., Zeguime, A., Dolo, A., Lyke, K.E., Plowe, C.V., et al. (2022). PfSPZ-CVac malaria vaccine demonstrates safety among malaria-experienced adults: A randomized, controlled phase 1 trial. *EclinicalMedicine* 52, 101579. 10.1016/j.eclinm.2022.101579.

18. Walliker, D., Quakyi, I.A., Wellems, T.E., McCutchan, T.F., Szarfman, A., London, W.T., Corcoran, L.M., Burkot, T.R., and Carter, R. (1987). Genetic analysis of the human malaria parasite *Plasmodium falciparum*. *Science* 236, 1661-1666. 10.1126/science.3299700.
19. Moser, K.A., Drabek, E.F., Dwivedi, A., Stucke, E.M., Crabtree, J., Dara, A., Shah, Z., Adams, M., Li, T., Rodrigues, P.T., et al. (2020). Strains used in whole organism *Plasmodium falciparum* vaccine trials differ in genome structure, sequence, and immunogenic potential. *Genome Med* 12, 6. 10.1186/s13073-019-0708-9.
20. Edlefsen, P.T., Rolland, M., Hertz, T., Tovanabutra, S., Gartland, A.J., deCamp, A.C., Magaret, C.A., Ahmed, H., Gottardo, R., Juraska, M., et al. (2015). Comprehensive sieve analysis of breakthrough HIV-1 sequences in the RV144 vaccine efficacy trial. *PLoS Comput Biol* 11, e1003973. 10.1371/journal.pcbi.1003973.
21. Lindner, S.E., Swearingen, K.E., Harupa, A., Vaughan, A.M., Sinnis, P., Moritz, R.L., and Kappe, S.H. (2013). Total and putative surface proteomics of malaria parasite salivary gland sporozoites. *Mol Cell Proteomics* 12, 1127-1143. 10.1074/mcp.M112.024505.
22. Zanghi, G., Patel, H., Camargo, N., Smith, J.L., Bae, Y., Flannery, E.L., Chuenchob, V., Fishbaugher, M.E., Mikolajczak, S.A., Roobsoong, W., et al. (2023). Global gene expression of human malaria parasite liver stages throughout intrahepatocytic development. *bioRxiv*. 10.1101/2023.01.05.522945.
23. Real, E., Howick, V.M., Dahalan, F.A., Witmer, K., Cudini, J., Andradi-Brown, C., Blight, J., Davidson, M.S., Dogga, S.K., Reid, A.J., et al. (2021). A single-cell atlas of *Plasmodium falciparum* transmission through the mosquito. *Nat Commun* 12, 3196. 10.1038/s41467-021-23434-z.
24. Kurup, S.P., Butler, N.S., and Harty, J.T. (2019). T cell-mediated immunity to malaria. *Nat Rev Immunol* 19, 457-471. 10.1038/s41577-019-0158-z.
25. Epstein, J.E., Tewari, K., Lyke, K.E., Sim, B.K., Billingsley, P.F., Laurens, M.B., Gunasekera, A., Chakravarty, S., James, E.R., Sedegah, M., et al. (2011). Live attenuated malaria vaccine designed to protect through hepatic CD8(+) T cell immunity. *Science* 334, 475-480. 10.1126/science.1211548.
26. Nussenzweig, R.S., Vanderberg, J., Most, H., and Orton, C. (1967). Protective immunity produced by the injection of x-irradiated sporozoites of *plasmodium berghei*. *Nature* 216, 160-162. 10.1038/216160a0.
27. Bettencourt, P. (2020). Current Challenges in the Identification of Pre-Erythrocytic Malaria Vaccine Candidate Antigens. *Front Immunol* 11, 190. 10.3389/fimmu.2020.00190.
28. Heide, J., Vaughan, K.C., Sette, A., Jacobs, T., and Schulze Zur Wiesch, J. (2019). Comprehensive Review of Human *Plasmodium falciparum*-Specific CD8+ T Cell Epitopes. *Front Immunol* 10, 397. 10.3389/fimmu.2019.00397.
