

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



**Biochemical and Bioinformatic Characterisation of Understudied
Erythrocyte Surface Expressed Hypervariable Protein Families in
Plasmodium falciparum.**

Hristina Chavdarova Vasileva

Thesis submitted in accordance with the requirements for the degree of

**Doctor of Philosophy
of the
University of London**

January 2025

Department of Infectious Biology

Faculty of Infectious and Tropical Diseases

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by the Joint Global Health Trials scheme (JGHTs)

(MRC/FCDO/NIRH/Wellcome Trust)

Research group affiliation(s): MRC Unit The Gambia at LSHTM
MATAMAL trial

Declaration

I declare that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated within the thesis.



Signature Date: 13.01.2025

Preface

According to the submission guidelines provided by the London School of Hygiene and Tropical Medicine, this thesis is presented as a “Research Paper Style Thesis.’ Three of the chapters in this thesis contain a total of two papers that are submitted for publication and in pre-print, and one chapter that is currently being prepared for submission to specific peer-reviewed journal. Due to the varying requirements of the journals, there may be some repetition of material and differences in formatting within these chapters. The publication details and acknowledgements for co-author contributions are provide on cover sheets for each individual paper. The rest of the thesis consists of additional material which includes two linking chapters between the research papers, and a background and discussion section to the research project.

All materials within this thesis were written by Hristina Vasileva.

Abstract

Pathogenesis of *Plasmodium falciparum* (*Pf*) malaria infection is dependent on parasite and host factors, and geographic and social factors, causing different clinical outcomes and disease severity. Parasite virulence is partly caused by evasion of the human host immune system during the blood stage of infection. Sequestration of parasite infected erythrocytes (IE) through cytoadherence are characteristic *Pf* virulence factors, enabled by parasite-derived proteins expressed on the surface of IEs. These proteins are antigenic and are associated with acquired immunity to *Pf*. IE surface-expressed antigens are associated with antigenic variability, called Variant Surface Antigens (VSAs). RIFIN and STEVOR are VSA protein families encoded respectively by approximately 180 *rif* and 40 *stevor* gene copies per parasite, expressing a single variant per parasite. Members of each family differ mostly in their hypervariable regions, which are exposed to the circulation and possess antigenic epitopes. Both variable domains are associated with *Pf* exposure and potentially clinical outcome. Seroreactivity and serorecognition to both protein families are age and exposure dependent, with higher reactivity in adults and higher domain recognition in individuals with clinical disease.

This study demonstrates the successful expression of isolated domains from two RIFIN and five STEVOR proteins as recombinant antigens, characterises their antigenicity, and demonstrates age-dependent immunity acquisition to the recombinants. Furthermore, the study reports the development of a specific *in-silico* model for the characterisation of STEVOR variants into clusters, after exploring other conventional methods for variant grouping. This model is then used to develop a library of eleven STEVOR variants as recombinant antigens and to further explore the breadth of antibody responses to the library in Sub-Saharan African populations, characterised by contrasting endemicity levels: high in Uganda and low in The Gambia and Guinea-Bissau.

The overall aim of the study was to develop a library of recombinant antigens from one of the understudied *P. falciparum* infected erythrocyte hypervariable protein families and explore their immunological profile in populations with contrasting malaria endemicity levels, aiming to investigate whether antibodies to these proteins contribute to the infection immunity to *Pf*.

Acknowledgements

Firstly, I would like to dedicate this work to my late grandmother, Goroka Vasileva, may she rest in peace. As a highly successful mathematician born in the 1940s in a small village of Bulgaria, she has always inspired me to be ambitious and pursue a career in academia. I wish she was among us to see me finish my doctoral degree, something she dreamed of seeing before she passed away.

