

1 **Proteome-wide antigenic profiling in Ugandan cohorts identifies associations between age,**
2 **exposure intensity, and responses to repeat-containing antigens in *Plasmodium falciparum***

3
4 Madhura Raghavan¹, Katrina L. Kalantar², Elias Duarte³, Noam Teyssier¹, Saki Takahashi¹, Andrew F.
5 Kung¹, Jayant V Rajan¹, John Rek⁴, Kevin K.A. Tetteh⁵, Chris Drakeley⁵, Isaac Ssewanyana^{4,5}, Isabel
6 Rodriguez-Barraquer^{1,6}, Bryan Greenhouse^{1,6*} and Joseph L. DeRisi^{1,6*}

7
8 Affiliations:

9 ¹ University of California San Francisco, San Francisco, CA, USA

10 ² Chan Zuckerberg Initiative, Redwood City, CA, USA

11 ³ University of California Berkeley, Berkeley, CA, USA

12 ⁴ Infectious Disease Research Collaboration, Kampala, Uganda

13 ⁵ London School of Hygiene & Tropical Medicine, London, United Kingdom

14 ⁶ Chan Zuckerberg Biohub, San Francisco, CA, USA

15
16 *Corresponding authors

17 (Joseph L. DeRisi - joe@derisilab.ucsf.edu, Bryan Greenhouse - bryan.greenhouse@ucsf.edu)

18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

47 **ABSTRACT**

48

49 Protection against *Plasmodium falciparum*, which is primarily antibody-mediated, requires recurrent
50 exposure to develop. The study of both naturally acquired limited immunity and vaccine induced protection
51 against malaria remains critical for ongoing eradication efforts. Towards this goal, we deployed a
52 customized *P. falciparum* PhIP-seq T7 phage display library containing 238,068 tiled 62-amino acid
53 peptides, covering all known coding regions, including antigenic variants, to systematically profile antibody
54 targets in 198 Ugandan children and adults from high and moderate transmission settings. Repeat elements
55 – short amino acid sequences repeated within a protein – were significantly enriched in antibody targets.
56 While breadth of responses to repeat-containing peptides was twofold higher in children living in the high
57 versus moderate exposure setting, no such differences were observed for peptides without repeats,
58 suggesting that antibody responses to repeat-containing regions may be more exposure dependent and/or
59 less durable in children than responses to regions without repeats. Additionally, short motifs associated
60 with seroreactivity were extensively shared among hundreds of antigens, potentially representing cross-
61 reactive epitopes. PfEMP1 shared motifs with the greatest number of other antigens, partly driven by the
62 diversity of PfEMP1 sequences. These data suggest that the large number of repeat elements and potential
63 cross-reactive epitopes found within antigenic regions of *P. falciparum* could contribute to the inefficient
64 nature of malaria immunity.

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91 INTRODUCTION

92

93 Malaria, a disease caused by the single-celled eukaryotic parasite *Plasmodium*, caused an estimated 241
94 million cases and 627,000 deaths in 2020, mostly by the species *Plasmodium falciparum* (*P. falciparum*)
95 (WHO Report 2021). Various strategies have been adopted for elimination of malaria, focusing on vector
96 control, chemoprevention and vaccines. In 2021, the World Health Organization (WHO) made its first
97 recommendation for widespread use of a malaria vaccine, RTS,S. While this is an encouraging step, there
98 is nevertheless need for improvement as the efficacy of RTS,S is only 30-40% and protection wanes in a
99 few months despite a four-dose regimen (Clinical Trials Partnership, 2015; Olotu et al., 2016). To design
100 more effective vaccines, a deeper understanding of the nature of acquired immunity to malaria is critical.
101 Natural protection against malaria requires multiple exposures and wanes upon cessation of exposure
102 (Doolan et al., 2009). This naturally acquired immunity develops gradually with age and increasing
103 cumulative exposure to *P. falciparum* in endemic settings, where adults may obtain substantial protection
104 from disease, and children under 5 face the highest risk of death(Doolan et al., 2009)(WHO Report 2020).
105 While a comprehensive understanding of the factors influencing the slow development of immunity is still
106 lacking, certain properties of parasite antigens have been proposed to contribute (Portugal et al., 2013).
107 These include, amongst other properties, antigenic variation, antigens containing repeat elements and
108 cross-reactive epitopes (Anders, 1986; Reeder & Brown, 1996; Schofield, 1991). While antigenic variation
109 has been extensively studied in malaria, a systematic investigation of repeat-containing antigens and cross-
110 reactive epitopes has been lacking.

111
112 Repeat elements are those where identical or similar motifs are repeated in tandem or with spaces within a
113 protein. Repeat elements are widely prevalent in the proteome of *P. falciparum* and have been described to
114 be highly immunogenic in a few antigens (Davies et al., 2017), such as the short, linear “NANP” repeats
115 from circumsporozoite protein (CSP) present in the RTS,S and R21 vaccines (Cockburn & Seder, 2018).
116 Due to increased valency, epitopes in repeat elements can behave differently in comparison to the
117 presentation of the same epitope as a single copy and have the potential to alter the nature of the resulting
118 response. For instance, increased valency may lead to increased plasmablast formation by increasing the
119 strength of the antigen-B-cell receptor (BCR) interaction, potentially altering the T-dependent response and
120 inducing a T-independent response (Feldmann & Basten, 1971; Kato et al., 2020; O’Connor et al., 2006;
121 Ochiai et al., 2013; Paus et al., 2006; Schofield, 1991; Schwickert et al., 2011). Although a few repeat
122 antigens in *P. falciparum* have been well characterized, there has not been a comprehensive investigation
123 of repeat elements with respect to their seroreactivity and their associations with humoral development.

124
125 The presence of biochemically similar epitopes can lead to cross-reactivity with antibodies and B-cell
126 receptors (BCR). While non-identical repeat elements may represent such potential cross-reactive epitopes
127 within a protein, similar epitopes may also be present across different proteins. How the quality of humoral
128 response may be impacted by the presence of cross-reactive epitopes remains largely unexplored, although
129 a study with viral variant antigens points to a frustrated affinity maturation process due to conflicting
130 selection forces from variant epitopes (Wang et al., 2015). A handful of cross-reactive epitopes have been
131 reported in *P. falciparum* (Wählén et al., 1992) and have been proposed to negatively impact the affinity
132 maturation process, although direct evidence is lacking (Anders, 1986). To obtain a deeper understanding
133 of how cross-reactive epitopes influence B cell immunity to malaria, a comprehensive atlas of cross-reactive
134 epitopes across the *P. falciparum* proteome is first needed.

135
136 A systematic proteome-wide investigation of the humoral response to *P. falciparum* would provide
137 important insights to our understanding of malaria immunity, including features such as repeat elements
138 and cross-reactive epitopes. Specific technical challenges have impeded progress in this area. Although
139 high-throughput approaches like protein arrays and alpha screens have reached a high coverage of the *P.*
140 *falciparum* proteome (Camponovo et al., 2020; Morita et al., 2017), they do not allow for high-resolution,
141 characterization of regions within antigenic proteins. In contrast, peptide arrays offer high resolution
142 antigenic profiling but are inherently limited to the numbers of targets that can be produced and printed on
143 an array, usually in the order of tens of proteins(Hou et al., 2020; Jaenisch et al., 2019).

144
145 Here, we customized a programmable phage display system (PhIP-seq) (Larman et al., 2011), previously
146 used for antigenic profiling in many diseases, including autoimmune disorders and viral infections (Mandel-
147 Brehm et al., 2019; Rajan et al., 2021; Vazquez et al., 2020; Zamecnik et al., 2020), for interrogation of the
148 humoral response to *P. falciparum* infection. We designed a custom library (“Falciparome”) that features
149 over 238,000 individual 62 amino acid peptides encoded in T7 Phage, tiled every 25 amino acids across all
150 annotated *P. falciparum* open reading frames from 3D7/IT genomes with additional variant antigenic
151 sequences. Importantly, PhIP-seq leverages advances in next generation sequencing to effectively convert
152 serological assays into digital sequence counts. Furthermore, programmable phage display allows iterative
153 enrichment, driving a high signal to background ratio with high specificity and sensitivity (O’Donovan et
154 al., 2020).

155
156 We performed PhIP-seq with the Falciparome phage library to characterize the targets of the naturally
157 acquired antibody response to *P. falciparum* in high resolution, leveraging well defined cohorts composed
158 of 198 Ugandan children and adults from two different transmission settings and compared these to a large
159 set of US anonymous blood donors. The resulting high-resolution atlas of seroreactive peptides suggest that
160 antibody responses to repeat-containing regions are more exposure-dependent and/or less durable in
161 children, compared to antibody responses to regions without repeats. Further, an extensive presence of
162 potential cross-reactive motifs was identified among antigenic peptides from many proteins highly targeted
163 by the immune system. These results have important implications for understanding the nature of humoral
164 response in malaria and the future vaccine designs.

165 166 **RESULTS**

167
168 PhIP-seq was performed on plasma samples selected from two Ugandan cohorts with household level data
169 on entomologic exposure as well as detailed individual characteristics (Table 1). For this study, we selected
170 a single sample from each of 200 age-stratified individuals (children aged 2-11 years and adults) from two
171 different sites in Uganda: Tororo, a region which had very high malaria transmission at the time of sampling
172 (annual entomological inoculation rate (EIR) - 49 infective bites per person) and Kanungu, a region of
173 moderately high transmission (EIR - 5 infective bites per person)(Kamya et al., 2015). While the majority
174 of individuals from Tororo were positive for infection at the time of sampling, those from Kanungu were
175 sampled at a median of 100 days after their previous infection. We have previously shown that children
176 acquire clinical immunity to malaria more rapidly in Tororo than Kanungu, consistent with higher rates of
177 exposure(Rodriguez-Barraquer et al., 2018), and that adults at both sites have substantial immunity (Rek et
178 al., 2016).

179

180 **Falciparome Library Design**

181

182 We constructed a T7 phage display library programmed to display the entire proteome of *P. falciparum* in
183 62-amino acid peptides with 25-amino acid step size, resulting in 37-amino acid tiling, referred to as the
184 Falciparome. (Methods, Figure 1). The complete design files are available at Dryad doi:10.7272/Q69S1P9G
185 and protocol at dx.doi.org/10.17504/protocols.io.j8nlkkr515r/v1. Overall, the library includes 238,068
186 peptides from 8,980 protein sequences, including all known protein sequences from 2 reference strains
187 (3D7 and IT) and extensive diversity of variant sequences from key antigens including PfEMP1s, RIFINs
188 and STEVORs (Table 2, Figure S1a, Methods). Greater than 99.5% of the programmed peptides were
189 represented in the final packaged phage library with relatively uniform distribution of abundance, with 90%
190 of the peptides within a 16-fold difference (Figure S2a).

191

192 **PhIP-seq using the Falciparome library robustly identifies peptides that differentiate individuals in** 193 **malaria endemic areas from US controls**

194

195 PhIP-seq was performed with Falciparome on less than 1ul plasma from the 200 Ugandan cohort samples,
196 and 86 samples from New York Blood Center (US controls) were run for non-specific background
197 correction, assuming most were unlikely to have been exposed. Two rounds of enrichment were performed.
198 The scalability of the assay allowed for high-throughput processing of all 286 samples in replicates. High
199 correlation observed between technical replicates (Pearson r: median (IQR) = 0.96 (0.92-0.98)) indicated
200 that the technique was highly reproducible (Figure S2b). Prior to any filtering, a clean separation of
201 Ugandan and US controls was observed (Figure S2b). Furthermore, expected target peptides were enriched
202 in a sample-specific manner - PhIP-seq with a polyclonal control antibody α -GFAP highly enriched for
203 GFAP peptides and seroreactivity against a common virus, Epstein-Barr virus (EBV), was higher across all
204 human samples than in the control α -GFAP experiment (Figure S2c). Two Ugandan samples were dropped
205 due to low quality data, resulting in 198 Ugandan samples for further analysis.

206

207 A stringent analysis pipeline was implemented to identify malaria-specific enriched peptides while
208 minimizing the potential for false positives. An increase in base read counts (enrichment) compared to US
209 controls (Z-score ≥ 3 in both replicates in a given sample) was implemented, plus a requirement that the
210 enrichment be present in at least 5 Ugandan samples (Methods, Figure 1, Figure S2d) to identify malaria-
211 specific enriched peptides ('seroreactive peptides'). Using this conservative approach, a total of 9927
212 peptides were identified as seroreactive across all samples, representing the identified targets of antibodies
213 in this study (Supplementary table 1).

214

215 **Overview of the malaria-specific seroreactive peptides identified with PhIP-seq**

216

217 The 9,927 seroreactive peptides identified by the pipeline were derived from 1,648 parasite proteins
218 ('seroreactive proteins') and antigenic variants, many of which showed broad seroreactivity across pediatric
219 and adult Ugandan samples, whereas these same peptides showed no or rare seroreactivity in US controls
220 (Figure 2A). The number of peptides enriched ("breadth") in children from moderate transmission settings
221 was less than half of that in children in high transmission settings or adults in either setting (Figure 2B), an
222 observation that we examined in greater detail below.

223
224 The 1648 seroreactive proteins identified here have reported expression across the lifecycle stages occurring
225 in the human - sporozoite, asexual and sexual blood stages (Figure 2c). Although liver stage proteomic *P.*
226 *falciparum* datasets were not available for comparison, several known liver stage antigens in the dataset
227 (LSA1, LSA3, etc.) were detected. Notably, none of the proteins expressed in the mosquito oocyst stage
228 were identified as seroreactive. Among the 40 seroreactive proteins with the highest seropositivity (percent
229 of people enriching for at least one seroreactive peptide in that protein), protective antibodies have been
230 reported for 20 of them (Supplementary table 2). Moreover, as expected, and consistent with previous
231 studies, the top seroreactive proteins were enriched for those at the host-parasite interface (GO analysis –
232 Fig S2e).

233
234 The proteins identified here overlapped substantially with antigenic proteins identified in previous protein
235 array screens (28%, 49% and 44% of those reported in (Camponovo et al., 2020; Crompton et al., 2010;
236 Helb et al., 2015) respectively). However, this whole-proteome approach also identified 952 proteins not
237 identified in the above studies. Antigens identified in previous studies may have not been enriched here
238 because of a known limitation of PhIP-seq - it detects predominantly linear epitopes, as opposed to
239 conformational epitopes, which would account for loss of sensitivity with respect to particular proteins. In
240 addition, prior studies were performed in different populations and may include false positives, e.g. by not
241 accounting for cross-reactivity from non-malaria specific antibodies with unexposed human samples.

242
243 **Expected and new relationships between age, exposure, and breadth of seroreactive regions captured**
244 **at high resolution by Falciparome**

245
246 Since our cohort was stratified by age and exposure, we next set out to investigate how the overall breadth
247 of seroreactive regions varied with age and exposure. Breadth was evaluated in two ways – i) the total
248 number of seroreactive peptides per person ii) the number of non-redundant seroreactive peptide groups in
249 each person. The latter was calculated to minimize redundant counting of potential shared linear epitopes
250 between seroreactive peptides due to the tiled nature of the library as well as common sequences across
251 peptides (Methods). Breadth increased with age in both settings, occurring more rapidly in the higher
252 transmission setting such that children reached a similar breadth as adults by age 11 (Figure 2b, Figure S2f).
253 As a result, children in the higher transmission setting had a significantly higher breadth than children in
254 the moderate transmission setting. In contrast, adults in both settings had comparable breadth. Overall, these
255 results are consistent with expected expansion of the repertoire of antibody targets with recurrent exposure
256 to *P. falciparum* (Crompton et al., 2010; Helb et al., 2015).

257
258 Variant surface antigen (VSA) families are highly diverse, multi-member gene families in *P. falciparum*
259 that are expressed on the surface of host erythrocytes and facilitate important functions of the parasite
260 (Niang et al., 2014; Reeder & Brown, 1996; Saito et al., 2017; Tan et al., 2015; Xie et al., 2021). Expression
261 of these genes is typically limited to one or few members of each family per parasite, presumably to evade
262 the host immune system. Multiple variants from three VSA families were represented in the library -
263 PfEMP1s (431 members from seven strains), RIFINs (157 3D7 + 118 IT) and STEVORs (32 3D7 + 32 IT
264 variants), and the breadth of seroreactive variants was investigated across age and exposure based on the
265 number of variant proteins in each family to which the VSA seroreactive peptides belonged in each person
266 (Methods) (Figure 2d, Figure S2g). In the moderate transmission setting, adults had a significantly higher

267 breadth of PfEMP1 variants recognized than children, suggesting an age and/or cumulative exposure
268 dependent increase in PfEMP1 breadth in this setting, as previously observed in (Cham et al., 2009). On
269 the other hand, both children and adults in this setting poorly recognized RIFINs and STEVORs. In contrast,
270 in the high transmission setting, children had a significantly higher breadth of variants recognized than
271 adults for all three VSAs. Children of 2–6 years had the broadest responses to RIFINs (including in the
272 variable V2 region) and STEVORS and children of 4–11 years to PfEMP1s (including in the variable DBL
273 domains), suggesting a decline in responses to variants as children develop into adults in this setting. This
274 is consistent with observations from a previous study investigating antibody responses to PfEMP1 DBL α
275 domains in Papua New Guinea (Barry et al., 2011). The loss of VSA breadth in adults in the high
276 transmission setting could be due to various reasons, including a decline in antibody levels to VSA variants
277 due to reduced antigenic exposure, as adults have a lower parasitic load than children in this setting, or a
278 shift in the focus of the immune response to less variant targets.

279

280 **Tiled design of library facilitates high resolution characterization of seroreactive proteins**

281

282 The short length and tiling design of the peptides in this library facilitated high resolution characterization
283 of antigenic regions within seroreactive proteins. Representative examples from previously characterized
284 proteins, such as Falciparum Interspersed Repeat Antigen (FIRA), Circumsporozoite Protein (CSP) and
285 Liver Stage Antigen (LSA3) are shown (Figure 3a), where known epitopes consisted of short amino-acid
286 motifs repeated multiple times within the proteins ('repeat elements'). Comparison with a previous study
287 using a high density linear peptide array covering a subset of antigens showed substantial overlap of the
288 seroreactive regions within these antigens (Jaenisch et al., 2019) (Figure S3a), although some differences
289 were apparent. Differences in the length of peptides as well as nature of display (linear 15-aa peptides on
290 an array versus phage display of 62-aa peptides) are potential explanations for these discrepancies.

291

292 Importantly, high-resolution maps of seroreactivity for over 1000 proteins were characterized for the first
293 time in our dataset (Figure 3b). A notable example is, PHISTc (PF3D7_0801000), which has previously
294 been described as an antigenic protein, but not dissected at high resolution (Baum et al., 2013; Dent et al.,
295 2015). It is exported from the parasite during the asexual blood stage and has unknown function, though
296 mildly protective antibodies have been described against the N-terminal segment (Nagaoka et al., 2021).
297 Another example is RON4 (Rhoptry Neck Protein 4), part of the moving junction during merozoite invasion
298 of the host (Morahan 2009) and is also critical for sporozoite invasion of hepatocytes (Giovannini 2011).

299

300 Beyond the overall breadth of seroreactive peptides, the dataset facilitated a high-resolution lens for
301 investigating the effect of age and exposure on seroreactivity to individual proteins. For instance, as
302 expected (Kazmin et al., 2017), we observed exposure-dependent seropositivity at the B-cell epitope in CSP
303 targeted by the RTS,S vaccine (NANP repeating sequence) (Figure 3a). The magnitude of exposure- and
304 age-related differences in proportion seropositive varied by individual protein and even within different
305 regions of specific proteins (Figure 3, Supplementary Table S3), highlighting the importance of dissecting
306 responses to different antigenic regions within seroreactive proteins.

307

308 **Seroreactive proteins contain more repeat elements than non-seroreactive proteins.**

309

310 A prominent feature that stood out following high-resolution characterization of seroreactive regions was
311 the presence of repeat elements, where identical or similar motifs were repeated in tandem or with gaps
312 within a given protein (Fig 3). Previous studies focused on individual or targeted subsets of antigens in *P.*
313 *falciparum* have highlighted the immunogenic nature of short amino acid repeat sequences (Davies et al.,
314 2017). The proteome of *P. falciparum* is notoriously rich in such sequences; however, their functions have
315 remained enigmatic and their properties have been difficult to characterize. To systematically investigate
316 the association of seroreactivity with these elements, repeats throughout all coding sequences were first
317 identified using RADAR (Rapid Automatic Detection and Alignment of Repeats) (Madeira et al., 2019)
318 and then compared to both PhIP-seq seroreactive and non-seroreactive proteins. The number of repeats in
319 each protein sequence was significantly higher in the seroreactive proteins in comparison to non-
320 seroreactive proteins (median number of repeats per protein: seroreactive proteins – 20; non-seroreactive
321 proteins – 6; p-value = <0.001, based on 1000 iterations (1,636 proteins per iteration) of random sampling
322 of the non-seroreactive set) (Figure 4A).

323
324 **Seroreactive peptides, within seroreactive proteins, contain more repeat elements than non-**
325 **seroreactive peptides.**

326
327 Next, we investigated if seroreactive regions within seroreactive proteins were enriched for repeat elements.
328 Because the Falciparome is composed of overlapping peptides tiled across each gene, the contribution of
329 individual peptide sequences within each seroreactive protein can be further classified into those that are
330 seroreactive vs. those from the same protein that are non-seroreactive. This enables a comparison of repeat
331 elements among seroreactive and non-seroreactive peptides within each protein sequence.