29. Pihlajamaa, T., Kajander, T., Knuuti, J., Horkka, K., Sharma, A., and Permi, P. (2013). Structure of *Plasmodium falciparum* TRAP (thrombospondin-related anonymous protein) A domain highlights distinct features in apicomplexan von Willebrand factor A homologues. *Biochem J* 450, 469-476. 10.1042/BJ20121058.
30. Paoletta, M.S., and Wilkowsky, S.E. (2022). Thrombospondin Related Anonymous Protein Superfamily in Vector-Borne Apicomplexans: The Parasite's Toolkit for Cell Invasion. *Front Cell Infect Microbiol* 12, 831592. 10.3389/fcimb.2022.831592.
31. Cheng, J., Novati, G., Pan, J., Bycroft, C., Zemgulyte, A., Applebaum, T., Pritzel, A., Wong, L.H., Zielinski, M., Sargeant, T., et al. (2023). Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* 381, eadg7492. 10.1126/science.adg7492.
32. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25, 1605-1612. 10.1002/jcc.20084.
33. Gilbert, P., Self, S., Rao, M., Naficy, A., and Clemens, J. (2001). Sieve analysis: methods for assessing from vaccine trial data how vaccine efficacy varies with genotypic and phenotypic pathogen variation. *J Clin Epidemiol* 54, 68-85. 10.1016/s0895-4356(00)00258-4.
34. Ouattara, A., Takala-Harrison, S., Thera, M.A., Coulibaly, D., Niangaly, A., Saye, R., Tolo, Y., Dutta, S., Heppner, D.G., Soisson, L., et al. (2013). Molecular basis of allele-specific efficacy of a blood-stage malaria vaccine: vaccine development implications. *J Infect Dis* 207, 511-519. 10.1093/infdis/jis709.
35. Ouattara, A., Niangaly, A., Adams, M., Coulibaly, D., Kone, A.K., Traore, K., Laurens, M.B., Tolo, Y., Kouriba, B., Diallo, D.A., et al. (2020). Epitope-based sieve analysis of *Plasmodium falciparum* sequences from a FMP2.1/AS02A vaccine trial is consistent with differential vaccine efficacy against immunologically relevant AMA1 variants. *Vaccine* 38, 5700-5706. 10.1016/j.vaccine.2020.06.035.

36. Thera, M.A., Doumbo, O.K., Coulibaly, D., Laurens, M.B., Ouattara, A., Kone, A.K., Guindo, A.B., Traore, K., Traore, I., Kouriba, B., et al. (2011). A field trial to assess a blood-stage malaria vaccine. *N Engl J Med* 365, 1004-1013. 10.1056/NEJMoa1008115.
37. Shah, Z., Naung, M.T., Moser, K.A., Adams, M., Buchwald, A.G., Dwivedi, A., Ouattara, A., Seydel, K.B., Mathanga, D.P., Barry, A.E., et al. (2021). Whole-genome analysis of Malawian *Plasmodium falciparum* isolates identifies possible targets of allele-specific immunity to clinical malaria. *PLoS Genet* 17, e1009576. 10.1371/journal.pgen.1009576.
38. Ogowang, C., Kimani, D., Edwards, N.J., Roberts, R., Mwacharo, J., Bowyer, G., Bliss, C., Hodgson, S.H., Njuguna, P., Viebig, N.K., et al. (2015). Prime-boost vaccination with chimpanzee adenovirus and modified vaccinia Ankara encoding TRAP provides partial protection against *Plasmodium falciparum* infection in Kenyan adults. *Sci Transl Med* 7, 286re285. 10.1126/scitranslmed.aaa2373.
39. Aghazadeh, B., Zhu, K., Kubiseski, T.J., Liu, G.A., Pawson, T., Zheng, Y., and Rosen, M.K. (1998). Structure and mutagenesis of the Dbl homology domain. *Nat Struct Biol* 5, 1098-1107. 10.1038/4209.