I am deeply grateful to many people who have supported and assisted me throughout my PhD journey. This is especially true for my supervisors, Dr Kevin Tetteh and Dr Anna Last. Their experience, knowledge, and encouragement have been pivotal in shaping my research project and their constructive feedback have been crucial for helping me achieve my set research aims and goals. Additionally, I would like to thank my diverse advisory panel, consisting of Lindsey Wu, Michael Ooko and Ernest Diez Benavente, for their advises, support and contributions throughout my studies. Furthermore, I want to express my gratitude to Dr David Mabey for his continuous support and inspiration throughout my PhD journey and for the opportunities he created for me to attend scientific meetings to present my findings. I also want to thank Dr Martin Holland and Dr Martin Goodier for their continued mentoring and advice. Additionally, I would like to thank Eleanor Martins for the enormous support with various aspects apart from research activities, including travel arrangements, procurement of materials and reagents, and documentation, without which this project would not have been possible. I would like to acknowledge Elin Dumont, Helena Brazal-Monzó, Jack Bickford-Smith, James Ashall, and my MSc student Ana Chopo-Pizarro from LSHTM for their help with the laboratory aspects of my project. As well as Fatou K. Jaiteh, Sainabou Drammeh, Dr Mamadou-Ousmane Ndiath, Eniyou Oriero, Benjamin Kobna Njie and the data science department from the MRC Unit The Gambia for their support and encouragement during my time spent in the Unit as part of my project. Additionally, I would like to mention the entire team of the MATAMAL clinical trial

operating in the Bijagos Islands of Guinea-Bissau for their guidance and support during my fieldwork, involving the collection of dried blood spots samples, as well as supporting my participation in the ASTMH 2023 conference in Chicago, Illinois, where I presented my PhD research. I have also greatly benefited from the brainstorming sessions with my fellow PhD colleagues Christian Kositz, Ismaila Manneh and Alex Keeley which have helped me critically think, analyse, and interpret my data results. Many thanks to Dr Chris Drakeley and Dr Seyi Soremekun for overseeing my PhD progress and providing timely input when needed.

I would finally like to thank my family, especially my mother and father, as well as my friends, for the immense emotional support throughout this journey. Without them, I would not have reached where I am today.

Table of Content

Declaration	2
Preface	3
Abstract	4
Acknowledgements	6
List of Figures	13
List of Tables	13
Chapter 1: Background	39
1.1 Malaria: History, Burden, Diagnosis and Prevention.....	39
1.2 <i>P. falciparum</i> serology as a tool for understanding transmission, protection, and infection exposure.	47
1.3 Naturally acquired immunity to <i>Plasmodium falciparum</i> malaria disease.....	53
1.4 Variant surface antigens	57
1.5 STEVOR and RIFIN: structure, function, and localisation.....	63
1.6 STEVOR: variability and host immune response	67
1.7 Importance of pathogen protein hypervariability	70
1.8 References	71
Chapter 2: Study Aim and Objectives	77
2.1 Aim.....	77
i. Hypothesis.....	77
2.2 Objectives.....	77
2.2.1 Objective one.....	77
i. Hypothesis.....	77
2.2.2 Objective two	78
i. Hypothesis.....	78
2.2.3 Objective three.....	78
i. Hypothesis.....	78
Chapter 3: Research Paper 1: Novel <i>Plasmodium falciparum</i> hypervariable erythrocyte surface expressed recombinant proteins and their potential as serological markers of infection exposure.	79
Abstract	80
3.1 Introduction	81
3.2 Methods.....	84
3.2.1 Ethics	84
3.2.2 Design and expression of SETVOR and RIFIN recombinant proteins	84

3.2.3 Serum pools.....	86
3.2.4 Detection of optimum recombinant protein antigenicity.....	87
3.2.5 Multiplex bead-based serological assay.....	88
3.2.6 Data analysis.....	89
3.3 Results.....	90
3.3.1 Design and expression of STEVOR and RIFIN recombinant proteins.....	90
3.3.2 Sample summary statistics.....	93
3.3.3 Age-related seropositivity.....	94
3.4 Discussion.....	99
3.5 Funding.....	103
3.6 Acknowledgements.....	103
3.7 Conflict of interests.....	103
3.8 List of non-standard abbreviations.....	104
3.9 References.....	106
3.10 Supplementary materials.....	114
Research Paper 1 cover sheet.....	125
Chapter 4: Methods for analysis of STEVOR protein sequences (Linking material to Chapter 5)	127
4.1 Introduction.....	127
4.2 Methods.....	129
4.2.1 Protein sequences alignment and hypervariable domain isolation.....	129
4.2.2 Clustering using phylogenetic trees.....	130
4.2.3 Clustering in <i>k</i> -means using dissimilarity matrix.....	130
4.2.4 Clustering using sequence digitalisation and principal component analysis.....	131
4.3 Results.....	132
4.4 Discussion.....	139
4.5 Conclusion.....	141
4.6 References.....	142
Chapter 5: Research paper 2: <i>In-silico</i> classification and antigen library expression of <i>Plasmodium falciparum</i> STEVOR hypervariable infected erythrocyte surface-expressed multivariant protein family	145
Abstract.....	146
5.1 Introduction.....	148
5.2 Methods.....	152
5.2.1 Protein sequences alignment and inspection.....	152