332
333 To accomplish this, a k-mer approach was used to characterize repeat elements (Figure 4b, Methods).
334 Briefly, the frequency of all biochemically similar k-mers of sizes 6-9aa (approximately the size of a linear
335 B-cell epitope) was calculated for each protein. Then, each peptide in the protein was assigned a repeat
336 index based on the maximum intra-protein frequency of any repeat element it encompassed. To minimize
337 redundant representation, multiple peptides from a given protein deriving their repeat indices from the same
338 repeat element were collapsed such that a repeat element was represented only once for each protein (Figure
339 4b). In this manner, the set of all 5171 non-VSA seroreactive-peptides was collapsed based on their repeat
340 elements to a set of 3091 non-redundant seroreactive peptides. The non-seroreactive peptides within each
341 seroreactive protein were also collapsed similarly.

342
343 Overall, seroreactive peptides yielded significantly higher repeat indices than non-seroreactive peptides
344 from seroreactive proteins, and this trend was more pronounced as a function of seropositivity (Figure 4c).
345 The median repeat index for non-seroreactive peptides was 1, while the median index for >10% and >40%
346 seropositivity was 3 and 13 respectively, for a kmer of size 7 (KS test p-value < 0.05 between successive
347 distributions). These results suggest that seroreactive peptides are dominated by repeat elements and those
348 with higher seropositivity also have progressively higher repeat indices. Examination of individual
349 proteins, including well characterized repeat-containing antigens such as FIRA, LSA1, LSA3, MESA and
350 GLURP, illustrate the relationship between seropositivity and repeat index (Figure 4d). This relationship
351 was consistently observed, regardless of kmer size from 6 to 9aa, and was insensitive to the level of
352 degeneracy or biochemical similarity used for determining repeat matches (Fig S4a). However, the presence
353 of a repeat element within any given peptide does not necessarily imply that the peptide will be seroreactive.

354

355 Taken together, these data indicate that seroreactive proteins tend to be repeat-containing proteins, and
356 within these proteins, the individual seroreactive peptides tend to be those that contain the repeats.
357 Furthermore, seroreactive regions that are shared widely among individuals tend to feature higher numbers
358 of repeat elements.

359

360 **Seropositivity is more dependent on exposure for peptides containing repeat elements than those** 361 **without repeat elements**

362

363 To investigate whether the breadth of seroreactive repeat-containing peptides differed depending on
364 exposure-setting and age, seroreactive peptides were first binned into two categories: those with repeats,
365 and those without. Specifically, seroreactive peptides with a 7-mer repeat index ≥ 3 were binned together
366 as “repeat-containing peptides” and those with a 7-mer repeat index ≤ 2 were binned as “non-repeat
367 peptides”. For the set of non-repeat containing peptides, breadth (number of non-repeat peptides enriched
368 per person) was significantly higher in adults than children in both exposure settings (percent increase in
369 median breadth in adults over 4-6 year old children: moderate setting – 28%; high setting – 20%) (Fig 5A).
370 However, within each set of age groups, there was no significant difference in breadth between the two
371 exposure settings.

372

373 For repeat-containing seroreactive peptides, breadth was calculated as follows. Each repeat-containing
374 seroreactive peptide was defined by the 7-mer (repeat element) that was used to calculate its repeat index
375 as described above. To avoid redundant counting, all repeat-containing peptides from a given protein
376 defined by the same repeat element were collapsed and counted only once. Similar to non-repeat peptides,
377 breadth of these peptides was higher in adults than children, reaching a similar level in both exposure
378 settings (percent increase in median breadth in adults over 4–6 year-old children: moderate setting – 193%;
379 high setting – 56%). In contrast to non-repeat peptides however, there was an exposure dependence in the
380 responses to repeat-containing peptides with age, such that children living in the high versus moderate
381 exposure setting had twice the breadth of repeat-containing peptides, reaching the same level in adulthood
382 in both settings (Figure 5B). These results were consistently observed with different thresholds for
383 categorizing repeat-containing peptides (repeat index ≥ 4 or 5) (Fig S5A).

384

385 Investigation of individual repeat elements recapitulated this trend and showed higher seropositivity in the
386 high exposure setting compared to moderate exposure in children, but not adults (Fig S5B). There were a
387 small number of notable exceptions, including repeat elements from PHISTc (PF3D7_0801000), LSA3
388 (PF3D7_0220000), FIRA (PF3D7_0501400), all of which did not show a transmission setting-dependent
389 response in children (Supplementary Table S4). Overall, the above data show that antibody responses to
390 repeat-containing peptides may be more efficiently acquired and/or maintained in children living in settings
391 of high vs. moderate exposure, but plateau at the same level in adulthood.

392

393 **Extensive sharing of motifs observed between seroreactive proteins, particularly the PfEMP1 family**

394

395 While repeat elements within individual proteins were explored in the previous section, similar or identical
396 motifs may also be shared among different proteins. If these motifs are a part of an epitope, then antibodies
397 and B-cell receptors (BCR) specific to a motif can potentially cross-react with the motif variants in different

398 proteins, depending on accessibility and other factors. Identifying such shared motifs serves as the first step
399 in exploring potential cross-reactivity between individual seroreactive proteins, and to identify them, a
400 systematic investigation was performed.

401
402 First, enriched kmers (6-9 amino acids) were identified by collecting those present in a significantly (FDR-
403 adjusted p-value < 0.001) higher number of seroreactive peptides (9927) than a random sampling (1000
404 iterations) of 9927 peptides from the whole library. From this collection, enriched kmers that were shared
405 by different seroreactive proteins were identified as “inter-protein motifs” (Fig 6a). Using a kmer size of 7,
406 and allowing for up to two biochemically conservative substitutions, a total of 911 significantly enriched
407 inter-protein motifs were identified, representing 509 seroreactive proteins (Supplementary table S5).
408 Limiting the selection of inter-protein motifs to only the most significantly enriched motif per seroreactive
409 peptide (the motif with the lowest p-value among all motifs in each peptide) yielded 417 significant inter-
410 protein motifs, from a similar number of proteins (507). As expected, increasing the kmer size, or further
411 constraining the number of allowed substitutions resulted in fewer identified motifs (Supplementary table
412 S6). For the subsequent analysis, we show results with a kmer size of 7, which is in the range of average
413 length of a linear B-cell epitope (Buus et al., 2012). As expected, previously described cross-reactive
414 epitopes between antigens were well represented, such as the glutamate-rich motifs in Pfl 1-1 and Ag332
415 (Mattei et al. 1989), among others (Fig 6a). Taken as a group, the collected motifs had a lower
416 hydrophobicity index (mean Kyle-Doolittle = -1.95), a lower net charge (mean = -0.47) (at pH 7),
417 enrichment of charged glutamate, lysine, asparagine and aspartate residues and depletion of cysteine and
418 hydrophobic residues than a random set of motifs in the proteome (Fig S6a). These biochemical
419 characteristics are consistent with those observed in prior studies of residues in B cell epitopes (Akbar et
420 al., 2021; Rubinstein et al., 2008).

421
422 The design of the programmable phage display library used here features 62 amino acid peptides tiled with
423 a 25 amino acid step size, yielding an overlap of 37 amino acids for sequential fragments, and 12 amino
424 acids for every second fragment (Fig S6b). The design provides for localization of seroreactive sequences
425 to the region of overlap when considering adjacent fragments. For all except the first and last two peptides
426 in each protein (85% of peptides in the library), the seroreactive region can theoretically be narrowed down
427 to a 12-13aa segment within the peptide. Given that B cell linear epitopes are typically 5 -12 amino acids
428 in length (Buus et al., 2012), the 12-13aa mapping provides a near-epitope resolution.

429
430 To test the notion that the inter-protein motifs within each peptide are actually the elements associated with
431 the observed seroreactivity, we leveraged the tiled peptide library design by comparing inter-protein motif
432 carrying peptides with overlapping and adjacent peptides (Fig S6c). The maximum seropositivity among
433 peptides containing an inter-protein motif was on average 54-fold higher than the maximum seropositivity
434 among overlapping peptides not containing the motif (using a pseudo-seropositivity of 0.1% for peptides
435 with 0% seropositivity to facilitate fold change calculation), suggesting a strong association between
436 seroreactivity and the inter-protein motif itself, not just the whole peptide within which it resides
437 (comparison of median seropositivity yielded a similar result). Furthermore, a similar result was observed
438 when the same analysis was done with all the significantly enriched kmers (Fig S6d).

439
440 On average, each inter-protein motif was shared by 3 seroreactive proteins. Among the 509 seroreactive
441 proteins, each of them shared inter-protein motifs with 6 other proteins on average (median = 3),

442 (Supplementary file 1, Fig S6e). Visualized as a network (Fig 6b), the PfEMP1 family of proteins formed
443 a central hub to which a large number of other seroreactive proteins were connected. The PfEMP1 family
444 of proteins possessed at least 90 shared inter-protein motifs, and this family shared those motifs with the
445 greatest number of other seroreactive proteins (57) compared to all other proteins in this analysis.
446 Approximately 5 times as many proteins shared connections with PfEMP1 than would be expected by
447 chance (PfEMP1 shared motifs with 12-16 other proteins using a set of 9927 peptides consisting of PfEMP1
448 seroreactive peptides + random non-PfEMP1 peptides). Seroreactive proteins sharing motifs with PfEMP1
449 included many of the proteins with the highest measured seropositivity, such as RIFINs, SURFINs, FIRA,
450 and PHISTc. This extent of sharing was driven, in part, by the number of PfEMP1 sequences included in
451 the analysis. This was apparent when the same analysis performed with a reduced diversity of PfEMP1
452 sequences in the seroreactive peptide set (using PfEMP1 peptides from only PF3D7 and PFIT genomes
453 instead of 7 genomes) resulted in PfEMP1 sharing motifs with 32 seroreactive proteins instead of 57. This
454 suggests that the extent of sharing for PfEMP1 observed in this study may only be a small fraction of that
455 occurring in the extensive natural diversity of PfEMP1 variants in circulating parasites.

456
457 Outside the main network driven by PfEMP1, 495 seroreactive proteins were also found to be highly
458 connected to each other through motif sharing (Fig S6f). A large proportion of proteins with high
459 seropositivity were connected (80% and 58% of proteins with >30% and 10-30% seropositivity
460 respectively). This included proteins like GARP, LSA3, Pf332, Pf11-1, and MESA (Fig6b, Fig S6g). As
461 observed for the full set of inter protein motifs, motifs shared by the subset of proteins with >30%
462 seropositivity also consisted predominantly of charged glutamate, lysine, asparagine and aspartate residues
463 (Fig S6g). Since the analysis used here to identify inter-protein motifs allowed only up to two conservative
464 substitutions in 7-mer motifs, the similarity of motifs in the network in Fig. S6g suggests that with a less
465 stringent threshold of identifying motifs, these proteins would be even more highly connected. Moreover,
466 80% of proteins in this network had reported expression in the asexual blood stage of the lifecycle of *P.*
467 *falciparum* (PlasmoDB), suggesting temporally concordant presence of proteins sharing motifs within their
468 seroreactive regions.

469
470 These results indicate that the interprotein motifs are strongly associated with seroreactivity and are
471 extensively shared across seroreactive proteins, including among regions highly targeted by the antibodies.
472 Furthermore, PfEMP1 shares motifs with the greatest number of other seroreactive proteins, partly driven
473 by the sequence diversity of PfEMP1 variants.

474 475 **DISCUSSION**

476
477 Using a customized programmable phage display (PhIP-seq) library, we have evaluated the proteome-wide
478 antigenic landscape of the malaria parasite *P. falciparum*, using the sera of 198 individuals living in two
479 distinct malaria endemic areas. This approach readily identified previously known antigens, including
480 proteins that are targets of protective antibodies, as well as novel antigens. In our study, we characterized
481 features of *P. falciparum* antigens that could potentially contribute to the inefficient acquisition and
482 maintenance of immunity to malaria. Repeat elements were found to be commonly targeted by antibodies,
483 and had patterns of seropositivity that were more dependent on exposure than non-repeat peptides.
484 Furthermore, extensive sharing of motifs associated with seroreactivity was observed among hundreds of
485 parasite proteins, indicating potential for extensive cross-reactivity among antigens in *P. falciparum*. These

486 data suggest that repeat elements— a common feature of the *P. falciparum* proteome, and shared motifs
487 between antigenic proteins could have important roles in shaping the nature and development of the immune
488 response to malaria.

489
490 To map the antigenic landscape, PhIP-seq for *P. falciparum* offers several attractive advantages. The library
491 described here contains >99.5% of the proteome, including variants for several antigenic families,
492 surpassing the coverage of other existing proteome-wide tools for *P. falciparum* (Camponovo et al., 2020;
493 Morita et al., 2017), while simultaneously providing high-resolution characterization of antigens (up to 12-
494 13 aa regions within peptides). Unlike peptide arrays, the platform converts a proteomic assay into a
495 genomic assay, leveraging the massive scale and low-cost nature of next-generation short-read sequencing.
496 The result is a cost-effective and scalable system, allowing for the processing of hundreds of samples in
497 parallel. Finally, an important aspect of all phage display systems is the ability to sequentially enrich, release
498 phage, and repeat, thus greatly amplifying the signal to noise (O'Donovan et al., 2020; Smith & Petrenko,
499 1997). Only one published study to date evaluates responses to more than a quarter of the proteome
500 (Camponovo et al., 2020), inherently limiting the scope of potential targets interrogated. 28% of the hits
501 described in this study of individuals living in Tanzania and exposed to various doses of PfSpz vaccine
502 overlapped with the hit proteins described in our study (Camponovo et al., 2020)). The limited overlap
503 may be due to multiple factors, including differences in the characteristics of individuals sampled, the use
504 of vaccine, and determination of seropositivity based on technical as opposed to biological controls (sera
505 from unexposed individuals).

506
507 The near-epitope resolution provided by this platform allowed a systematic investigation of targets of
508 antibodies. Targets with high seropositivity were observed to be significantly enriched for repeat elements.
509 In some previous reports, the elevated antigenic potential of repeat elements has been noted (Davies et al.,
510 2017), however the proteome-wide approach described here demonstrates that a large collection of proteins
511 containing these elements are highly immunogenic. The high immunogenicity of repeat sequences observed
512 here may be the result of competitive advantages that B cell clones could encounter when binding to higher
513 valency epitopes, as opposed to single copy epitopes. Evidence from experimental inoculations of antigens
514 with differing repeat numbers support this notion (Kato et al., 2020). Moreover, tandem repeat regions are
515 predicted to be intrinsically disordered, which in turn have favorable predictions as linear B cell epitopes
516 (Guy et al., 2015). Notably, this high immunogenicity can potentially restrict responses to other epitopes
517 within the antigen, as has been reported for responses to protective non-repeat epitopes in the
518 circumsporozoite protein (CSP). (Chatterjee et al., 2021).

519
520 A key finding of this study is the exposure-setting dependent difference in seroreactivity to repeat-
521 containing peptides, with the breadth of seroreactivity increasing more quickly with age in the high versus
522 moderate exposure setting. We note that the samples analyzed in this study differed between the two cohorts
523 not just by exposure, but also with respect to time since most recent infection, reflecting the differing
524 epidemiology of infection in these settings. In the moderate exposure setting, the median number of days
525 since last infection was 100, whereas over 65% of the samples from the high exposure setting were taken
526 during periods of active infection. The difference in seroreactivity to repeat-containing peptides observed
527 here between the settings could therefore emerge from two related mechanisms. In the first, the difference
528 could be driven by a requirement for a minimum level of cumulative exposure to the target repeats to
529 generate a robust response. In the second, the antibody response to repeats may be inherently less durable,

530 leading to rapid waning in the absence of frequent exposure. Future longitudinal studies may be required
531 to distinguish between these two possibilities. There were a few exceptions, including repeats from FIRA,
532 PHISTc and LSA3, that did not show an exposure-setting dependent difference in seropositivity, suggesting
533 that factors beyond the repeated nature of the epitope could influence the nature of the response. Whether
534 either or both potential mechanisms contribute, the predominance of repeat containing peptides in antibody
535 targets, along with the remarkable abundance of these peptides in the *P. falciparum* proteome, suggests a
536 possible strategy evolved by the parasite for the purpose of diverting the humoral response towards short-
537 lived or exposure-dependent responses.

538
539 The hypothesis of less durable antibody responses to repeat antigens in *P. falciparum* can be reconciled
540 with a model in which repeating epitopes favor extrafollicular B cell responses, which are typically short-
541 lived (Cockburn & Seder, 2018; Schofield, 1991). This is based on the potential of repeat epitopes in an
542 antigen to interact with multiple B-cell receptors (BCRs) on naïve B-cells, thereby conferring high binding
543 strength and sufficient activation to direct these cells into an early extrafollicular response and production
544 of short-lived plasmablasts. Several studies provide support to this model, where strong binding of BCR to
545 the antigen, including through increased valency, increases the production of plasmablasts (Kato et al.,
546 2020; O'Connor et al., 2006; Ochiai et al., 2013; Paus et al., 2006; Schwickert et al., 2011). This could also
547 happen via a T-cell independent response, as has been reported for some repeat antigenic structures
548 (Schofield, 1991; Schofield & Uadia, 1990). On the other hand, the effect on germinal centers (GC), which
549 result in long-lived plasma cells (LLPCs) and isotype-switched memory cells, is unclear. While some
550 studies have reported no change or a decrease in the formation of GCs (O'Connor et al., 2006; Ochiai et
551 al., 2013; Paus et al., 2006) with increased strength of interaction between antigen and B-cells, some have
552 reported an increase (Kato et al., 2020; Schwickert et al., 2011), though it is unclear whether the latter were
553 productive GCs. More insights come from a few studies that measured the outcome of GCs following
554 increased strength of BCR-antigen interaction and these have reported a decrease in LLPCs (Fink et al.,
555 2007) and IgG-switched memory cells (Pape et al., 2018; Taylor et al., 2015). If repeat antigens in *P.*
556 *falciparum* follow this pattern, an expected outcome would be defective formation of LLPCs and memory
557 B cells. On the other hand, the finding that adults from both exposure settings ultimately developed a similar
558 breadth of response to repeat regions could argue for the hypothesis that greater cumulative exposure is
559 required to develop responses to these regions, but the difference between adults and children could also be
560 driven by age-intrinsic factors (Baird et al., 1991, 1993).

561
562 Another major finding of this study is the extensive presence of inter-protein motifs among seroreactive
563 proteins. Since a strong association with seroreactivity was observed for these motifs, they may represent
564 cross-reactive epitopes. Whether these inter-protein motifs are cross-reactive *in vivo* is unclear and may
565 depend on expression timing and accessibility to the immune system, among other factors. Analogously,
566 seroreactive repeat elements with non-identical repeating units could represent cross-reactive epitopes
567 within proteins. Extensive presence of potential cross-reactive epitopes in *P. falciparum* antigens may play
568 an important role in influencing the quality of the immune response to malaria. While it could be
569 advantageous for the host if multiple parasite proteins could be targeted by antibodies through cross-
570 reactivity, simultaneous presence of cross-reactive epitopes could alternatively frustrate the affinity
571 maturation process due to conflicting selection forces, as was observed for variant HIV antigens (Wang et
572 al., 2015). Further, recurrent exposure may be necessary for the generation of strong cross-binding
573 antibodies to cross-reactive epitopes (Murugan et al., 2020). Thus, the extensive presence of cross-reactive

574 epitopes, both within and between antigenic proteins in *P. falciparum*, could represent an evolutionary
575 strategy aimed at limiting high-affinity antibodies in favor of lower affinity, cross reactive antibodies. In
576 essence, the large number of shared seroreactive sequences in *P. falciparum* may represent a complex
577 immune counter measure, resulting in inefficient immunity acquisition which requires extensive exposure.
578 The atlas of seroreactive repeat elements and inter-protein motifs from this study will be useful for future
579 investigations in understanding their impact on the quality of immune response to malaria.

580
581 The PfEMP1 family shared inter-protein motifs with the greatest number of other antigens in this study.
582 This was driven in part by the wide diversity of PfEMP1 variants, indicating that as one becomes naturally
583 exposed to different PfEMP1 variants (Cham et al., 2009), there may be an increase in not only the sequence
584 diversity, but the number of cross-reactive epitopes that the immune system encounters. Possessing cross-
585 reactive epitopes with other antigens could result in binding of pre-existing antibodies to the new variants,
586 which could be disadvantageous to the host if binding strength is weak. Further, cross-reactivity may inhibit
587 generation of antibodies specific to the new variant due to original antigenic sin (Vatti et al., 2017). Thus,
588 the mechanism of evolving PfEMP1 variants within a network of shared sequences with other antigens
589 could be another strategy evolved by the parasite for immune evasion. On the other hand, binding of new
590 variants to pre-existing antibodies may be advantageous to the host if those antibodies are effective against
591 the new variants.