40. Schepis, A., Kumar, S., and Kappe, S.H.I. (2023). Malaria parasites harness Rho GTPase signaling and host cell membrane ruffling for productive invasion of hepatocytes. *Cell Rep* 42, 111927. 10.1016/j.celrep.2022.111927.
41. Jacot, D., Tosetti, N., Pires, I., Stock, J., Graindorge, A., Hung, Y.F., Han, H., Tewari, R., Kursula, I., and Soldati-Favre, D. (2016). An Apicomplexan Actin-Binding Protein Serves as a Connector and Lipid Sensor to Coordinate Motility and Invasion. *Cell Host Microbe* 20, 731-743. 10.1016/j.chom.2016.10.020.
42. Shrestha, S., Li, X., Ning, G., Miao, J., and Cui, L. (2016). The RNA-binding protein Puf1 functions in the maintenance of gametocytes in *Plasmodium falciparum*. *J Cell Sci* 129, 3144-3152. 10.1242/jcs.186908.
43. Tomavo, S., Slomianny, C., Meissner, M., and Carruthers, V.B. (2013). Protein trafficking through the endosomal system prepares intracellular parasites for a home invasion. *PLoS Pathog* 9, e1003629. 10.1371/journal.ppat.1003629.
44. Nishimura, N., Plutner, H., Hahn, K., and Balch, W.E. (2002). The delta subunit of AP-3 is required for efficient transport of VSV-G from the trans-Golgi network to the cell surface. *Proc Natl Acad Sci U S A* 99, 6755-6760. 10.1073/pnas.092150699.
45. Shah, Z., Adams, M., Moser, K.A., Shrestha, B., Stucke, E.M., Laufer, M.K., Serre, D., Silva, J.C., and Takala-Harrison, S. (2020). Optimization of parasite DNA enrichment approaches to generate whole genome sequencing data for *Plasmodium falciparum* from low parasitaemia samples. *Malar J* 19, 135. 10.1186/s12936-020-03195-8.
46. Warrenfeltz, S., Basenko, E.Y., Crouch, K., Harb, O.S., Kissinger, J.C., Roos, D.S., Shanmugasundram, A., and Silva-Franco, F. (2018). EuPathDB: The Eukaryotic Pathogen Genomics Database Resource. *Methods Mol Biol* 1757, 69-113. 10.1007/978-1-4939-7737-6_5.
47. Langdon, W.B. (2015). Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. *BioData Min* 8, 1. 10.1186/s13040-014-0034-0.
48. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20, 1297-1303. 10.1101/gr.107524.110.
49. Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6, 80-92. 10.4161/fly.19695.
50. Lee, S.A., Harrison, A.G., Tessier, N., Tavul, L., Miotto, O., Siba, P.M., Mueller, I., Barry, A.E., and Bahlo, M. (2016). Assessing clonality in malaria parasites using massively parallel sequencing. <http://bahlolab.github.io/moimix/>.
51. Zhu, S.J., Hendry, J.A., Almagro-Garcia, J., Pearson, R.D., Amato, R., Miles, A., Weiss, D.J., Lucas, T.C., Nguyen, M., Gething, P.W., et al. (2019). The origins and relatedness structure of mixed infections vary with local prevalence of *P. falciparum* malaria. *Elife* 8, 10.7554/eLife.40845.
52. Weir, B.S., and Cockerham, C.C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution* 38, 1358-1370. 10.1111/j.1558-5646.1984.tb05657.x.
53. Reynisson, B., Alvarez, B., Paul, S., Peters, B., and Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res* 48, W449-W454. 10.1093/nar/gkaa379.

54. Gonzalez-Galarza, F.F., McCabe, A., Santos, E., Jones, J., Takeshita, L., Ortega-Rivera, N.D., Cid-Pavon, G.M.D., Ramsbottom, K., Ghattaoraya, G., Alfirevic, A., et al. (2020). Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res* 48, D783-D788. 10.1093/nar/gkz1029.
55. Ge, S.X., Jung, D., and Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* 36, 2628-2629. 10.1093/bioinformatics/btz931.