592
593 While phage display of small peptides yields high resolution discrimination of linear epitopes, this approach
594 may not capture antibodies binding to conformational epitopes. Therefore, such epitopes are likely to be
595 missed by this assay, although polyclonal responses are frequently a mixture of linear, partially linear, and
596 conformational epitopes. Reassuringly, we observed a large-scale enrichment of *P. falciparum* peptide
597 sequences in exposed individuals when compared with control sera from the US. This suggests that the
598 humoral immune system of exposed individuals acquires an extensive and diverse set of *P. falciparum*
599 targets, including thousands of linear sequences. The bias towards linear epitopes may have increased the
600 relative detectability of repeat regions by this assay since they often form intrinsically disordered regions.
601 However, that would not account for the observed differences between exposure settings for children and
602 adults. Another limitation of our study is that it did not provide quantitative measures of absolute antibody
603 reactivity to individual peptides per person. Therefore, enrichment counts for peptides were only used in a
604 semi-quantitative way to determine seropositivity. Lastly, given the breadth and sensitivity of the PhIP-seq
605 technique, 86 control sera were used to remove non-specific enrichments. We imposed a stringent filter to
606 minimize false positives by requiring that each seroreactive peptide be enriched in at least five Ugandan
607 samples over control sera. While this excluded possible seroreactive peptides unique to a single sample, the
608 resulting sequences that passed were those that exhibited a minimum level of sharing among multiple
609 individuals, thereby enriching for those seroreactive peptides that represent common serological responses
610 to malaria.

611
612 With the rapid success of mRNA vaccines for SARS-CoV-2(Chaudhary et al., 2021), an optimistic future
613 for malaria vaccines is possible as well. Findings from this study could have important implications on
614 malaria vaccine design. Results from our study suggest that that in natural infections in children, repeat
615 regions in *P. falciparum* could lead to an exposure-dependent and/or short-lived antibody response to a
616 higher degree than for non-repeat regions. While we recognize that vaccine induced immunity is distinct
617 from naturally acquired immunity, this potential limitation should be considered when evaluating repeat-

618 containing antigens as vaccine targets. Further, given that highly immunogenic regions in natural immunity
619 to malaria are predominantly repeats and there is widespread presence of potential cross-reactive epitopes
620 across many proteins, whole-parasite vaccines may also be susceptible to similar limitations. If the findings
621 from this study translate to vaccine-induce immune responses, non-repetitive, unique antigenic regions may
622 be more effective targets.

623

624 **MATERIALS AND METHODS**

625

626 **Ethical Approval**

627 The study protocol was reviewed and approved by the Makerere University School of Medicine Research
628 and Ethics Committee (Identification numbers 2011–149 and 2011–167), the London School of Hygiene
629 and Tropical Medicine Ethics Committee (Identification numbers 5943 and 5944), the University of
630 California, San Francisco, Committee on Human Research (Identification numbers 11–05539 and 11–
631 05995) and the Uganda National Council for Science and Technology (Identification numbers HS-978 and
632 HS-1019). Written informed consent was obtained from all participants in the study. For children, this was
633 obtained from the parents or guardians.

634

635 **Study Sites and Participants**

636 Plasma samples for the study were obtained from the Kanungu and Tororo sites of the Program for
637 Resistance, Immunology, and Surveillance of Malaria (PRISM) cohort studies, part of the East African
638 International Centers of Excellence in Malaria Research (Kamya et al., 2015). Kihhi sub-county in
639 Kanungu district is a rural highland area in southwestern Uganda characterized by moderate transmission;
640 samples used from this region were collected between 2012-2016. Nagongera sub-county in Tororo district
641 is a rural area in southeastern Uganda with high transmission and samples used from this region were
642 collected between Aug-Sep 2012. Samples from Tororo were restricted to individuals with fewer than 6
643 malaria cases per year to exclude individuals with very high incidence. Entomological inoculation rates
644 (EIR) used in the study were calculated for each household based on entomological surveys involving
645 collection of mosquitoes with CDC light traps and quantifying the number of *P. falciparum*-containing
646 female anopheles mosquitoes along with sporozoite rates (Kilama et al., 2014). These cohorts and study
647 sites have been described extensively in prior publications (Helb et al., 2015; Kamya et al., 2015; Rek et
648 al., 2016; Yeka et al., 2015); briefly, follow up consisted of continuous passive surveillance for malaria at
649 a study clinic open 7 days a week where all routine medical care was provided, routine active surveillance
650 for parasitemia, and routine entomologic surveillance. One plasma sample was selected from each of 100
651 participants, stratified by age, from each of the 2 cohorts. The 86 US controls were de-identified plasma
652 obtained from adults who donated blood to the New York Blood Center.

653

654 **Bioinformatic Construction of Falciparome Phage Library**

655 The pipeline for library construction is shown in Fig S2a. To construct the library, raw protein sequence
656 files were downloaded from their respective public databases. Coding sequences from 3D7 and IT strains
657 were downloaded from PlasmoDB (Amos et al., 2022) and vaccine/viral sequences were downloaded from
658 the RefSeq database (O’Leary et al., 2016). Antigenic variant sequences were curated from multiple
659 sources. The entire collection of protein sequences used as input for designing the peptides in the study can
660 be found in the Dryad dataset doi:10.7272/Q69S1P9G. Pseudogenes were removed and any remaining stop
661 codons within coding sequences were replaced with Alanine residues. These sequences were combined and

662 filtered using CD-HIT(Fu et al., 2012; Li & Godzik, 2006) to remove sequences with > x% identity, where
663 the threshold X used varied for different sets of sequences are in Table 2.

664
665 The final set of protein sequences (n=8,980) was then merged and short sequences (<30 aa long) were
666 removed prior to collapsing at 100% sequence identity (n = 8534). Next, all sequences were split into 62-
667 amino acid peptide fragments with 25-amino acid step size. Fragments with homopolymer runs of >= 8
668 exact amino acid matches in a row were removed, X amino acids were substituted to Alanine and Z amino
669 acids (Glutamic acid or Glutamine) to Q (Glutamine), and finally, lzw compression was used to identify
670 and remove low-complexity sequences with a compression ratio less than 0.4. Lastly, sequence headers
671 were renamed to remove spaces and the resulting peptide fragments were converted to nucleotide
672 sequences. Adapter sequences were appended, with a library-specific linker on the 5' end
673 (GTGGTTGGTGTAGGAGCA) and a 3' linker sequence coding for two stop codons and a 17mer (-
674 TGATAA- GCATATGCCATGGCCTC). This file was then iteratively scanned for restriction enzyme sites
675 (EcoRI, HindIII), which were eliminated by replacement with synonymous codons to facilitate cloning.
676 The final set of nucleotide sequences was collapsed at 100% nucleotide sequence identity (n = 238,068)
677 and then ordered from Agilent Technologies.

678
679 **Cloning and Packaging into T7 Phage**

680 A single vial of lyophilized DNA was received from Agilent. The lyophilized oligonucleotides were
681 resuspended in 10 mM Tris-HCl-1 mM EDTA, pH 8.0 to a final concentration of 0.2nM and PCR amplified
682 for cloning into T7 phage vector arms (Novagen/EMD Millipore Inc. T7 Select 10-3 Cloning kit). Detailed
683 protocol can be found in [dx.doi.org/10.17504/protocols.io.j8nlkkr515r/v1](https://doi.org/10.17504/protocols.io.j8nlkkr515r/v1). Four 30ul packaging reactions
684 were performed and all were pooled in the end. Plaque assays were done with the packaging reaction to
685 determine the titer of infectious phage in the packaging reaction and estimated to be 2×10^8 pfu/ml. Phage
686 libraries were then prepared and amplified fresh from packaging reactions. Resulting phage libraries were
687 tittered by plaque assay and adjusted to a working concentration of 10^{10} pfu/mL before incubation with
688 patient plasma.

689 **Immunoprecipitation of antibody-bound phage**

690 Plasma samples were first diluted in 1:1 storage buffer (0.04% NaN₃, 40% Glycerol, 40mM HEPES (pH
691 7.3), 1x PBS (-Ca and -Mg)) to preserve antibody integrity. Then, a 1:2.5x dilution of that stock was made
692 in 1x PBS resulting in a final 1:5 dilution and 1 ul of this was used in PhIP-seq. The protocol was followed
693 as in [dx.doi.org/10.17504/protocols.io.j8nlkkr515r/v1](https://doi.org/10.17504/protocols.io.j8nlkkr515r/v1). 40 ul of Pierce Protein A/G Bead slurry
694 (ThermoFisher Scientific) were used per sample. After round 1 of IP, the eluted phage were amplified in *E.*
695 *coli* and enriched through a second round of IP. The final lysate was spun and stored at -20°C for NGS
696 library prep. Immunoprecipitated phage lysate was heated to 70°C for 15 minutes to expose DNA. DNA
697 was then amplified in two subsequent reactions. All samples had a minimum of two technical replicates.

698
699 **Bioinformatic Analysis of PhIP-Seq Data**

700
701 **Identification of seroreactive peptides**

702 Sequencing reads were first trimmed to cut out adaptors with Cutadapt(Martin, 2011):
703 `call(['cutadapt', '-g', r1_linker_dict[index_library], '-o', read1_trimmed, read1_location]))`
704 Trimmed reads were then aligned to the full FalciParome peptide library using GSNAP(Wu & Nacu, 2010)
705 paired end alignment, outputting a SAM file:

706 call(["/data/bin/bin/gsnap", "--gunzip", "-A", "sam", "--gmap-mode=none", "--batch=2", "--nofails", "--
707 npaths=1", "-t", "12", "--use-shared-memory=0", "-d", (gsnap_library), "-D", gsnap_libraries,
708 read1_location, read2_location], stdout = f)

709 For each aligned sequence, the CIGAR string was examined, and all alignments where the CIGAR string
710 did not indicate a perfect match were removed. The final set of peptides was tabulated to generate counts
711 for each peptide in each individual sample. Samples with less than 250,000 aligned reads were dropped
712 from further analysis and any resulting samples with only one technical replicate were also dropped (2 of
713 the 200 Ugandan samples were dropped). To keep the analysis restricted to *P. falciparum* peptides and limit
714 the influence from non-*P. falciparum* peptides, reads mapping to all vaccine, viral and experimental control
715 peptides were excluded from analysis. The remaining peptide counts were normalized for read depth and
716 multiplied by 500,000, resulting in reads/500,000 total reads (RP5K) for each peptide. The null distribution
717 for each peptide was modelled using read counts from a set of 86 plasma from the US (New York Blood
718 Center) using a normal distribution, with the assumption that most of these individuals were likely
719 unexposed to malaria. To avoid inflation by division, if the standard deviation of read counts of any peptide
720 in the US samples was <1, then that was set as 1. Z-score enrichments ((x-mean US)/std. dev US) were
721 then calculated for each peptide in each sample using the US distribution and Z-score ≥ 3 in both technical
722 replicates (or more than 75% of the replicates if there were more than 2 technical replicates) of a sample
723 was used to identify enriched peptides within a given sample. To call malaria-specific peptide enrichments
724 ('seroreactive peptides'), enrichment was required in a minimum of 5 Ugandan samples. Seropositivity for
725 a peptide was calculated by the percent of Ugandan samples enriching for that peptide. Seropositivity for a
726 protein was calculated by the percent of Ugandan samples enriching for any peptide within that protein.

727

728 **Calculation of breadth of non-redundant peptide groups per person**

729 Seroreactive peptides in each person were collapsed based on shared sequences (7-mer identical motifs)
730 using the network approach described in AVARDA (Monaco et al., 2021) to get a conservative estimate of
731 the number of non-redundant peptide groups in each person.

732

733 **Calculation of VSA breadth per person**

734 VSA breadth was calculated as the number of variant proteins in each VSA family that were seroreactive
735 in a given person and was calculated as follows. Since all these families possess conserved as well as
736 variable regions, during library design, peptides across conserved regions from many variants that share
737 identical sequences were filtered out to avoid redundant representation and only one representative peptide
738 was retained in the final Falciparome library. Therefore, to accurately calculate the number of VSA proteins
739 recognized in a person, all seroreactive VSA peptides were mapped back to the sequences from the full
740 collection of VSA protein sequences to identify all the variant proteins each seroreactive peptide sequence
741 mapped to. This information was then used to get the number of variant proteins reactive to a person's
742 plasma. Domain classification for PfEMP1s was done using the VarDom server (Rask et al., 2010). Domain
743 classification for RIFINs was done based on (Joannin et al., 2008).

744

745 **Repeat analysis**

746 Only unique 3D7/IT proteins in the library (if both 3D7 and IT homologs were present in the library, only
747 the 3D7 homolog was considered) that were not members of Variant Surface Antigens (PfEMP1, RIFIN,
748 STEVOR, SURFIN, Pfmc-2TM) were considered for all repeat analysis to avoid redundancy of
749 representation.

750

751 *Cumulative repeat frequency in proteins* - For calculation of cumulative repeat frequency in proteins,
752 amino acid sequences of proteins were input into the RADAR (Madeira et al., 2019) program for denovo
753 identification of repeats using default settings. Cumulative frequency of repeats in a protein was then
754 determined by adding the repeat counts of all reported repeats in the protein. To compare to the non-
755 seroreactive set, the same number of proteins as the seroreactive protein set was randomly sampled from
756 the total non-seroreactive protein set 1000 times and the distribution of cumulative frequencies between the
757 seroreactive and non-seroreactive sets were compared using a 2-sample KS-test in each iteration.

758

759 *Repeat index calculation* - To systematically compare the distribution of repeats between seroreactive and
760 non-seroreactive peptides within seroreactive proteins, the following approach was adopted. Firstly, for
761 each protein, repeats and their frequency within that protein was calculated using a k-mer approach. K-mers
762 were fixed length sequences (6/7/8/9-aa) with any number of conservative substitutions (AG, DE, RHK,
763 ST, NQ, LVI, YFW) and did not include polymeric stretches of single amino acids from N/Q/D/E/R/H/K.
764 For each protein sequence, all possible kmers in the protein and their frequency (number of non-overlapping
765 occurrences) in the protein (intra-protein repeat frequency) was calculated. Then for each peptide in the
766 protein, all k-mers in the peptide sequence were taken and the k-mer with the highest intra-protein repeat
767 frequency was identified. This frequency was assigned as the repeat index for the peptide. Once all peptides
768 across all seroreactive proteins were assigned a repeat index, they were subsequently classified according
769 to seropositivity. In each seropositivity group, since peptides from the same protein could have the same
770 highest intra-protein repeat k-mer, to avoid redundancy of representation, peptides sharing the same highest
771 k-mer were collapsed and counted only once. For the non-seroreactive peptide set, random sampling of
772 peptides from all non-seroreactive peptides was performed (1000 iterations). The 2-sample KS test was
773 then used to compare distributions.

774

775 **Inter-protein motif analysis**

776 First, all motifs with wildcards (any amino acid allowed at that position) or conservative substitutions (AG,
777 DE, RHK, ST, NQ, LVI, YFW), shared by at least two seroreactive peptides were identified using the
778 SLiMfinder program (Edwards et al., 2007), a part of the SLiMSuite package. The following parameters
779 were used for running the program with the seroreactive peptide sequences as input:

780 teiresias=T efilter=F blastf=F masking=F fmask=F imask=F compmask=F metmask=F slimlen=7
781 absmin=2 absminamb=2 slimchance=F maxwild=1 maxseq=10000 walltime=240 minocc=0.0002
782 ambocc=0.0003 wildvar=False equiv=<txt file that lists the allowed conservative substitutions - AG, DE,
783 RHK, ST, NQ, LVI, YFW>

784

785 Following this, a custom script was used to parse motifs with desired length and degeneracy threshold and
786 identify those enriched over background. First, motifs of length K with at least N fixed positions and
787 allowed number of conservative substitutions and wildcards were filtered depending on the degeneracy
788 thresholds used. Motifs with homopolymeric stretches of KKKKK/ NNNNN/ EEEEE were not considered
789 as this is a common feature in the proteome of plasmodium. Then, for each motif, the number of seroreactive
790 peptides possessing that motif was determined (frequency in the seroreactive set). Next, enrichment in the
791 seroreactive set over background was estimated with the following approach. Random sampling was
792 performed on the whole library to get the same number of random peptides as seroreactive peptides
793 (n=9927) and the occurrence frequency of each motif was calculated in the random set each time. This was

794 bootstrapped 1000 times and this represented the background frequency of the motifs in 1000 iterations. A
795 p-value for enrichment in the seroreactive set was then calculated using a Poisson model for the background
796 frequency distribution. Significantly enriched ones were then identified following multiple hypothesis
797 correction (FDR of 0.1%). This set of motifs represented the final collection of significantly enriched
798 motifs. From this set, those that were shared by at least two seroreactive proteins were identified as inter-
799 protein motifs. Network visualizations were performed with Cytoscape (Shannon et al., 2003). For the
800 analysis on PfEMP1 with random set of peptides, all PfEMP1 peptides from the seroreactive set (n=3001)
801 were combined with random peptides (n=6926) to a total of 9927 peptides. This was treated as the
802 ‘seroreactive’ set and a similar analysis was performed to identify significantly enriched motifs in this set.
803

804 **DATA AVAILABILITY**

805
806 The data associated with this study can be accessed in the Dryad repository with the
807 doi:10.7272/Q69S1P9G
808 (https://datadryad.org/stash/share/YuYmQNKNvrWmoMX8n99wle_2bFyrtweAGclxYPhkPjY).
809

810 **ACKNOWLEDGEMENTS**

811
812 We thank all study participants who participated in this study and their families. We thank the New York
813 Blood Center for providing us with the de-identified human control plasma samples. We thank Caleigh
814 Mandel-Brehm for advice and discussions on PhIP-seq. We thank members of the DeRisi and Greenhouse
815 labs for helpful discussions.
816

817 **FUNDING**

818
819 Joseph L. DeRisi - Chan Zuckerberg Biohub
820 Bryan Greenhouse - CZB Investigator program, NIH/NIAID awards A1089674 (East Africa ICEMR),
821 AI119019, and AI144048
822 The funders had no role in study design, data collection and interpretation, or the decision to submit the
823 work for publication.
824

825 **COMPETING INTERESTS**

826
827 Joseph L DeRisi: Paid scientific advisor for Allen & Co. Paid scientific advisor for the Public Health
828 Company, Inc. and holds stock options. Founder and holding stock options in VeriPhi Health, Inc..
829 All other authors declare no competing interests.
830

831 **REFERENCES**

832
833 Akbar, R., Robert, P. A., Pavlović, M., Jeliaskov, J. R., Snapkov, I., Slabodkin, A., Weber, C. R.,
834 Scheffer, L., Miho, E., Haff, I. H., Haug, D. T. T., Lund-Johansen, F., Safonova, Y., Sandve, G. K.,
835 & Greiff, V. (2021). A compact vocabulary of paratope-epitope interactions enables predictability of
836 antibody-antigen binding. *Cell Reports*, 34(11), 108856.
837 <https://doi.org/10.1016/J.CELREP.2021.108856>
838 Amos, B., Aurrecochea, C., Barba, M., Barreto, A., Basenko, E. Y., Ba' Zant, W., Belnap, R., Blevins,

- 839 A. S., Ohme, U. B. ", Brestelli, J., Brunk, B. P., Caddick, M., Callan, D., Campbell, L., Christensen,
840 M. B., Christophides, G. K., Crouch, K., Davis, K., Debarry, J., ... Zheng, J. (2022). VEuPathDB:
841 the eukaryotic pathogen, vector and host bioinformatics resource center. *Nucleic Acids Research*,
842 50(D1), D898–D911. <https://doi.org/10.1093/NAR/GKAB929>
- 843 Anders, R. F. (1986). Point of view Multiple cross-reactivities amongst antigens of Plasmodium
844 falciparum impair the development of protective immunity against malaria. In *Parasite Immunology*
845 (Vol. 8).
- 846 Baird, J. K., Jones, T. R., Danudirgo, E. W., Annis, B. A., Bangs, M. J., Basri, P. H., & Masbar, S.
847 (1991). Age-Dependent Acquired Protection against Plasmodium Falciparum in People Having Two
848 Years Exposure to Hyperendemic Malaria. *The American Journal of Tropical Medicine and*
849 *Hygiene*, 45(1), 65–76. <https://doi.org/10.4269/AJTMH.1991.45.65>
- 850 Baird, J. K., Purnomo, Basri, H., Bangs, M. J., Andersen, E. M., Jones, T. R., Masbar, S., Harjosuwarno,
851 S., Subianto, B., & Arbani, P. R. (1993). Age-Specific Prevalence of Plasmodium falciparum
852 Among Six Populations with Limited Histories of Exposure to Endemic Malaria. *The American*
853 *Journal of Tropical Medicine and Hygiene*, 49(6), 707–719.
854 <https://doi.org/10.4269/AJTMH.1993.49.707>
- 855 Barry, A. E., Trieu, A., Fowkes, F. J. I., Pablo, J., Kalantari-Dehaghi, M., Jasinskas, A., Tan, X., Kayala,
856 M. A., Tavul, L., Siba, P. M., Day, K. P., Baldi, P., Felgner, P. L., & Doolan, D. L. (2011). The
857 stability and complexity of antibody responses to the major surface antigen of Plasmodium
858 falciparum are associated with age in a malaria endemic area. *Molecular and Cellular Proteomics*,
859 10(11). <https://doi.org/10.1074/mcp.M111.008326>
- 860 Baum, E., Badu, K., Molina, D. M., Liang, X., Felgner, P. L., & Yan, G. (2013). Protein Microarray
861 Analysis of Antibody Responses to Plasmodium falciparum in Western Kenyan Highland Sites with
862 Differing Transmission Levels. *PLoS ONE*, 8(12), e82246.
863 <https://doi.org/10.1371/journal.pone.0082246>
- 864 Buus, S., Rockberg, J., Forsström, B., Nilsson, P., Uhlen, M., & Schafer-Nielsen, C. (2012). High-
865 resolution Mapping of Linear Antibody Epitopes Using Ultrahigh-density Peptide Microarrays.
866 *Molecular & Cellular Proteomics : MCP*, 11(12), 1790. <https://doi.org/10.1074/MCP.M112.020800>
- 867 Camponovo, F., Campo, J. J., Le, T. Q., Oberai, A., Hung, C., Pablo, J. V., Teng, A. A., Liang, X., Sim,
868 B. K. L., Jongo, S., Abdulla, S., Tanner, M., Hoffman, S. L., Daubenberger, C., & Penny, M. A.
869 (2020). Proteome-wide analysis of a malaria vaccine study reveals personalized humoral immune
870 profiles in Tanzanian adults. *ELife*, 9. <https://doi.org/10.7554/eLife.53080>
- 871 Cham, G. K. K., Turner, L., Lusingu, J., Vestergaard, L., Mmbando, B. P., Kurtis, J. D., Jensen, A. T. R.,
872 Salanti, A., Lavstsen, T., & Theander, T. G. (2009). Sequential, Ordered Acquisition of Antibodies
873 to Plasmodium falciparum Erythrocyte Membrane Protein 1 Domains. *The Journal of Immunology*,
874 183(5), 3356–3363. <https://doi.org/10.4049/jimmunol.0901331>
- 875 Chatterjee, D., Lewis, F. J., Sutton, H. J., Kaczmarek, J. A., Gao, X., Cai, Y., McNamara, H. A., Jackson,
876 C. J., & Cockburn, I. A. (2021). Avid binding by B cells to the Plasmodium circumsporozoite
877 protein repeat suppresses responses to protective subdominant epitopes. *Cell Reports*, 35(2).
878 <https://doi.org/10.1016/J.CELREP.2021.108996>
- 879 Chaudhary, N., Weissman, D., & Whitehead, K. A. (2021). mRNA vaccines for infectious diseases:
880 principles, delivery and clinical translation. *Nature Reviews Drug Discovery* 2021 20:11, 20(11),
881 817–838. <https://doi.org/10.1038/s41573-021-00283-5>
- 882 Clinical Trials Partnership, S. (2015). Efficacy and safety of RTS,S/AS01 malaria vaccine with or without
883 a booster dose in infants and children in Africa: final results of a phase 3, individually randomised,
884 controlled trial. *The Lancet*, 386(9988), 31–45. [https://doi.org/10.1016/S0140-6736\(15\)60721-8](https://doi.org/10.1016/S0140-6736(15)60721-8)
- 885 Cockburn, I. A., & Seder, R. A. (2018). Malaria prevention: from immunological concepts to effective
886 vaccines and protective antibodies. *Nature Immunology* 2018 19:11, 19(11), 1199–1211.
887 <https://doi.org/10.1038/s41590-018-0228-6>
- 888 Crompton, P. D., Kayala, M. A., Traore, B., Kayentao, K., Ongoiba, A., Weiss, G. E., Molina, D. M.,
889 Burk, C. R., Waisberg, M., Jasinskas, A., Tan, X., Doumbo, S., Doumtabe, D., Kone, Y., Narum, D.

- 890 L., Liang, X., Doumbo, O. K., Miller, L. H., Doolan, D. L., ... Pierce, S. K. (2010). A prospective
891 analysis of the Ab response to Plasmodium falciparum before and after a malaria season by protein
892 microarray. *Proceedings of the National Academy of Sciences of the United States of America*,
893 *107*(15), 6958–6963. <https://doi.org/10.1073/pnas.1001323107>
- 894 Davies, H. M., Nofal, S. D., McLaughlin, E. J., & Osborne, A. R. (2017). Repetitive sequences in malaria
895 parasite proteins. *FEMS Microbiology Reviews*, *41*(6), 923.
896 <https://doi.org/10.1093/FEMSRE/FUX046>
- 897 Dent, A. E., Nakajima, R., Liang, L., Baum, E., Moormann, A. M., Sumba, P. O., Vulule, J., Babineau,
898 D., Randall, A., Davies, D. H., Felgner, P. L., & Kazura, J. W. (2015). *Plasmodium falciparum*
899 Protein Microarray Antibody Profiles Correlate With Protection From Symptomatic Malaria in
900 Kenya. *Journal of Infectious Diseases*, *212*(9), 1429–1438. <https://doi.org/10.1093/infdis/jiv224>
- 901 Doolan, D. L., Dobaño, C., & Baird, J. K. (2009). Acquired immunity to Malaria. In *Clinical*
902 *Microbiology Reviews* (Vol. 22, Issue 1, pp. 13–36). <https://doi.org/10.1128/CMR.00025-08>
- 903 Edwards, R. J., Davey, N. E., & Shields, D. C. (2007). SLiMfinder: A Probabilistic Method for
904 Identifying Over-Represented, Convergently Evolved, Short Linear Motifs in Proteins. *PLOS ONE*,
905 *2*(10), e967. <https://doi.org/10.1371/JOURNAL.PONE.0000967>
- 906 Feldmann, M., & Basten, A. (1971). THE RELATIONSHIP BETWEEN ANTIGENIC STRUCTURE
907 AND THE REQUIREMENT FOR THYMUS-DERIVED CELLS IN THE IMMUNE RESPONSE.
908 *Journal of Experimental Medicine*, *134*(1), 103–119. <https://doi.org/10.1084/JEM.134.1.103>
- 909 Fink, K., Manjarrez-Orduño, N., Schildknecht, A., Weber, J., Senn, B. M., Zinkernagel, R. M., &
910 Hengartner, H. (2007). B Cell Activation State-Governed Formation of Germinal Centers following
911 Viral Infection. *The Journal of Immunology*, *179*(9), 5877–5885.
912 <https://doi.org/10.4049/JIMMUNOL.179.9.5877>
- 913 Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation
914 sequencing data. *Bioinformatics (Oxford, England)*, *28*(23), 3150–3152.
915 <https://doi.org/10.1093/BIOINFORMATICS/BTS565>
- 916 Guy, A. J., Irani, V., MacRaild, C. A., Anders, R. F., Norton, R. S., Beeson, J. G., Richards, J. S., &
917 Ramsland, P. A. (2015). Insights into the Immunological Properties of Intrinsically Disordered
918 Malaria Proteins Using Proteome Scale Predictions. *PLoS ONE*, *10*(10).
919 <https://doi.org/10.1371/JOURNAL.PONE.0141729>
- 920 Helb, D. A., Tetteh, K. K. A., Felgner, P. L., Skinner, J., Hubbard, A., Arinaitwe, E., Mayanja-Kizza, H.,
921 Ssewanyana, I., Kamya, M. R., Beeson, J. G., Tappero, J., Smith, D. L., Crompton, P. D., Rosenthal,
922 P. J., Dorsey, G., Drakeley, C. J., Greenhouse, B., & Rayner, J. C. (2015). Novel serologic
923 biomarkers provide accurate estimates of recent Plasmodium falciparum exposure for individuals
924 and communities. *Proceedings of the National Academy of Sciences of the United States of America*,
925 *112*(32), E4438–E4447. <https://doi.org/10.1073/pnas.1501705112>
- 926 Hou, N., Jiang, N., Ma, Y., Zou, Y., Piao, X., Liu, S., & Chen, Q. (2020). Low-Complexity Repetitive
927 Epitopes of Plasmodium falciparum Are Decoys for Humoural Immune Responses. *Frontiers in*
928 *Immunology*, *11*. <https://doi.org/10.3389/fimmu.2020.00610>
- 929 Jaenisch, T., Heiss, K., Fischer, N., Geiger, C., Bischoff, F. R., Moldenhauer, G., Rychlewski, L., Sié, A.,
930 Coulibaly, B., Seeberger, P. H., Wyrwicz, L. S., Breitling, F., & Loeffler, F. F. (2019). High-density
931 peptide arrays help to identify linear immunogenic B-cell epitopes in individuals naturally exposed
932 to malaria infection. *Molecular and Cellular Proteomics*, *18*(4), 642–656.
933 <https://doi.org/10.1074/mcp.RA118.000992>
- 934 Joannin, N., Abhiman, S., Sonnhammer, E. L., & Wahlgren, M. (2008). Sub-grouping and sub-
935 functionalization of the RIFIN multi-copy protein family. *BMC Genomics*, *9*.
936 <https://doi.org/10.1186/1471-2164-9-19>
- 937 Kamya, M. R., Arinaitwe, E., Wanzira, H., Katureebe, A., Barusya, C., Kigozi, S. P., Kilama, M., Tatem,
938 A. J., Rosenthal, P. J., Drakeley, C., Lindsay, S. W., Staedke, S. G., Smith, D. L., Greenhouse, B., &
939 Dorsey, G. (2015). Malaria Transmission, Infection, and Disease at Three Sites with Varied
940 Transmission Intensity in Uganda: Implications for Malaria Control. *The American Journal of*

- 941 *Tropical Medicine and Hygiene*, 92(5), 903. <https://doi.org/10.4269/AJTMH.14-0312>
- 942 Kato, Y., Abbott, R. K., Freeman, B. L., Haupt, S., Groschel, B., Silva, M., Menis, S., Irvine, D. J.,
943 Schief, W. R., & Crotty, S. (2020). Multifaceted Effects of Antigen Valency on B Cell Response
944 Composition and Differentiation In Vivo. *Immunity*, 53(3), 548-563.e8.
945 <https://doi.org/10.1016/J.IMMUNI.2020.08.001>
- 946 Kazmin, D., Nakaya, H. I., Lee, E. K., Johnson, M. J., Van Der Most, R., Van Den Berg, R. A., Ballou,
947 W. R., Jongert, E., Wille-Reece, U., Ockenhouse, C., Aderem, A., Zak, D. E., Sadoff, J., Hendriks,
948 J., Wrammert, J., Ahmed, R., & Pulendran, B. (2017). Systems analysis of protective immune
949 responses to RTS,S malaria vaccination in humans. *Proceedings of the National Academy of*
950 *Sciences of the United States of America*, 114(9), 2425–2430.
951 <https://doi.org/10.1073/PNAS.1621489114/-/DCSUPPLEMENTAL>
- 952 Kilama, M., Smith, D. L., Hutchinson, R., Kigozi, R., Yeka, A., Lavoy, G., Kamya, M. R., Staedke, S. G.,
953 Donnelly, M. J., Drakeley, C., Greenhouse, B., Dorsey, G., & Lindsay, S. W. (2014). Estimating the
954 annual entomological inoculation rate for Plasmodium falciparum transmitted by Anopheles
955 gambiae s.l. using three sampling methods in three sites in Uganda. *Malaria Journal*, 13(1), 1–13.
956 <https://doi.org/10.1186/1475-2875-13-111/TABLES/3>
- 957 Larman, H. B., Zhao, Z., Laserson, U., Li, M. Z., Ciccia, A., Gakidis, M. A. M., Church, G. M., Kesari,
958 S., LeProust, E. M., Solimini, N. L., & Elledge, S. J. (2011). Autoantigen discovery with a synthetic
959 human peptidome. *Nature Biotechnology*, 29(6), 535–541. <https://doi.org/10.1038/nbt.1856>
- 960 Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or
961 nucleotide sequences. *Bioinformatics (Oxford, England)*, 22(13), 1658–1659.
962 <https://doi.org/10.1093/BIOINFORMATICS/BTL158>
- 963 Madeira, F., Park, Y., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A., Potter, S.,
964 Finn, R., & Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019.
965 *Nucleic Acids Research*, 47(W1), W636–W641. <https://doi.org/10.1093/NAR/GKZ268>
- 966 Mandel-Brehm, C., Dubey, D., Kryzer, T. J., O'Donovan, B. D., Tran, B., Vazquez, S. E., Sample, H. A.,
967 Zorn, K. C., Khan, L. M., Bledsoe, I. O., McKeon, A., Pleasure, S. J., Lennon, V. A., DeRisi, J. L.,
968 Wilson, M. R., & Pittock, S. J. (2019). Kelch-like Protein 11 Antibodies in Seminoma-Associated
969 Paraneoplastic Encephalitis. <https://doi.org/10.1056/NEJMoa1816721>, 381(1), 47–54.
970 <https://doi.org/10.1056/NEJMoa1816721>
- 971 Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads.
972 *EMBNet.Journal*, 17(1), 10–12. <https://doi.org/10.14806/EJ.17.1.200>
- 973 Mattei, D., Berzins, K., Wahlgren, M., Udomsangpetch, R., Perlmann, P., Griesser, H. W., Scherf, A.,
974 Mjuller-Hill, B., Bonnefoy, S., Guillotte, M., Langsley, G., Silva, L. P. Da, & Mercereau-Puijalon,
975 O. (1989). Cross-reactive antigenic determinants present on different Plasmodium falciparum
976 blood-stage antigens. *Parasite Immunology*, 11(1), 15–29. <https://doi.org/10.1111/J.1365-3024.1989.TB00645.X>
- 977
- 978 Monaco, D. R., Kottapalli, S. V., Breitwieser, F. P., Anderson, D. E., Wijaya, L., Tan, K., Chia, W. N.,
979 Kammers, K., Caturegli, P., Waugh, K., Roederer, M., Petri, M., Goldman, D. W., Rewers, M.,
980 Wang, L.-F., & Larman, H. B. (2021). Deconvoluting virome-wide antibody epitope reactivity
981 profiles. *EBioMedicine*, 75, 103747. <https://doi.org/10.1016/J.EBIOM.2021.103747>
- 982 Morita, M., Takashima, E., Ito, D., Miura, K., Thongkukiatkul, A., Diouf, A., Fairhurst, R. M., Diakite,
983 M., Long, C. A., Torii, M., & Tsuboi, T. (2017). Immunoscreening of Plasmodium falciparum
984 proteins expressed in a wheat germ cell-free system reveals a novel malaria vaccine candidate.
985 *Scientific Reports*, 7(1), 1–8. <https://doi.org/10.1038/srep46086>
- 986 Murugan, R., Scally, S. W., Costa, G., Mustafa, G., Thai, E., Decker, T., Bosch, A., Prieto, K., Levashina,
987 E. A., Julien, J. P., & Wardemann, H. (2020). Evolution of protective human antibodies against
988 Plasmodium falciparum circumsporozoite protein repeat motifs. *Nature Medicine*, 26(7), 1135–
989 1145. <https://doi.org/10.1038/s41591-020-0881-9>
- 990 Nagaoka, H., Kanoi, B. N., Morita, M., Nakata, T., Palacpac, N. M. Q., Egwang, T. G., Horii, T., Tsuboi,
991 T., & Takashima, E. (2021). Characterization of a Plasmodium falciparum PHISTc protein,

- 992 PF3D7_0801000, in blood- stage malaria parasites. *Parasitology International*, 80, 102240.
993 <https://doi.org/10.1016/J.PARINT.2020.102240>
- 994 Niang, M., Bei, A. K., Madnani, K. G., Pelly, S., Dankwa, S., Kanjee, U., Gunalan, K., Amaladoss, A.,
995 Yeo, K. P., Bob, N. S., Malleret, B., Duraisingh, M. T., & Preiser, P. R. (2014). STEVOR is a
996 plasmodium falciparum erythrocyte binding protein that mediates merozoite invasion and rosetting.
997 *Cell Host and Microbe*, 16(1), 81–93. <https://doi.org/10.1016/j.chom.2014.06.004>
- 998 O'Connor, B. P., Vogel, L. A., Zhang, W., Loo, W., Shnider, D., Lind, E. F., Ratliff, M., Noelle, R. J., &
999 Erickson, L. D. (2006). Imprinting the Fate of Antigen-Reactive B Cells through the Affinity of the
1000 B Cell Receptor. *Journal of Immunology (Baltimore, Md. : 1950)*, 177(11), 7723.
1001 <https://doi.org/10.4049/JIMMUNOL.177.11.7723>
- 1002 O'Donovan, B., Mandel-Brehm, C., Vazquez, S. E., Liu, J., Parent, A. V., Anderson, M. S., Kassimatis,
1003 T., Zekeridou, A., Hauser, S. L., Pittock, S. J., Chow, E., Wilson, M. R., & DeRisi, J. L. (2020).
1004 High-resolution epitope mapping of anti-Hu and anti-Yo autoimmunity by programmable phage
1005 display. *Brain Communications*, 2(2). <https://doi.org/10.1093/BRAINCOMMS/FCAA059>
- 1006 O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse,
1007 B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V.,
1008 Chetvermin, V., Choi, J., Cox, E., Ermolaeva, O., ... Pruitt, K. D. (2016). Reference sequence
1009 (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic
1010 Acids Research*, 44(D1), D733–D745. <https://doi.org/10.1093/NAR/GKV1189>
- 1011 Ochiai, K., Maienschein-Cline, M., Simonetti, G., Chen, J., Rosenthal, R., Brink, R., Chong, A. S., Klein,
1012 U., Dinner, A. R., Singh, H., & Sciammas, R. (2013). Transcriptional Regulation of Germinal
1013 Center B and Plasma Cell Fates by Dynamical Control of IRF4. *Immunity*, 38(5), 918–929.
1014 <https://doi.org/10.1016/J.IMMUNI.2013.04.009>
- 1015 Olotu, A., Fegan, G., Wambua, J., Nyangweso, G., Leach, A., Lievens, M., Kaslow, D. C., Njuguna, P.,
1016 Marsh, K., & Bejon, P. (2016). Seven-Year Efficacy of RTS,S/AS01 Malaria Vaccine among
1017 Young African Children. *The New England Journal of Medicine*, 374(26), 2519.
1018 <https://doi.org/10.1056/NEJM0A1515257>
- 1019 Pape, K. A., Maul, R. W., Dileepan, T., Paustian, A. S., Gearhart, P. J., & Jenkins, M. K. (2018). Naïve B
1020 cells with high-avidity germline-encoded antigen receptors produce persistent IgM+ and transient
1021 IgG+ memory B cells. *Immunity*, 48(6), 1135. <https://doi.org/10.1016/J.IMMUNI.2018.04.019>
- 1022 Paus, D., Tri, G. P., Chan, T. D., Gardam, S., Basten, A., & Brink, R. (2006). Antigen recognition
1023 strength regulates the choice between extrafollicular plasma cell and germinal center B cell
1024 differentiation. *Journal of Experimental Medicine*, 203(4), 1081–1091.
1025 <https://doi.org/10.1084/jem.20060087>
- 1026 Portugal, S., Pierce, S. K., & Crompton, P. D. (2013). Young Lives Lost as B Cells Falter: What We Are
1027 Learning About Antibody Responses in Malaria. *The Journal of Immunology*, 190(7), 3039–3046.
1028 <https://doi.org/10.4049/JIMMUNOL.1203067>
- 1029 Rajan, J. V., McCracken, M., Mandel-Brehm, C., Gromowski, G., Pollett, S., Jarman, R., & DeRisi, J. L.
1030 (2021). Phage display demonstrates durable differences in serological profile by route of inoculation
1031 in primary infections of non-human primates with Dengue Virus 1. *Scientific Reports 2021 11:1*,
1032 11(1), 1–12. <https://doi.org/10.1038/s41598-021-90318-z>
- 1033 Rask, T. S., Hansen, D. A., Theander, T. G., Pedersen, A. G., & Lavstsen, T. (2010). Plasmodium
1034 falciparum Erythrocyte Membrane Protein 1 Diversity in Seven Genomes – Divide and Conquer.
1035 *PLOS Computational Biology*, 6(9), e1000933. <https://doi.org/10.1371/JOURNAL.PCBI.1000933>
- 1036 Reeder, J., & Brown, G. (1996). Antigenic variation and immune evasion in Plasmodium falciparum
1037 malaria. *Immunology and Cell Biology*, 74(6), 546–554. <https://doi.org/10.1038/ICB.1996.88>
- 1038 Rek, J., Katrak, S., Obasi, H., Nayebare, P., Katureebe, A., Kakande, E., Arinaitwe, E., Nankabirwa, J. I.,
1039 Jagannathan, P., Drakeley, C., Staedke, S. G., Smith, D. L., Bousema, T., Kamya, M., Rosenthal, P.
1040 J., Dorsey, G., & Greenhouse, B. (2016). Characterizing microscopic and submicroscopic malaria
1041 parasitaemia at three sites with varied transmission intensity in Uganda. *Malaria Journal*, 15(1).
1042 <https://doi.org/10.1186/S12936-016-1519-8>

- 1043 Rodriguez-Barraquer, I., Arinaitwe, E., Jagannathan, P., Kanya, M. R., Rosenthal, P. J., Rek, J., Dorsey,
1044 G., Nankabirwa, J., Staedke, S. G., Kilama, M., Drakeley, C., Ssewanyana, I., Smith, D. L., &
1045 Greenhouse, B. (2018). Quantification of anti-parasite and anti-disease immunity to malaria as a
1046 function of age and exposure. *ELife*, 7. <https://doi.org/10.7554/eLife.35832>
1047 Rubinstein, N. D., Mayrose, I., Halperin, D., Yekutieli, D., Gershoni, J. M., & Pupko, T. (2008).
1048 Computational characterization of B-cell epitopes. *Molecular Immunology*, 45(12), 3477–3489.
1049 <https://doi.org/10.1016/j.molimm.2007.10.016>
1050 Saito, F., Hirayasu, K., Satoh, T., Wang, C. W., Lusingu, J., Arimori, T., Shida, K., Palacpac, N. M. Q.,
1051 Itagaki, S., Iwanaga, S., Takashima, E., Tsuboi, T., Kohyama, M., Suenaga, T., Colonna, M.,
1052 Takagi, J., Lavstsen, T., Horii, T., & Arase, H. (2017). Immune evasion of Plasmodium falciparum
1053 by RIFIN via inhibitory receptors. *Nature* 2017 552:7683, 552(7683), 101–105.
1054 <https://doi.org/10.1038/nature24994>
1055 Schofield, L. (1991). On the function of repetitive domains in protein antigens of Plasmodium and other
1056 eukaryotic parasites. In *Parasitology Today* (Vol. 7, Issue 5, pp. 99–105). Elsevier Current Trends.
1057 [https://doi.org/10.1016/0169-4758\(91\)90166-L](https://doi.org/10.1016/0169-4758(91)90166-L)
1058 Schofield, L., & Uadia, P. (1990). Lack of Ir gene control in the immune response to malaria. I. A
1059 thymus-independent antibody response to the repetitive surface protein of sporozoites. *The Journal*
1060 *of Immunology*, 144(7).
1061 Schwickert, T. A., Victora, G. D., Fooksman, D. R., Kamphorst, A. O., Mugnier, M. R., Gitlin, A. D.,
1062 Dustin, M. L., & Nussenzweig, M. C. (2011). A dynamic T cell–limited checkpoint regulates
1063 affinity-dependent B cell entry into the germinal center. *The Journal of Experimental Medicine*,
1064 208(6), 1243. <https://doi.org/10.1084/JEM.20102477>
1065 Shannon, P., Markiel, A., Ozier, O., Baliga, N., Wang, J., Ramage, D., Amin, N., Schwikowski, B., &
1066 Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular
1067 interaction networks. *Genome Research*, 13(11), 2498–2504. <https://doi.org/10.1101/GR.1239303>
1068 Smith, G. P., & Petrenko, V. A. (1997). Phage display. *Chemical Reviews*, 97(2), 391–410.
1069 <https://doi.org/10.1021/CR960065D/ASSET/IMAGES/LARGE/CR960065DF00004.JPEG>
1070 Tan, J., Pieper, K., Piccoli, L., Abdi, A., Foglierini, M., Geiger, R., Maria Tully, C., Jarrossay, D., Maina
1071 Ndungu, F., Wambua, J., Bejon, P., Silacci Fregni, C., Fernandez-Rodriguez, B., Barbieri, S.,
1072 Bianchi, S., Marsh, K., Thathy, V., Corti, D., Sallusto, F., ... Lanzavecchia, A. (2015). A LAIR1
1073 insertion generates broadly reactive antibodies against malaria variant antigens. *Nature* 2015
1074 529:7584, 529(7584), 105–109. <https://doi.org/10.1038/nature16450>
1075 Taylor, J. J., Pape, K. A., Steach, H. R., & Jenkins, M. K. (2015). Apoptosis and antigen affinity limits
1076 effector cell differentiation of a single naïve B cell. *Science (New York, N.Y.)*, 347(6223), 784.
1077 <https://doi.org/10.1126/SCIENCE.AAA1342>
1078 Vatti, A., Monsalve, D. M., Pacheco, Y., Chang, C., Anaya, J. M., & Gershwin, M. E. (2017). Original
1079 antigenic sin: A comprehensive review. *Journal of Autoimmunity*, 83, 12–21.
1080 <https://doi.org/10.1016/J.JAUT.2017.04.008>
1081 Vazquez, S. E., Ferré, E. M. N., Scheel, D. W., Sunshine, S., Miao, B., Mandel-Brehm, C., Quandt, Z.,
1082 Chan, A. Y., Cheng, M., German, M., Lionakis, M., Derisi, J. L., & Anderson, M. S. (2020).
1083 Identification of novel, clinically correlated autoantigens in the monogenic autoimmune syndrome
1084 APS1 by proteome-wide phip-seq. *ELife*, 9. <https://doi.org/10.7554/ELIFE.55053>
1085 Wåhlin, B., Sjölander, A., Ahlborg, N., Udomsangpetch, R., Scherf, A., Mattei, D., Berzins, K., &
1086 Perlmann, P. (1992). Involvement of Pf155/RESA and cross-reactive antigens in Plasmodium
1087 falciparum merozoite invasion in vitro. *Infection and Immunity*, 60(2), 443–449.
1088 <https://doi.org/10.1128/IAI.60.2.443-449.1992>
1089 Wang, S., Mata-Fink, J., Kriegsman, B., Hanson, M., Irvine, D. J., Eisen, H. N., Burton, D. R., Wittrup,
1090 K. D., Kardar, M., & Chakraborty, A. K. (2015). Manipulating the Selection Forces during Affinity
1091 Maturation to Generate Cross-Reactive HIV Antibodies. *Cell*, 160(4), 785–797.
1092 <https://doi.org/10.1016/J.CELL.2015.01.027>
1093 Wu, T. D., & Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short

- 1094 reads. *Bioinformatics*, 26(7), 873–881. <https://doi.org/10.1093/BIOINFORMATICS/BTQ057>
- 1095 Xie, Y., Li, X., Chai, Y., Song, H., Qi, J., & Gao, G. F. (2021). Structural basis of malarial parasite
- 1096 RIFIN-mediated immune escape against LAIR1. *Cell Reports*, 36(8), 109600.
- 1097 <https://doi.org/10.1016/J.CELREP.2021.109600>
- 1098 Yeka, A., Nankabirwa, J., Mpimbaza, A., Kigozi, R., Arinaitwe, E., Drakeley, C., Greenhouse, B.,
- 1099 Kanya, M. R., Dorsey, G., & Staedke, S. G. (2015). Factors Associated with Malaria Parasitemia,
- 1100 Anemia and Serological Responses in a Spectrum of Epidemiological Settings in Uganda. *PLOS*
- 1101 *ONE*, 10(3), e0118901. <https://doi.org/10.1371/JOURNAL.PONE.0118901>
- 1102 Zamecnik, C. R., Rajan, J. V., Yamauchi, K. A., Mann, S. A., Loudermilk, R. P., Sowa, G. M., Zorn, K.
- 1103 C., Alvarenga, B. D., Gaebler, C., Caskey, M., Stone, M., Norris, P. J., Gu, W., Chiu, C. Y., Ng, D.,
- 1104 Byrnes, J. R., Zhou, X. X., Wells, J. A., Robbiani, D. F., ... Wilson, M. R. (2020). ReScan, a
- 1105 Multiplex Diagnostic Pipeline, Pans Human Sera for SARS-CoV-2 Antigens. *Cell Reports*
- 1106 *Medicine*, 1(7), 100123. <https://doi.org/10.1016/j.xcrm.2020.100123>
- 1107

Table 1 – Characteristics of the Ugandan Cohorts

Region	Age group (yrs)	No. of people	Proportion positive for infection at the time of sample collection	Time since last infection (days) - median (IQR)	Incidence of symptomatic malaria per year - median (IQR)	Household annual EIR* (infective bites / person) - median (IQR)
Tororo	2 to 3	10	0.5	18.5 (0,85)	5.8 (2.9,7.7)	56 (33,148)
	4 to 6	30	0.66	0 (0,45)	3.6 (2.6,4.8)	59 (38,84)
	7 to 11	30	0.63	0 (0,45)	2.3 (2,4.3)	46 (30,110)
	> 18	30	0.7	0 (0,45)	1.2 (0.9,1.6)	49 (35,94)
Kanungu	2 to 3	10	0.1	155 (61,190)	1.7 (0.9,2)	4.3 (4, 14)
	4 to 6	30	0.2	114 (43,289)	1.5 (0.7, 2.3)	7.3 (4.5, 15)
	7 to 11	30	0.13	121 (41,263)	1.5 (0.6, 2)	5.2 (4, 7)
	> 18	30	0.2	109 (38, 223)	1.1 (0.8, 1.3)	6.8 (4.8, 15.4)

*EIR – Entomological Inoculation Rate

Table 2 – Composition of Falciparome phage library

	Input sequences before collapsing on similarity	Identity threshold for collapsing by CD-HIT	# Final collapsed Protein sequences
<i>P. falciparum</i> reference proteome	3D7, IT (10,771 total)	99%	6372
<i>P. falciparum</i> variant sequences	<ul style="list-style-type: none"> • PfEMP1 (431 from 3D7, IT, IGH, RAJ116, PFCLIN, IT4, DD2 genomes) • RIFIN (all 3D7 + IT) • STEVOR (all 3D7 + IT) • SURFIN (all 3D7 + IT+15) • AMA1 (2) • CSP (6) • MSPDBL1 (6) • MSPDBL2 (5) • PfMC2TM (all 3D7 + IT) 	100% (90% for CSP)	1205
Other variants	<i>P. reichenowi</i> PfEMP1 (PFREICH) <i>Anopheles</i> - CE5 (5), SG6 (5)		
<i>Anopheles</i> salivary proteins	53 proteins from 19 <i>Anopheles</i> species as described in Fig 1 of (Arcà et al. 2017)	98%	708
Vaccine/Viral/Toxin sequences	<ul style="list-style-type: none"> • Tetanus • Diphtheria • Pertussis • EBV • Measles • Mumps • Rubella • Polio • RotoAB 	98% (90% for RotoAB)	684
Laboratory positive controls	<ul style="list-style-type: none"> • GFAP • GFP • Gephyrin • MYC, NR1 • Tubulin (alpha/beta) 	98%	11
TOTAL PROTEINS			8,980
TOTAL PEPTIDES			238,068

References:

Arcà, Bruno, Fabrizio Lombardo, Claudio J. Struchiner, and José M.C. Ribeiro. 2017. "Anopheline Salivary Protein Genes and Gene Families: An Evolutionary Overview after the Whole Genome Sequence of Sixteen *Anopheles* Species." *BMC Genomics* 18 (1).

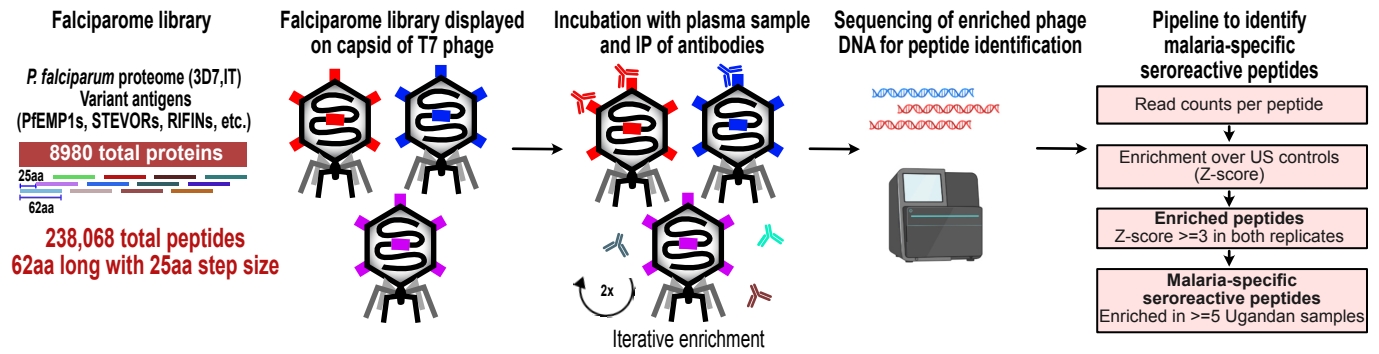


Fig 1 - PhIP-seq overview and analysis pipeline

Falciparome phage library displays the proteome of *Plasmodium falciparum* in 62-aa peptides with 25-aa step size on T7 phage and also includes variant sequences of many antigens, including major Variant Surface Antigens (VSA). PhIP-seq was performed with incubation of Falciparome library with human plasma, followed by IP of antibodies in the sample and enrichment of antibody binding phage. Two rounds of enrichment were performed and enriched phage were sequenced to obtain the identity of the

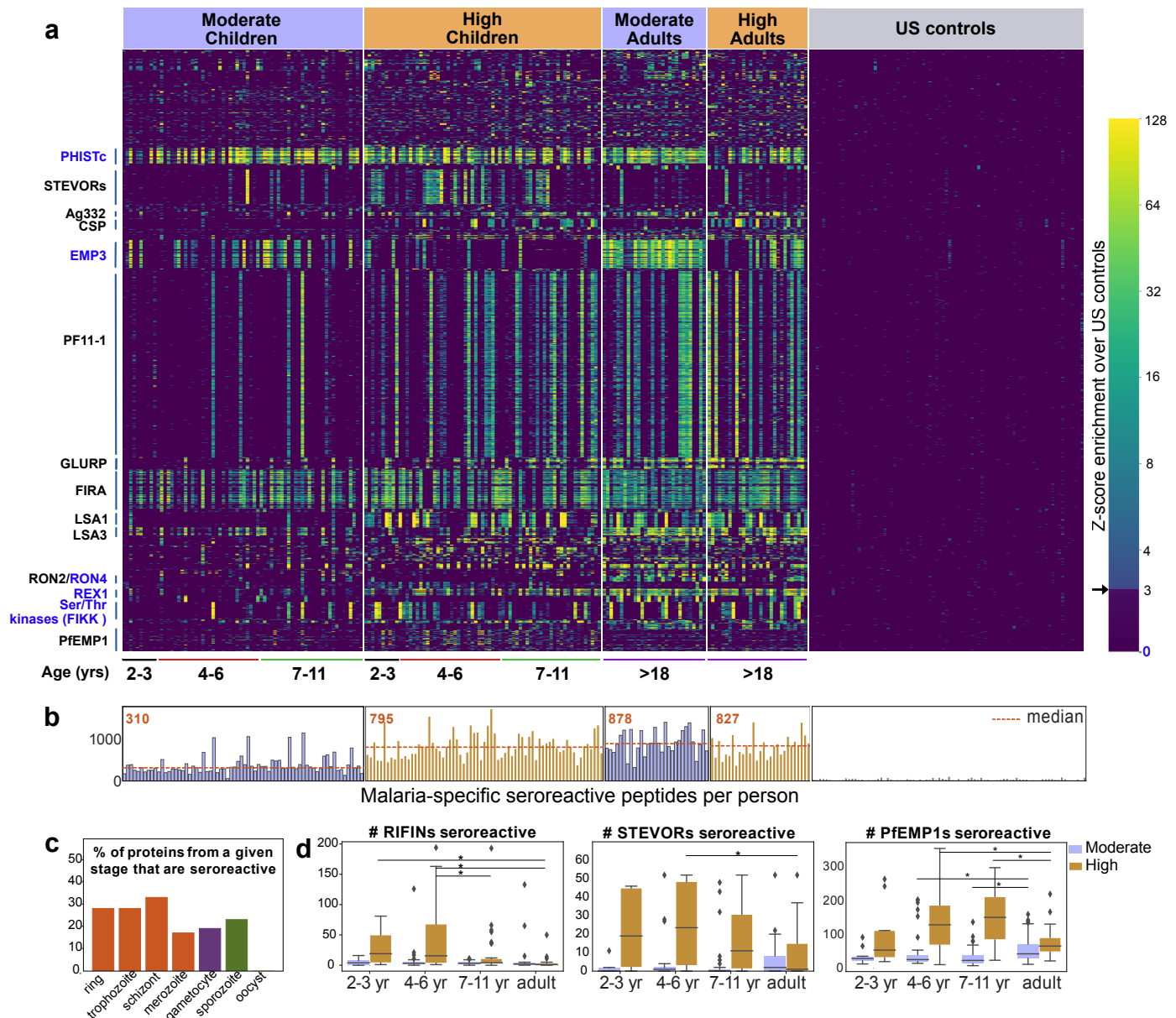


Fig2: PhIP-seq with Falciparome captures known, novel antigens and relationships between age, exposure and breadth of seroreactive regions

(a) Heatmap of Z-score enrichment over US controls for seroreactive peptides (rows) with >10% seropositivity across different age groups in the moderate and high exposure cohorts. Peptides are sorted by protein name and samples (columns) are ordered by increasing age in each group. Well-characterized (black labels) as well as under-characterized/novel (blue labels) antigens in *Plasmodium falciparum* were identified with this approach.

(b) Breadth of antibody reactivity, shown as number of seroreactive peptides in each person. Dotted red line and red text indicate median breadth for each population group. Children from the moderate transmission setting had significantly lower breadth than children from the high transmission setting as well as all adults (KS test p-value < 0.05).

(c) Percentage of proteins across different stages identified as seroreactive in this study. Stage classification is based on proteomic datasets in PlasmoDB (spectral count ≥ 2 for at least 1 peptide in a protein in a given stage is counted as expression) and shows enrichment of proteins from all life stages of *Plasmodium falciparum* in the human host.

(d) Breadth of VSA reactivity, shown as number of variant proteins of RIFINs, STEVORs and PfEMP1s seroreactive per person. In the moderate transmission setting, children had a significantly lower breadth than adults for PfEMP1 and both age groups poorly recognized RIFINs and STEVORs. In contrast, in the high transmission setting, children had a significantly (* KS test < 0.05) higher breadth than adults for all three families.

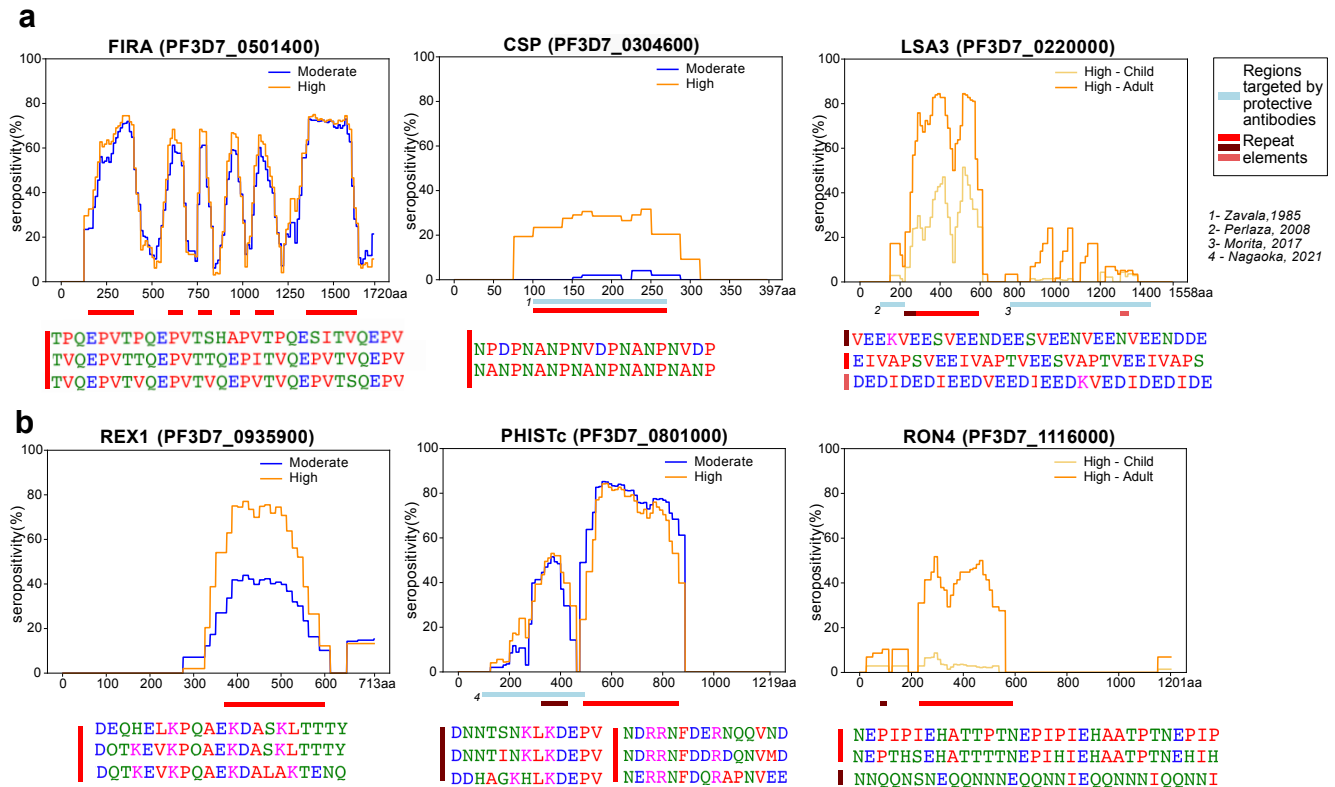


Fig3: Tiled design of library facilitates high resolution characterization of seroreactive proteins

(a) Examples of previously well-characterized antigens and (b) novel/previously under-characterized antigens identified in this dataset. Average percent of people seropositive at each residue (seropositivity) based on signal from peptides spanning it are shown for each protein for different groups in the cohort. The magnitude of exposure- and age-related differences in seropositivity varies by individual protein and even within different regions of specific proteins. Reddish bars underneath each protein represent repeat elements and blue bars represent examples of regions encompassing targets of protective antibodies described in previous studies. Snapshots of sequences of repeat elements present in a protein are represented beneath the protein.

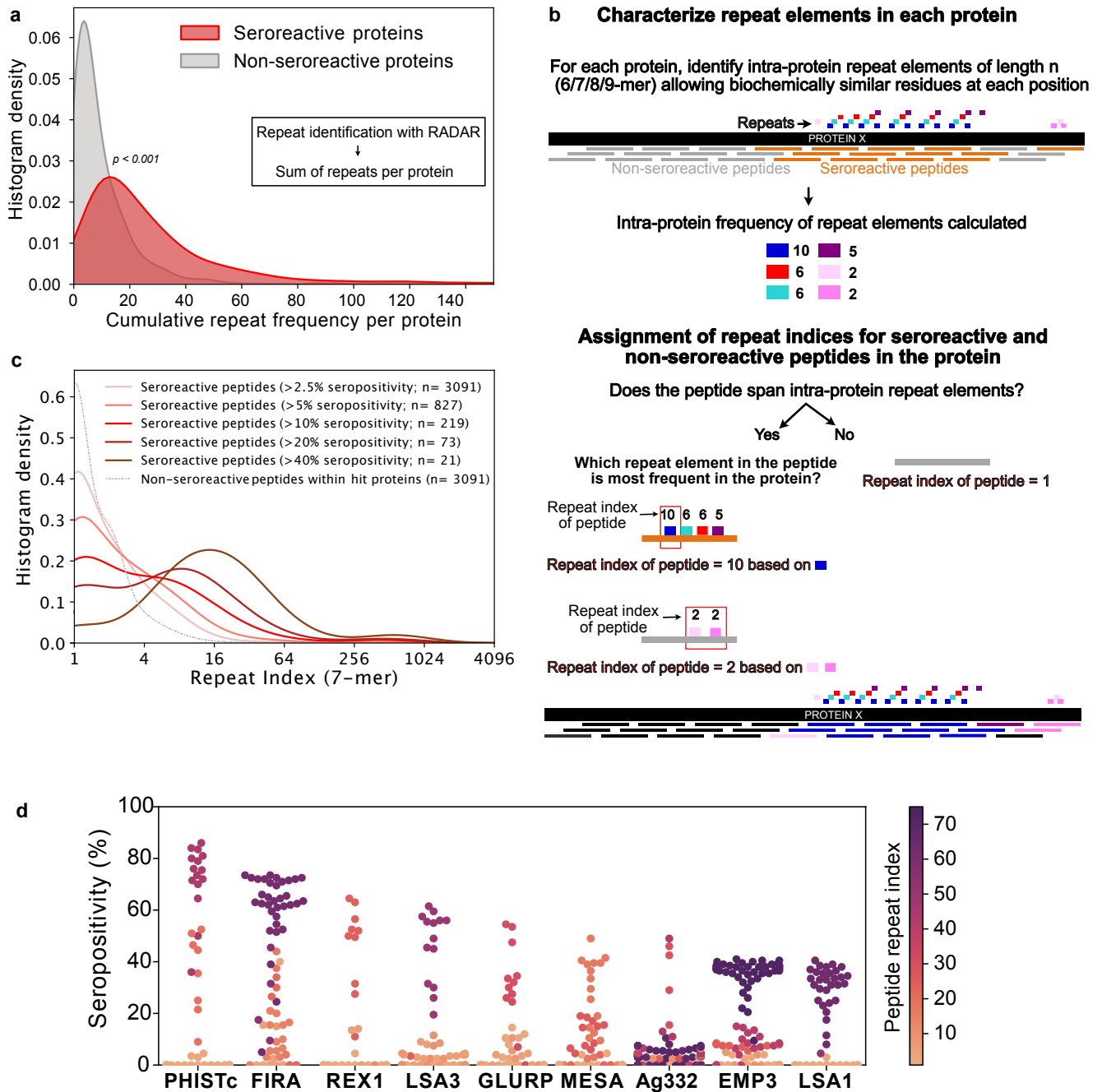


Fig4 - Repeat elements are more enriched in seroreactive peptides within seroreactive proteins than non-seroreactive peptides

(a) Distribution of cumulative frequency of repeat elements per protein is significantly higher (KS test p -value < 0.05) in the seroreactive protein set than a randomly sampled subset of non-seroreactive proteins (1000 iterations).
 (b) Pipeline to compute the representation of repeats in each peptide as repeat index.
 (c) Distribution of repeat indices is significantly higher (KS test p -value < 0.05) in seroreactive peptides than a randomly sampled subset of non-seroreactive peptides within seroreactive proteins (1000 iterations). Distribution of repeat indices also significantly increases with increase in seropositivity (KS test p -value < 0.05 between all successive distributions).
 (d) Seropositivity of all peptides (dots) colored by their repeat indices in the top 9 most seropositive repeat-containing proteins shows enrichment of repeat elements in peptides with high seropositivity.

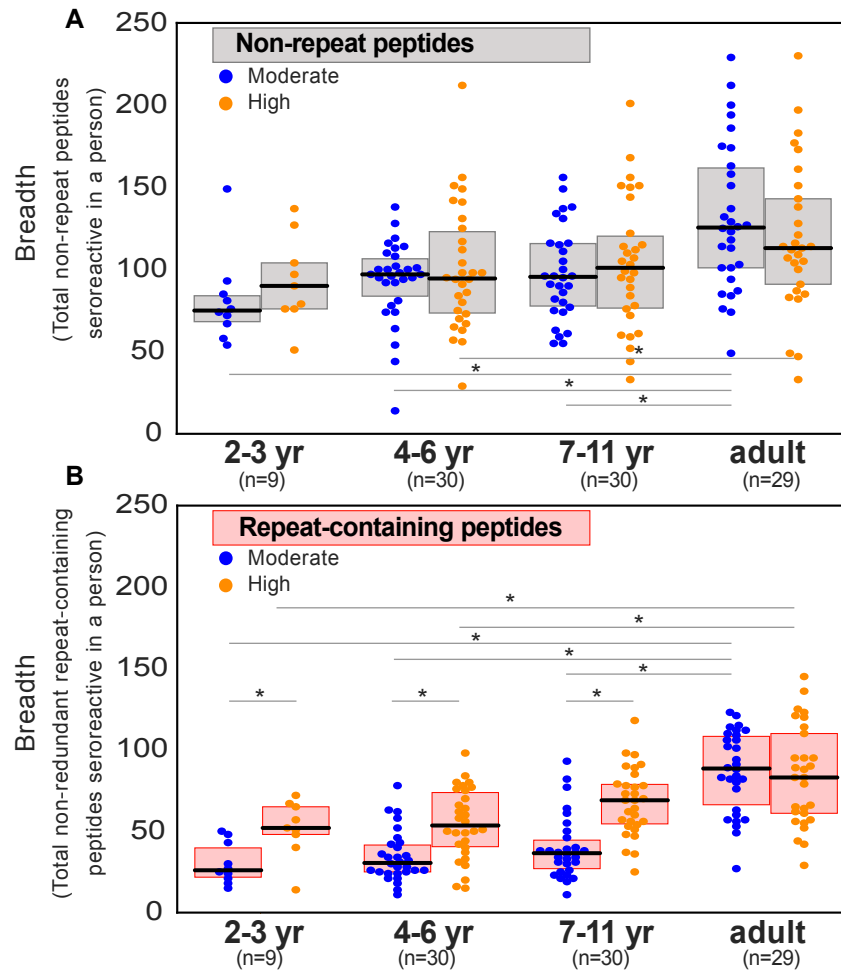


Fig 5: Exposure intensity is positively associated with breadth of seroreactive repeat-containing peptides, but not non-repeat peptides in children

(a) Breadth of seroreactive non-repeat peptides per person is not significantly different between the two exposure settings within each age group.

(b) Breadth of seroreactive repeat-containing peptides per person is significantly higher (KS-test p-value < 0.05) in the high exposure setting than in the moderate exposure setting within the three groups in children, but not adults.

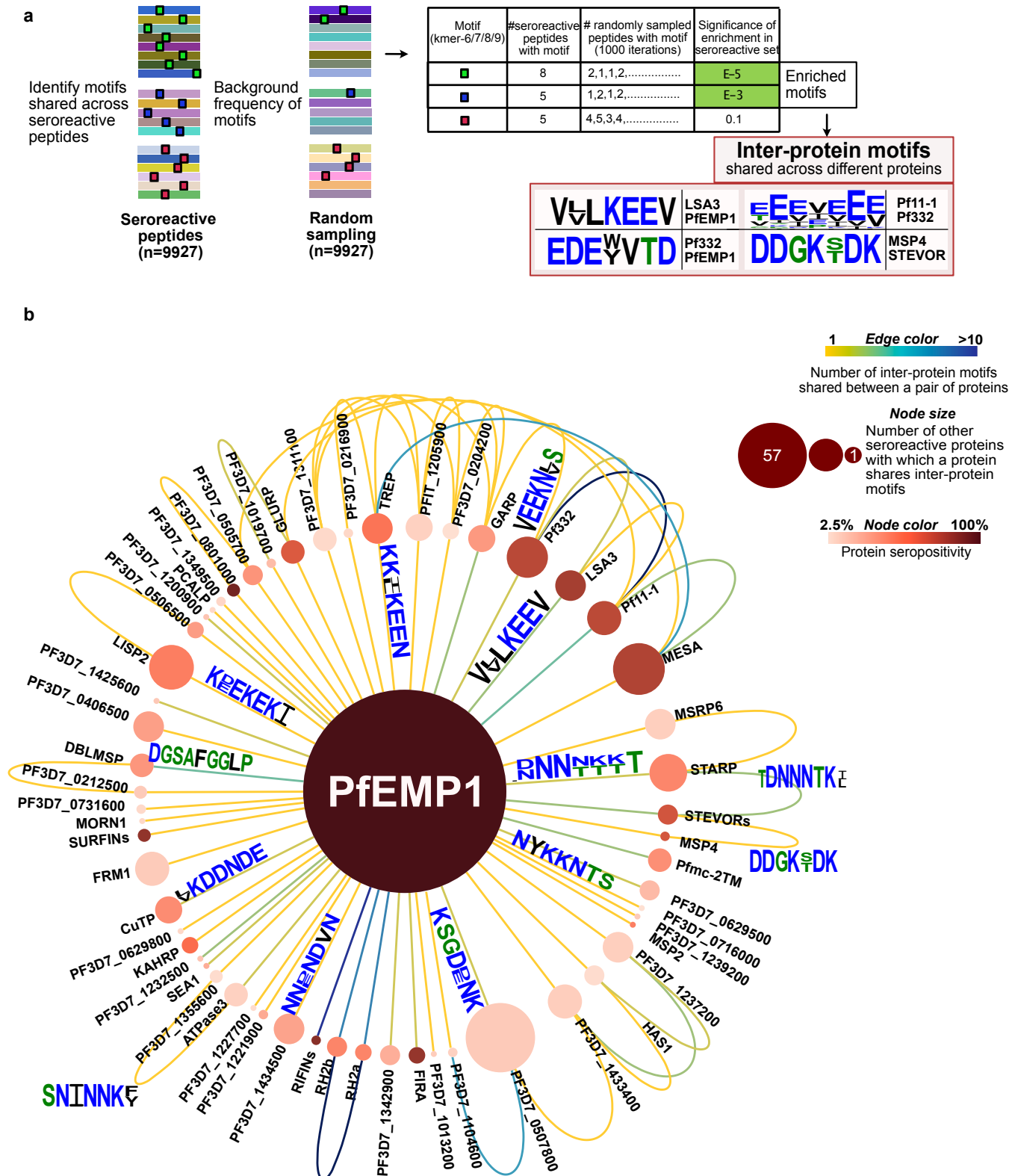
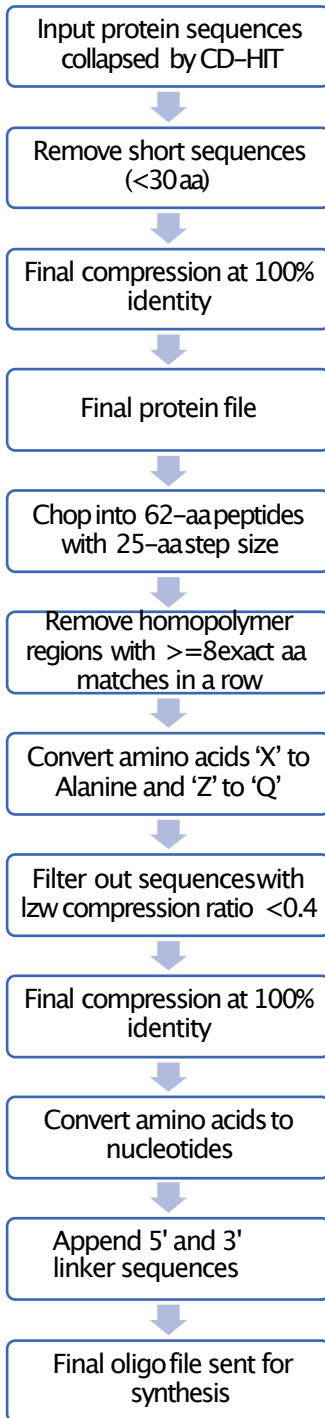


Fig6: Extensive sharing of motifs observed among seroreactive proteins, with the most shared with PfEMP1 family
 (a) Pipeline to identify inter-protein motifs (6-9aa) significantly enriched (FDR < 0.001) in seroreactive peptides from different seroreactive proteins (different colors) over background. Background for each motif was estimated based on the number of random peptides possessing the motif in 1000 random samplings of 9927 peptides. Examples of inter-protein motifs and seroreactive proteins sharing them are also shown.
 (b) Network of PfEMP1 sharing inter-protein motifs with other seroreactive proteins based on 7-aa motifs with up to 2 conservative substitutions. PfEMP1 shared inter-protein motifs with the greatest number of other seroreactive proteins.

SUPPLEMENTAL FIGURES

PEPTIDE PROCESSING STEPS



NT SEQUENCE VERIFICATION

Do both files (AA and NT) have the same number of sequences?

Are there linker sequences at the 5' and 3' ends of all sequences?

Are all sequences the same length?

Are there any long runs of (> 8 NT) of single nucleotide?

Do the translated sequences match those in the original peptide file?

Are all restriction sites (HINDIII, EcoR1) removed?

Fig S1a - Pipeline for library construction: Input sequences of different groups were filtered with CD-HIT to remove similar sequences with more than the indicated % identity in Table 2. The filtered sequences were then processed into peptides using the peptide processing pipeline and quality checks were performed as described in NT sequence verification.

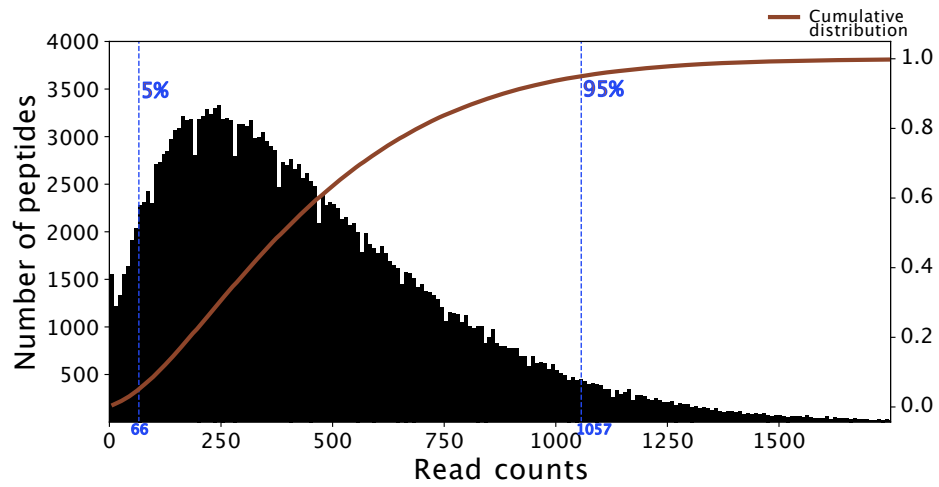


Fig S2a – Histogram of read counts of input library.

Read counts corresponding to the 5th and 95th percentile in the distribution (indicated in blue) are within a 15-fold difference. Cumulative density plot of the distribution is shown in red.

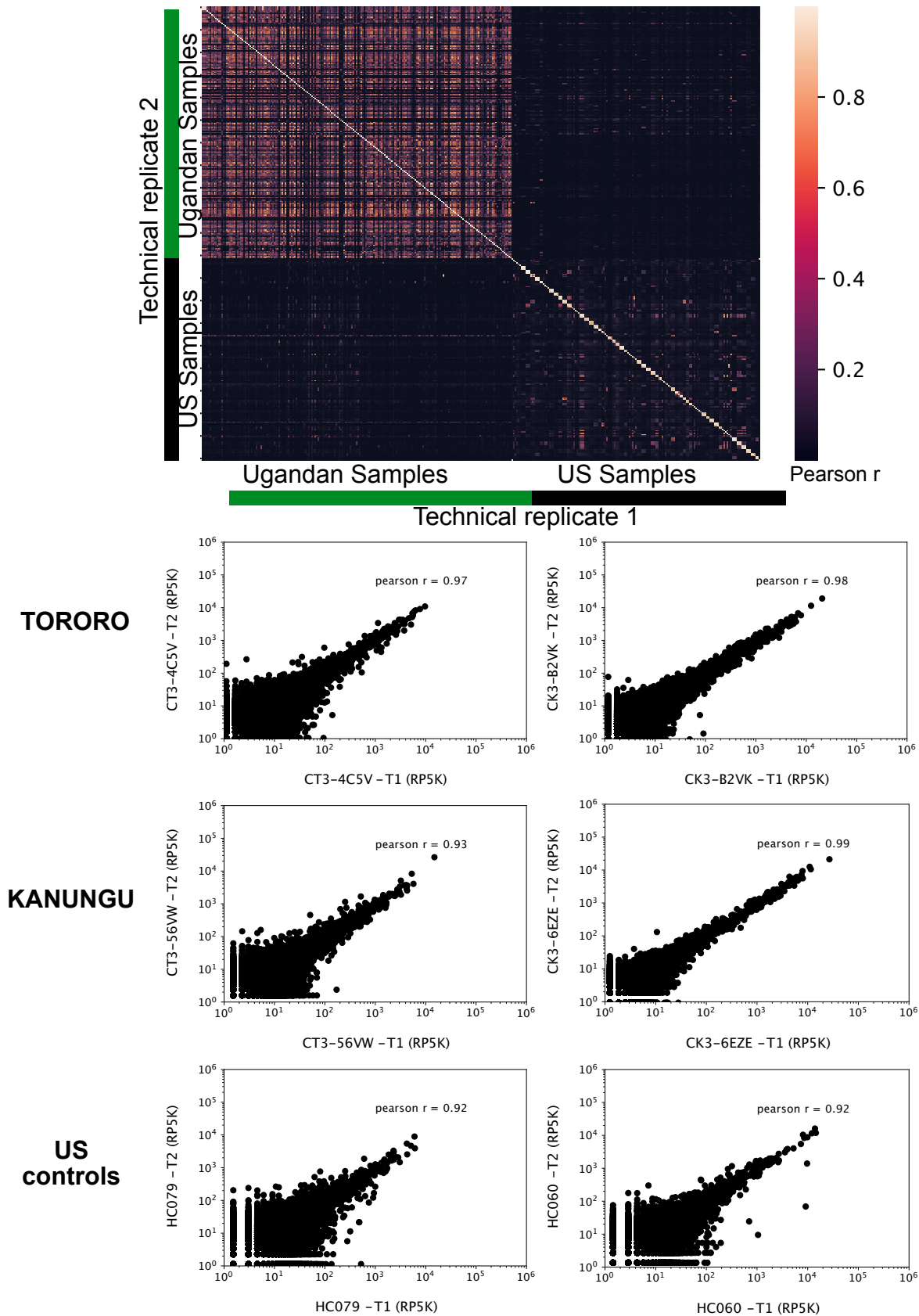


Fig. S2b - Technical replicates are well correlated

Top - Pearson correlation matrix of depth-adjusted read counts across all samples. Technical replicates are placed symmetrically on rows and columns.

Bottom three - Representative scatter plots of reads per 500,000 (RP5K) of technical replicates of samples from Tororo, Kanungu and US.

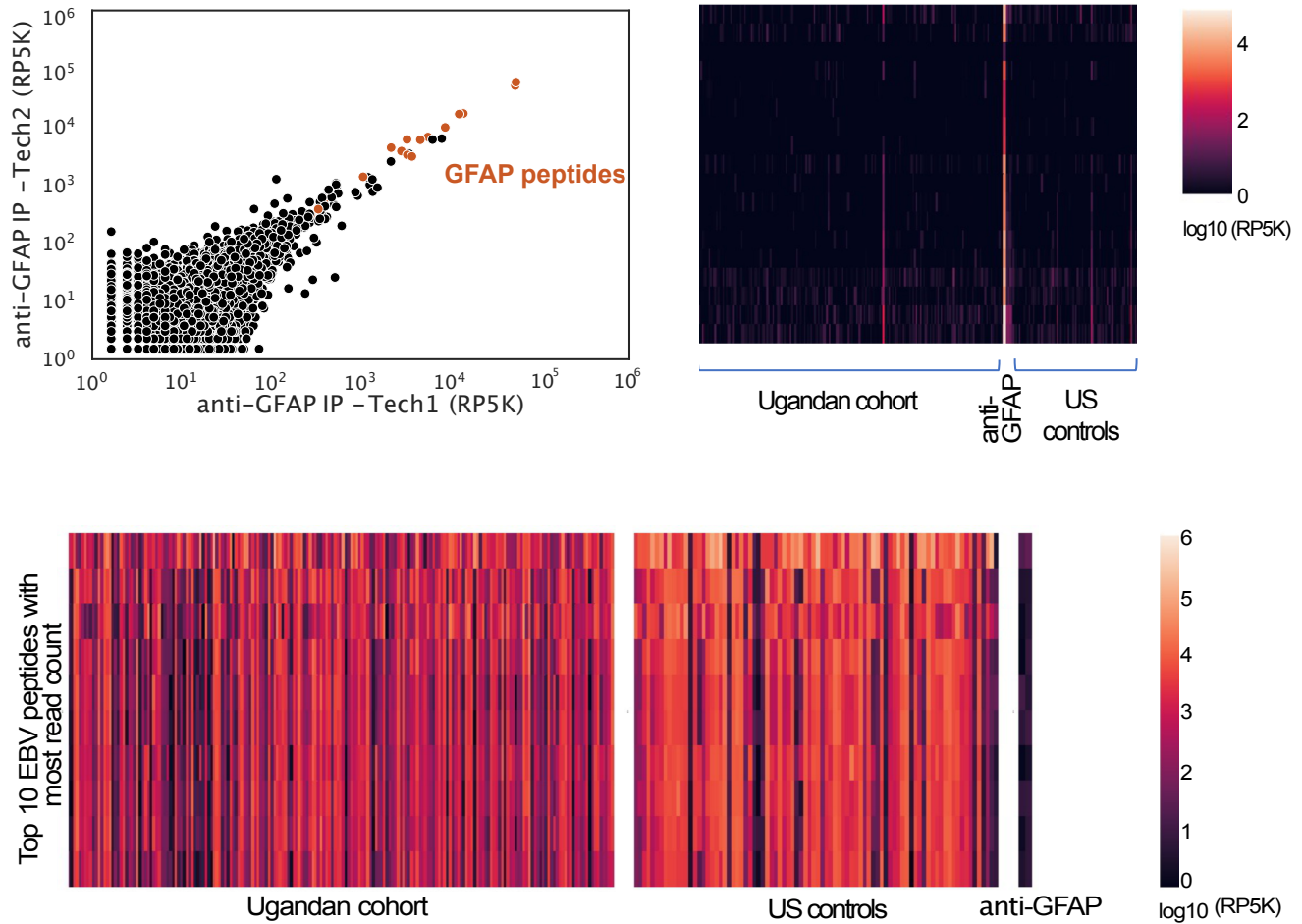
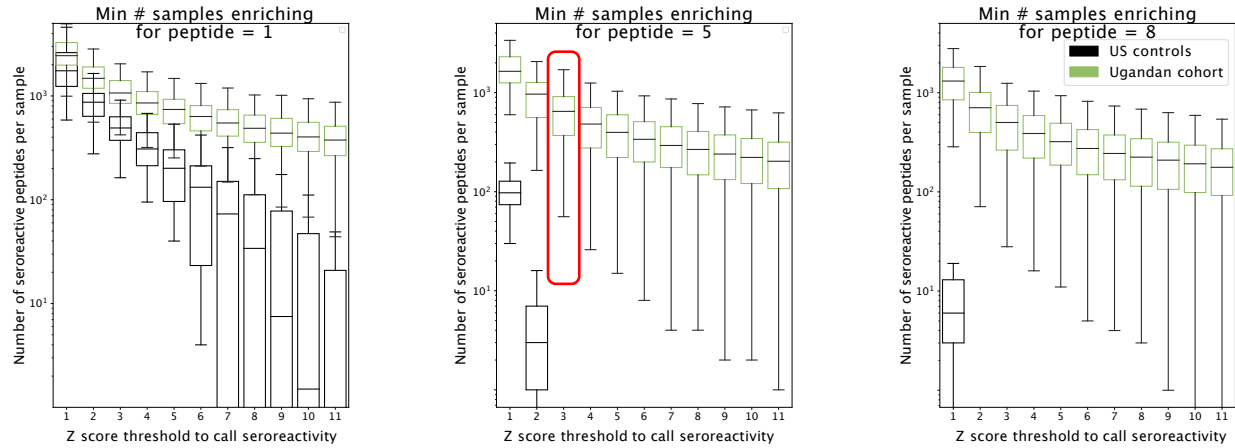


Fig. S2c

Top panel - PhIP-seq with polyclonal anti-GFAP enriches for GFAP peptides and enrichment is specific to IP with anti-GFAP, but is observed rarely in the Ugandan cohort and US controls.

Left - Scatter plot of Reads Per 500,000 (RP5K) of technical replicates of an IP with anti-GFAP. GFAP peptides are in red. Right - Heat map of RP5K of GFAP peptides (rows) in different samples (columns).

Bottom panel - Heat map of RP5K of top 10 Epstein-Barr virus (EBV) peptides (rows) with highest read counts in human samples. Enrichment is observed across Ugandan and US samples, but not in the IP with anti-GFAP.

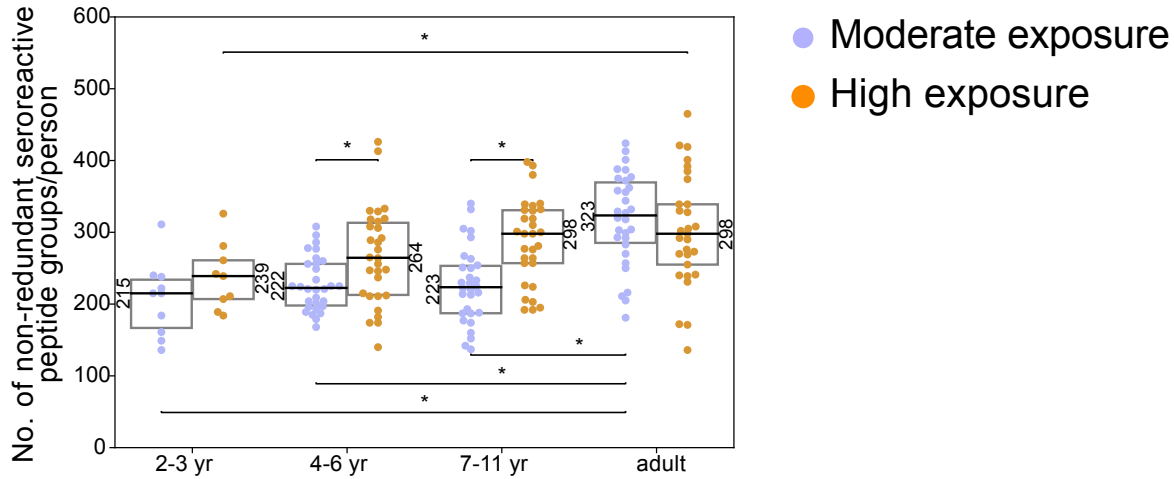


FigS2d - Moving threshold analysis to determine optimal thresholds for calling peptides as seroreactive based on minimum Z-score and enrichment in a minimum number of samples in a group. Box plots of resultant number of seroreactive peptides for corresponding thresholds are shown for Ugandan samples and US controls. The final thresholds for calling seroreactivity were selected based on minimizing the number of peptides identified as seroreactive in the US controls and is indicated by the red box.

GO cellular component complete	#	#	expected	Fold Enrichment	+/-	raw P value	FDR
anchored component of plasma membrane	25	6	.48	12.41	+	2.03E-05	8.19E-04
↳intrinsic component of plasma membrane	50	6	.97	6.20	+	6.08E-04	1.73E-02
↳plasma membrane	122	12	2.36	5.08	+	7.08E-06	3.88E-04
↳membrane	1489	70	28.81	2.43	+	1.19E-16	1.82E-14
↳cellular anatomical entity	3445	88	66.64	1.32	+	6.23E-06	3.68E-04
↳cell periphery	128	12	2.48	4.85	+	1.11E-05	5.70E-04
↳intrinsic component of membrane	1253	62	24.24	2.56	+	5.38E-15	4.13E-13
↳anchored component of membrane	28	6	.54	11.08	+	3.53E-05	1.29E-03
cell surface	34	8	.66	12.16	+	8.88E-07	5.68E-05
Maurer's cleft	90	20	1.74	11.49	+	6.23E-15	4.35E-13
↳host cell cytoplasm part	204	29	3.95	7.35	+	8.64E-17	1.66E-14
↳host cell cytoplasm	218	32	4.22	7.59	+	6.22E-19	4.78E-16
↳host intracellular part	218	32	4.22	7.59	+	6.22E-19	2.39E-16
↳host intracellular region	219	32	4.24	7.55	+	7.04E-19	1.80E-16
↳host cell part	361	35	6.98	5.01	+	1.57E-15	2.02E-13
↳host cellular component	365	35	7.06	4.96	+	2.16E-15	2.37E-13
↳host cell	365	35	7.06	4.96	+	2.16E-15	2.07E-13
symbiont-containing vacuole	54	7	1.04	6.70	+	1.36E-04	4.18E-03
↳extracellular membrane-bounded organelle	54	7	1.04	6.70	+	1.36E-04	4.36E-03
↳extracellular organelle	55	7	1.06	6.58	+	1.51E-04	4.46E-03
↳extracellular region	77	9	1.49	6.04	+	3.08E-05	1.18E-03
apical complex	58	6	1.12	5.35	+	1.24E-03	3.18E-02
↳apical part of cell	60	6	1.16	5.17	+	1.46E-03	3.62E-02
integral component of membrane	1236	62	23.91	2.59	+	2.79E-15	2.38E-13

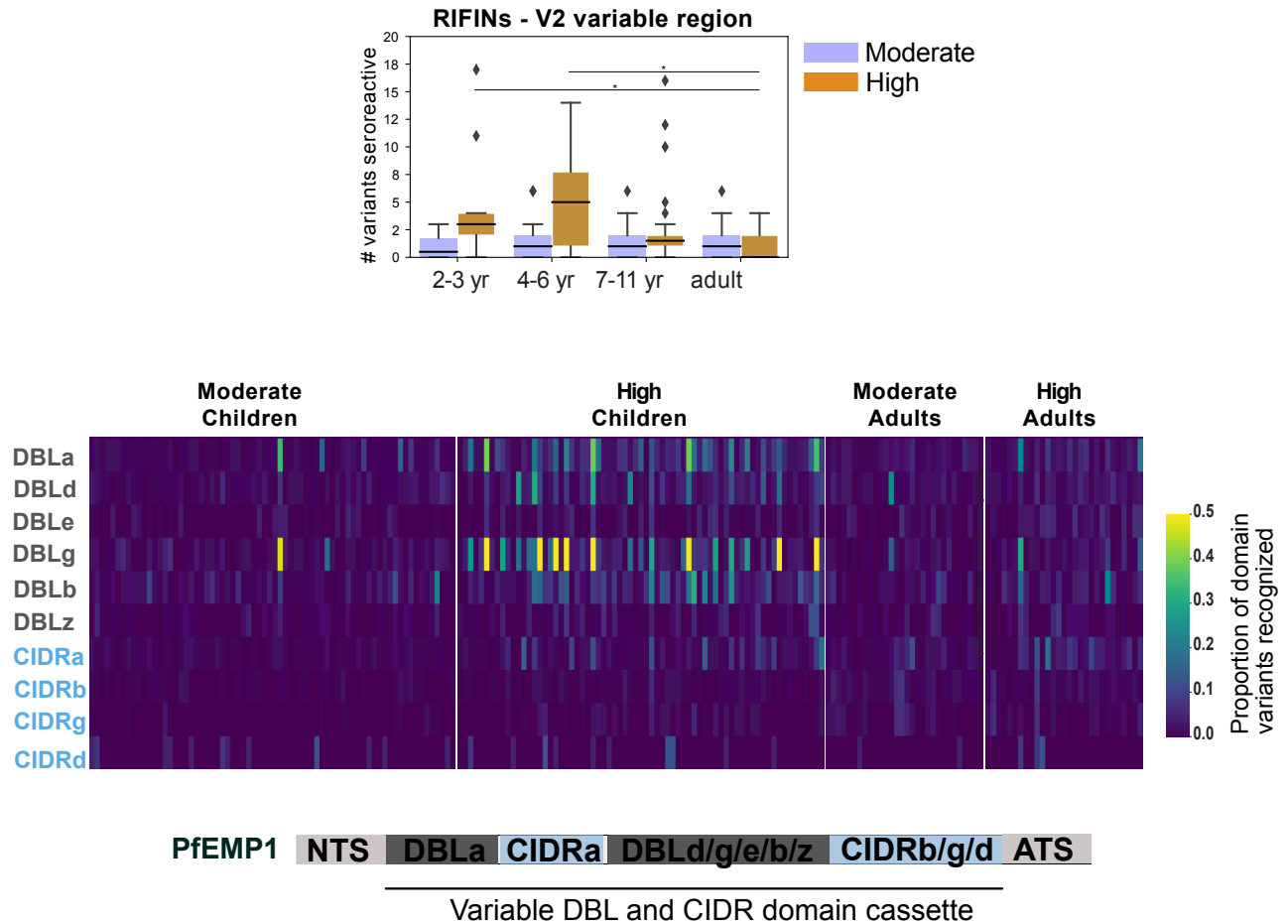
GO biological process complete	#	#	expected	Fold Enrichment	+/-	raw P value	FDR
multi-organism process	12	4	.23	17.23	+	1.87E-04	3.31E-02
entry into host	35	7	.68	10.34	+	1.13E-05	3.21E-03
↳movement in host environment	38	7	.74	9.52	+	1.82E-05	4.68E-03
↳biological process involved in interaction with host	253	24	4.89	4.90	+	1.75E-10	1.65E-07
↳biological process involved in symbiotic interaction	277	27	5.36	5.04	+	5.36E-12	7.60E-09
↳biological process involved in interspecies interaction between organisms	283	28	5.47	5.11	+	1.40E-12	3.97E-09
modulation by symbiont of host erythrocyte aggregation	73	14	1.41	9.91	+	5.57E-10	3.94E-07
↳modulation by symbiont of host cellular process	74	14	1.43	9.78	+	6.52E-10	3.08E-07
↳modulation by symbiont of host process	74	14	1.43	9.78	+	6.52E-10	2.64E-07
↳modulation of process of other organism involved in symbiotic interaction	74	14	1.43	9.78	+	6.52E-10	3.69E-07
↳modulation of process of other organism	74	14	1.43	9.78	+	6.52E-10	2.31E-07
cell-cell adhesion	63	7	1.22	5.74	+	3.25E-04	4.01E-02
↳cell adhesion	65	8	1.26	6.36	+	6.22E-05	1.47E-02
↳biological adhesion	91	11	1.76	6.25	+	2.86E-06	9.00E-04
pathogenesis	75	8	1.45	5.51	+	1.56E-04	2.95E-02
adhesion of symbiont to host	82	8	1.59	5.04	+	2.75E-04	3.71E-02

FigS2e - Results from GO enrichment analysis with the top 100 seroreactive proteins with the highest seroprevalence in the dataset shows enrichment of proteins at the host-parasite interface.
 Top - GO Cellular Component; Bottom - GO biological process.



FigS2f: Box plots of non-redundant seroreactive peptide groups per person in each group

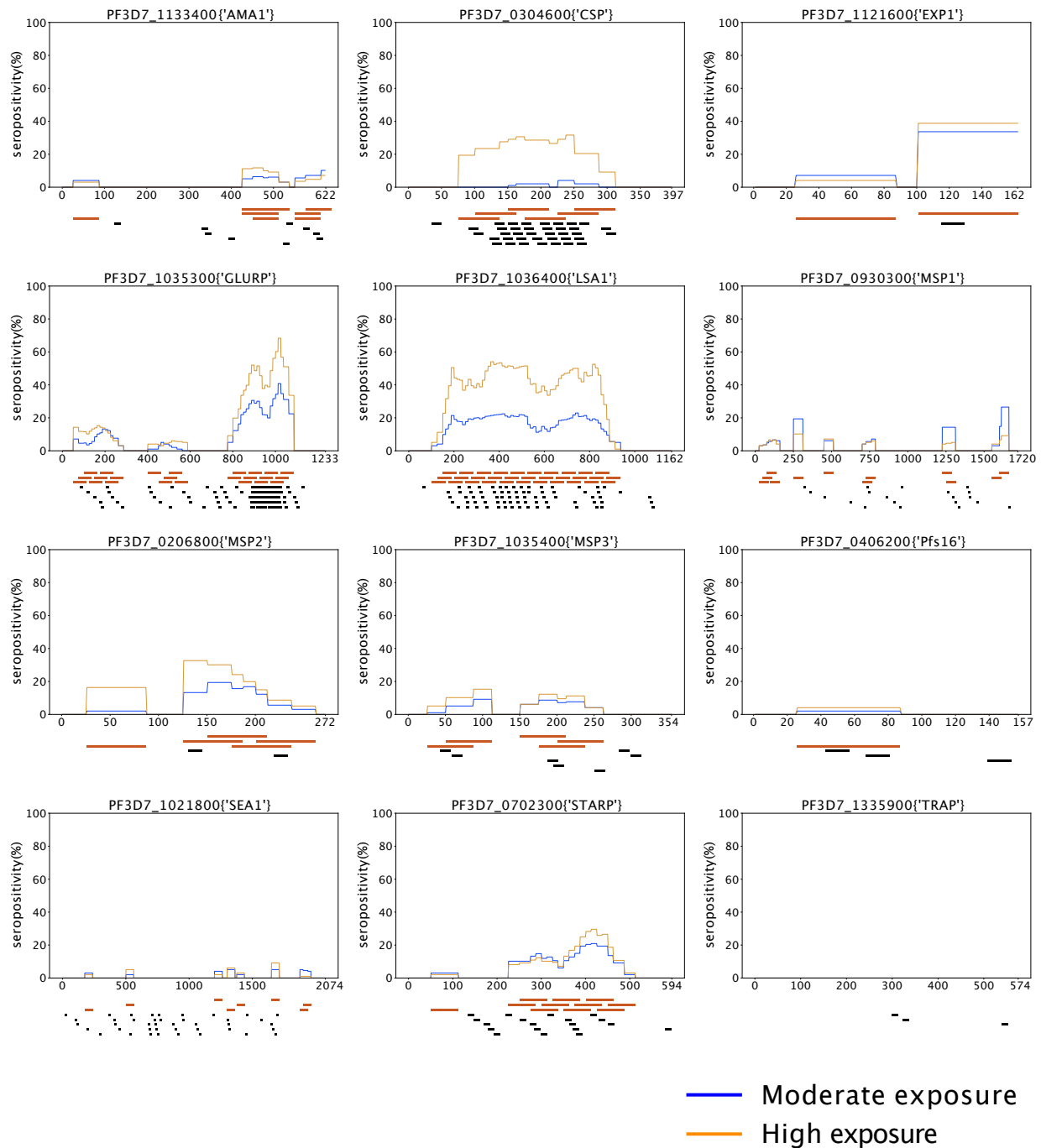
All seroreactive peptides in each person were collapsed based on sequence similarity (sharing of 7mer identical motifs). The resulting number of non-redundant groups were used as a measure of conservative non-shared breadth. Children from the moderate transmission setting had a significantly lower breadth than children from the high transmission setting and all adults by both approaches.* indicates p-value < 0.05 by KS-test. Median for each group is labeled on the side of the box.



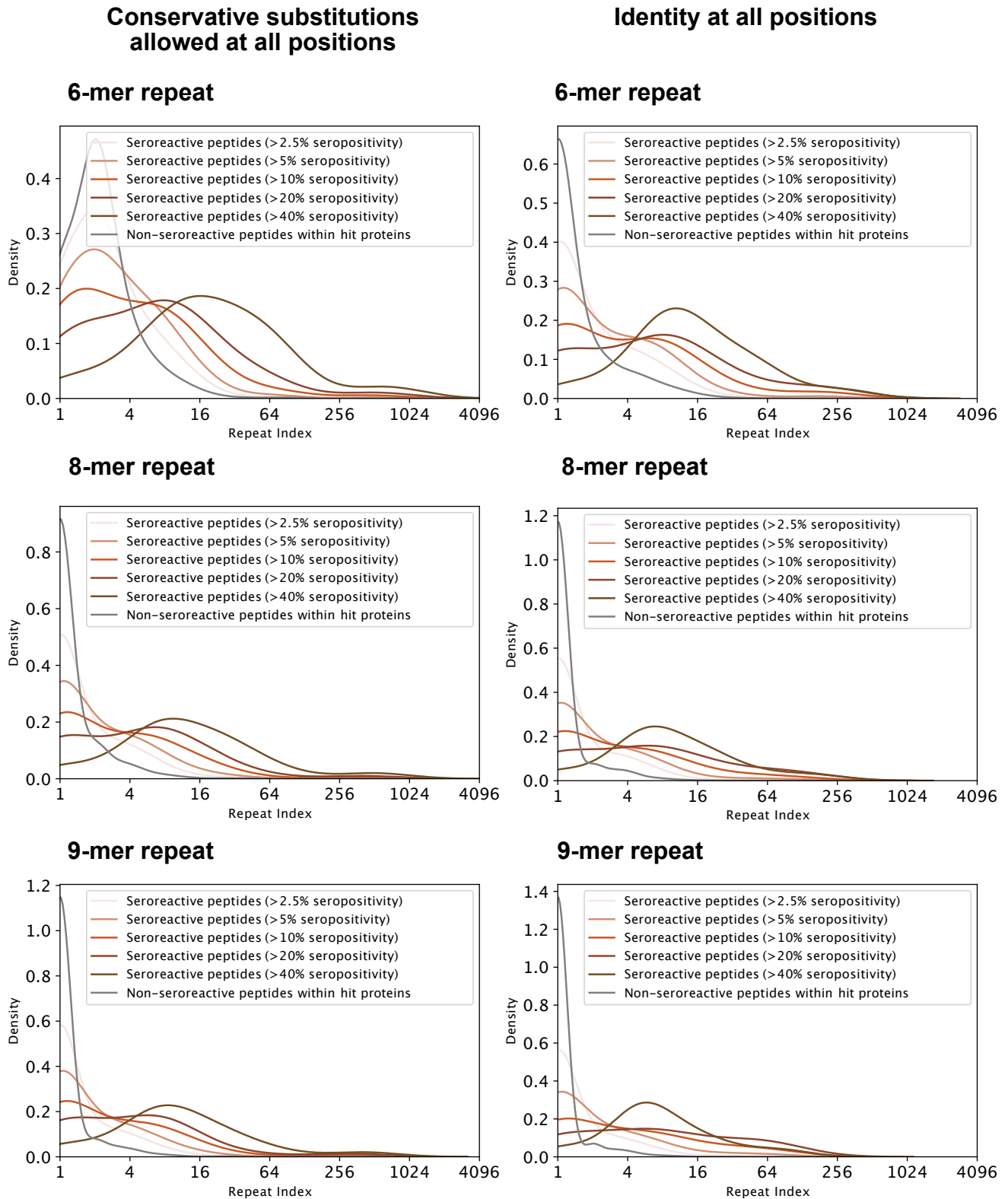
FigS2g: Breadth of seroreactivity in the variable regions of RIFIN and PfEMP1

Top - Number of domain variants seroreactive in the variable region V2 of RIFINs. Significantly different groups (KS test < 0.05) are marked with an *.

Bottom - Heatmap of proportion of variants from the library that are seroreactive in a given person for each PfEMP1 domain. Each column is a person. Schematic of domain structure of PfEMP1 is shown below the heatmap.



FigS3a - Location of seroreactive peptides identified in this dataset (red bar) and seroreactive 15-mer peptides identified using a high density peptide array (black bar) in Jaenisch et. al. (peptides with p-value <0.05 in (-) samples (malaria low parasitemia samples from Burkina Faso) over C (control - European samples)) for 12 vaccine candidates in that study. Average seropositivity per residue is plotted for moderate and high transmission samples in our study.



FigS4a: Distribution of repeat indices of seroreactive and non-seroreactive peptides within hit proteins for different lengths and degeneracy of repeating motif.

Left three: Conservative substitutions ([GA],[ST],[DE],[NQ],[RHK],[LVI],[YFW]) are allowed at all positions in the motif.

Right three: Identical residues at all positions in the motif.

For all 6 methods of defining repeats, all seroreactive regions were significantly different from the non-seroreactive set ($p < 0.01$ based on 1000 random samplings of non-seroreactive set).

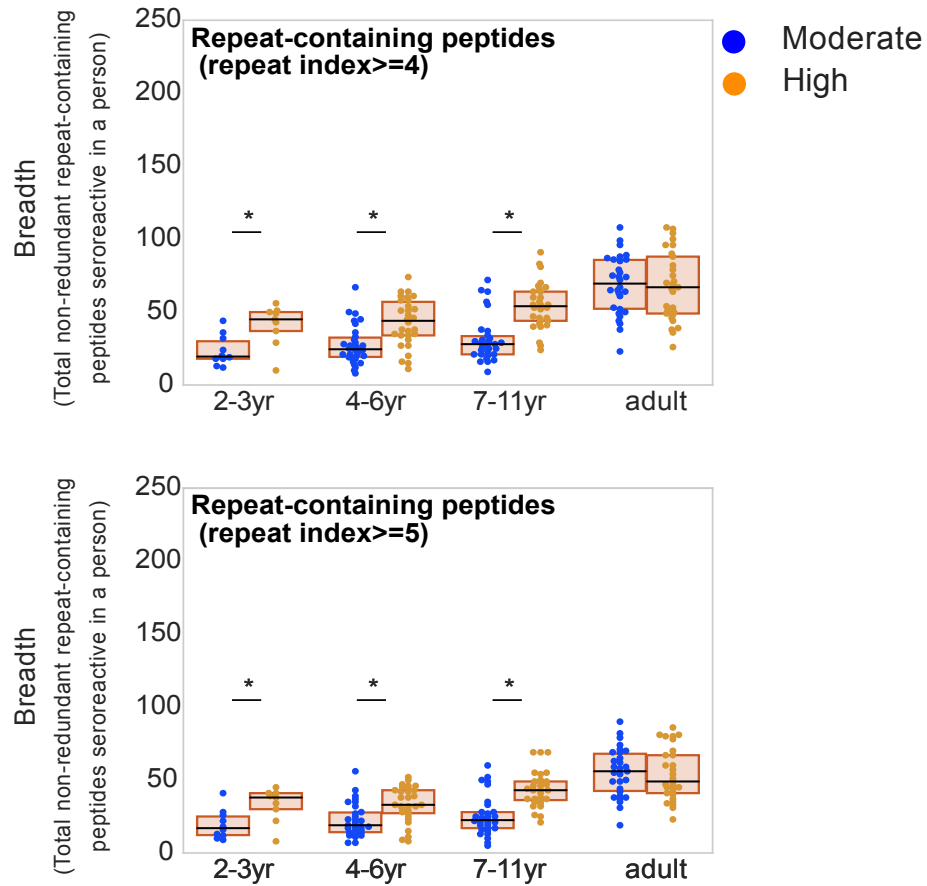
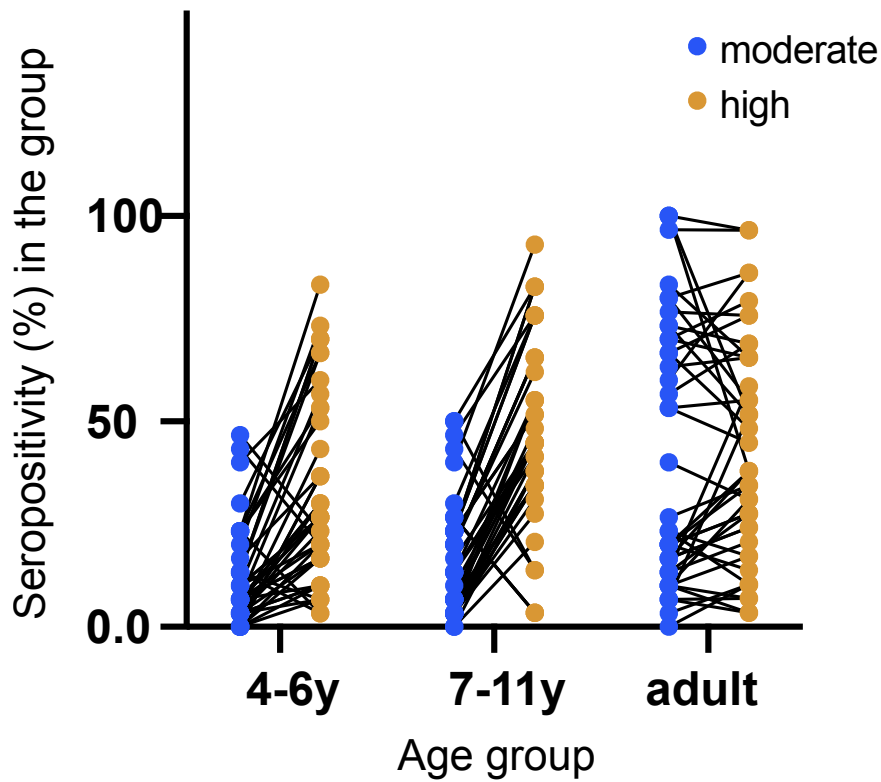


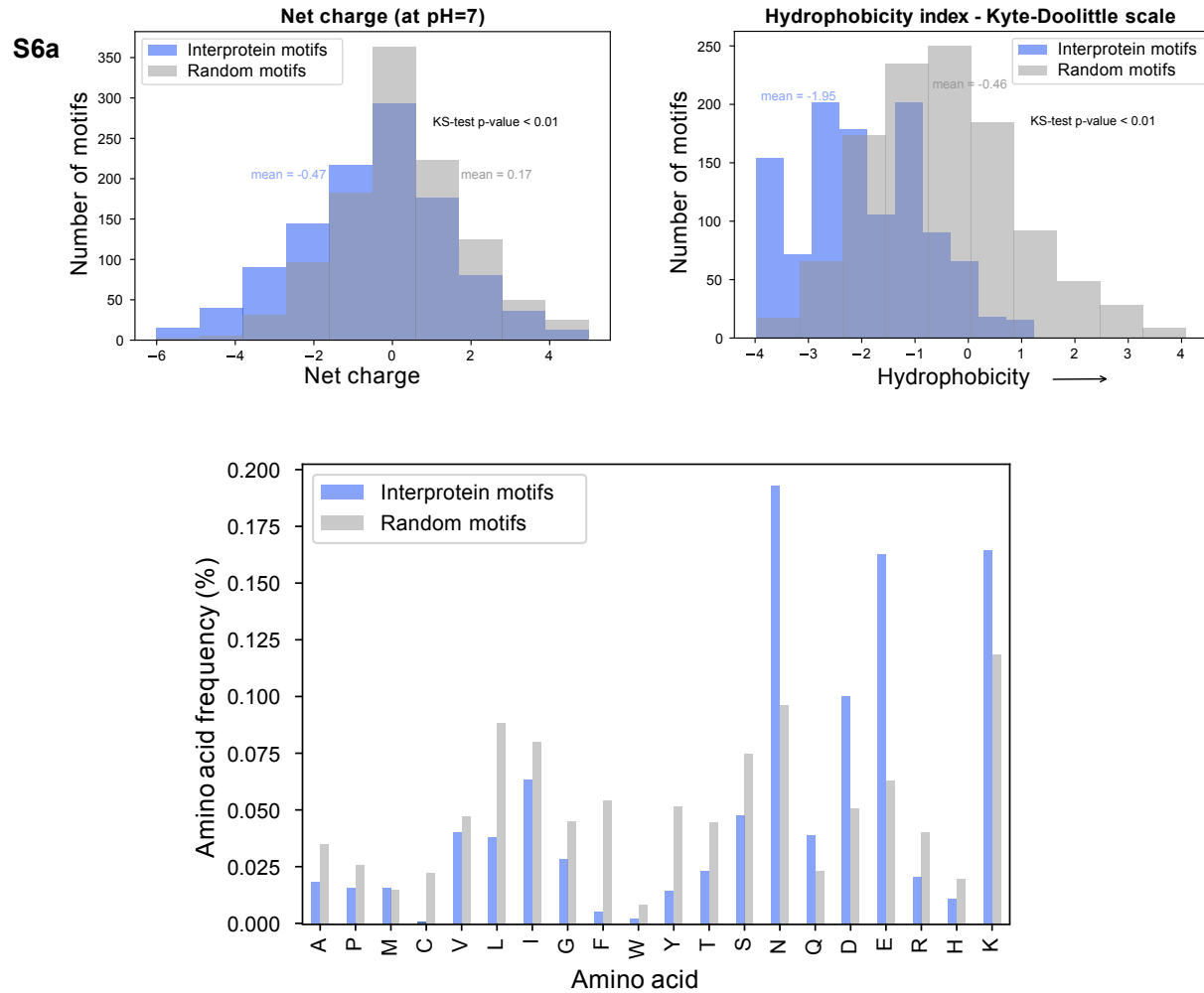
Fig S5a:

Breadth of repeat-containing peptides per person using different repeat index thresholds for categorizing repeat-containing peptides. Age groups showing significant difference between the two transmission settings are marked by * based on a KS-test p-value < 0.05 .

Repeat elements that are significantly different by exposure in 7-11 years



FigS5b: Seropositivity in all age and exposure groups for seroreactive repeat elements that are significantly different across exposure groups in 7-11 year olds. Each dot represents a seroreactive repeat element and seropositivity for the repeat element in a given group was calculated as the percent of people in that group enriching for any seroreactive peptide with that repeat element.

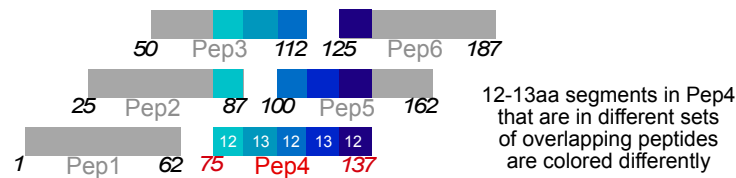


FigS6: All S6 figures are based on 7-aa motifs with at least 5 identical residues and up to 2 conservative substitutions.

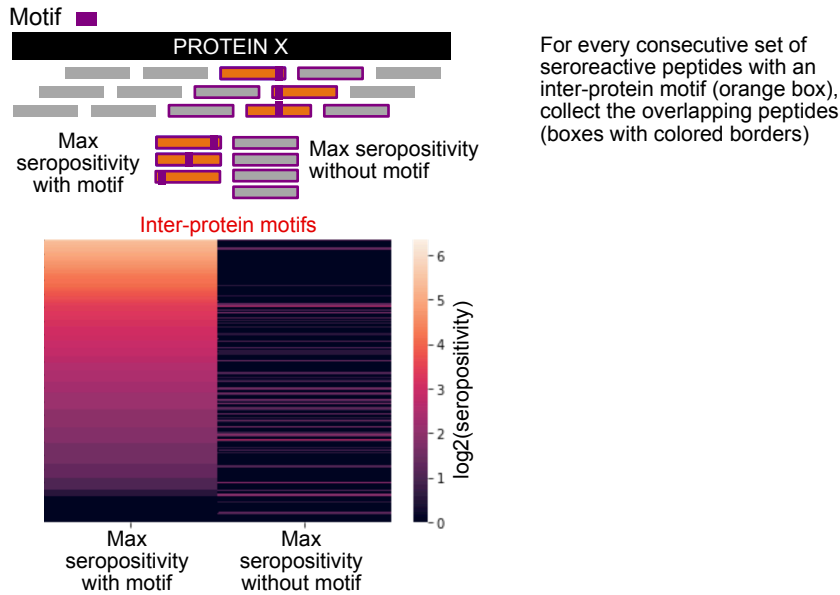
(a) Top - Histogram of net charge and hydrophobicity index of the 911 inter-protein motifs in comparison to a random set of 911 kmers of the same length from Pf proteome.

Bottom - Distribution of amino acid frequencies in interprotein and random motifs. All except Methionine (M) are significantly different between the two groups.

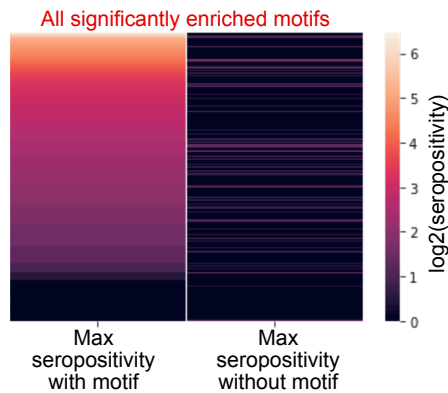
S6b



S6c



S6d



FigS6b-d

(b) Design of tiled peptide library and segments in overlapping peptides. Shown are overlapping segments shared with Peptide 4. Start and end amino acid positions of each peptide are marked at either ends.

(c) Comparison of maximum seropositivity of overlapping peptides with and without inter-protein motifs. Each row in the heatmap pertains to a collection of overlapping peptides surrounding a consecutive set of seroreactive peptides with an inter-protein motif.

(d) Same as in c, but for all 'enriched' motifs in seroreactive peptides.

S6e

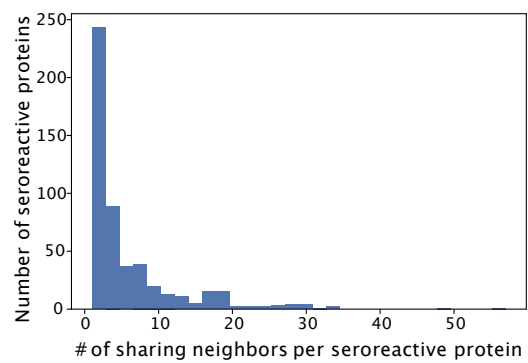
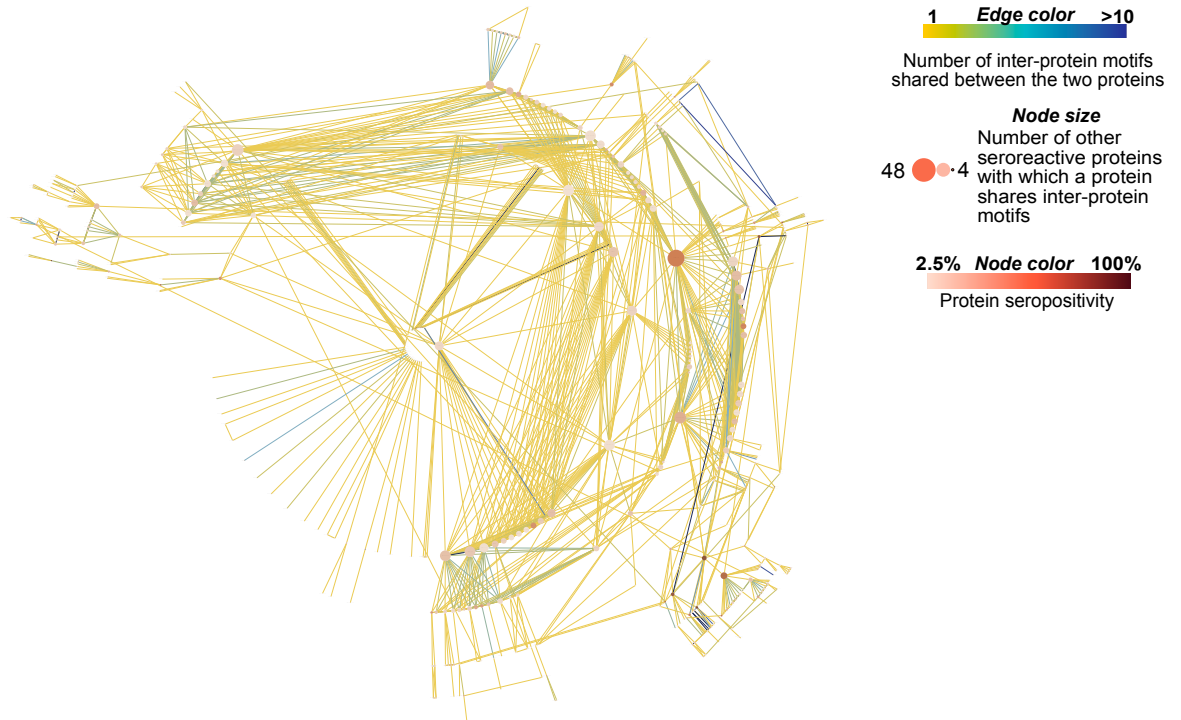
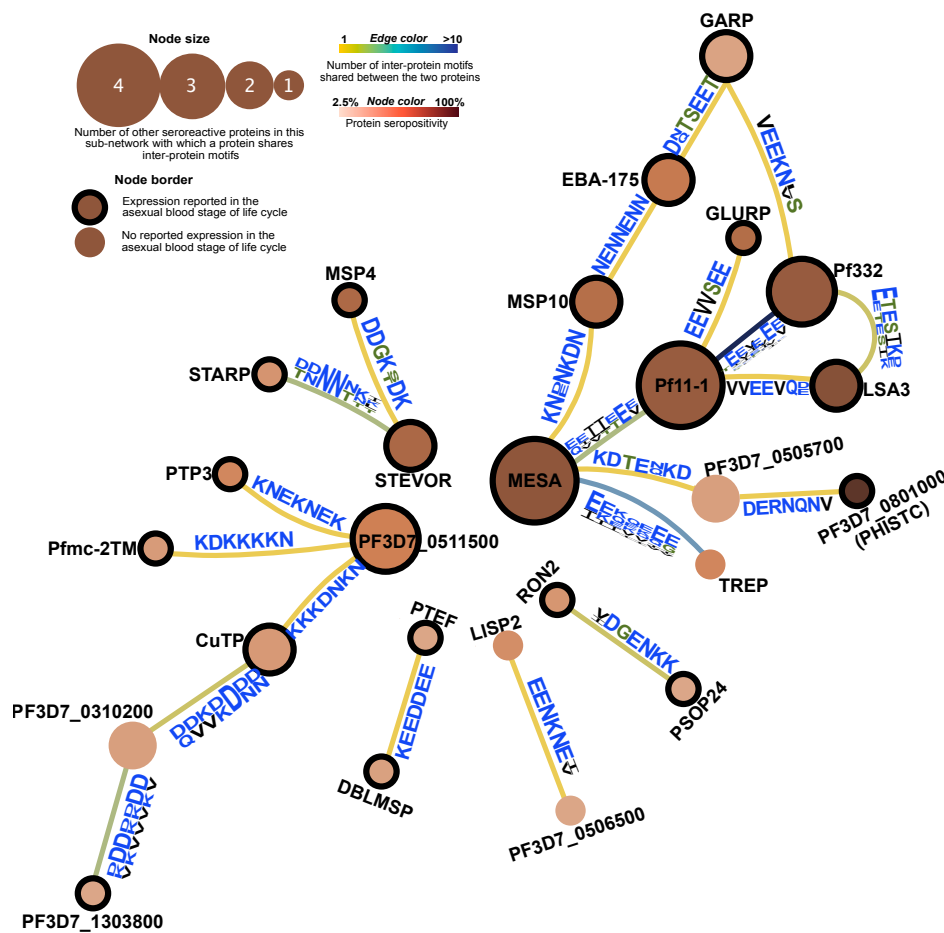


Fig S6e - Histogram of number of other seroreactive proteins with which a seroreactive protein shares inter-protein motifs.

S6f



S6g



FigS6f-g:

(f) Network of proteins outside the PfEMP1 network.

(g) Network of proteins with >30% seropositivity outside of PfEMP1 network