5.2.2 Domain sequences isolation	153
5.2.3 Clustering model and variant library selection	153
5.2.4 Recombinant antigens library expression	154
5.3 Results	156
5.3.1 STEVOR protein panel selection	156
5.3.2 Expression of selected recombinant proteins in <i>Escherichia coli</i>	161
5.4 Discussion	164
5.5 Limitations	168
5.6 Conclusion	168
5.7 Declarations	169
5.7.1 Ethical approval and consent to participate	169
5.7.2 Availability of data	169
5.7.3 Competing interests	169
5.7.4 Funding	169
5.7.5 Authors contributions	170
5.7.6 Acknowledgements	170
5.8 List of abbreviations	171
5.9 References	173
5.10 Supplementary materials	181
Research Paper 2 cover sheet	192
Chapter 6: Description of Study Sites for sample selection for Research Paper 3: “Breadth of endemic serum antibody responses to <i>Plasmodium falciparum</i> STEVOR recombinant protein antigen library” (Linking material to Chapter 7).....	194
6.1 Introduction	194
6.2 FIGHTMAL	195
6.2.1 Study site and malaria endemicity	195
6.2.2 Study design and population	196
6.3 MASSIV	198
6.3.1 Study site and malaria endemicity	198
6.3.2 Study design and population	199
6.4 MATAMAL	201
6.4.1 Study site and endemicity	201
6.4.2 Study design and population	203
6.5 Discussion	205
6.6 References	206

Chapter 7: Research Paper 3: Breadth of endemic serum antibody responses to <i>Plasmodium falciparum</i> STEVOR recombinant protein antigens library	209
Abstract	210
7.1 Introduction	212
7.2 Methods	215
7.2.1 Samples source	215
7.2.2 Study populations	215
7.2.3 Recombinant protein panel	217
7.2.4 Multiplex bead-based serological assay	217
7.2.5 Statistical Analysis	218
7.3 Results	219
7.3.1 Samples statistics	219
7.3.2 FIGHTMAL seroprevalence	221
7.3.3 MASSIV/MATAMAL seroprevalence	225
7.3.4 Recombinant seroreactivity profiles	229
7.3.5 Breadth of responses to STEVOR V2 recombinants	233
7.3.6 Seropositivity correlation	236
7.4 Discussion	238
7.5 Conclusion	243
7.6 Limitations	244
7.7 Declarations	245
7.7.1 Competing interests	245
7.7.2 Authors contributions	245
7.8 List of abbreviations	246
7.9 References	247
7.10 Supplementary materials	254
Research Paper 3 cover sheet	279
Chapter 8: Discussion, Future Work and Limitations	281
8.1 Summary of research findings	281
8.1.1 Research Paper 1	281
Aim	281
Objectives	281
Discussion	281
8.1.2 Research Paper 2	284
Aim	284

Objectives	284
Discussion	284
8.1.3 Research Paper 3	286
Aim	286
Objectives	286
Discussion	286
8.2 Future work	288
8.3 Limitations	288
8.4 References	291
Appendices	296

List of Figures

Chapter 1: Background

- **Figure 1:** *Plasmodium falciparum* life cycle. Illustration of the complete *P. falciparum* life cycle, highlighting the stages in the human host, and the mosquito vector.
- **Figure 2:** Figure illustrating the general structure of the main immunoglobulin isotypes. Each rectangle represents a domain, and the C-region, which determines the isotype is illustrated as the main colour for each isotype.
- **Figure 3:** Illustration of the proposed localisation and 2TM topology of hypervariable surface proteins on the membrane of *P. falciparum* infected erythrocytes.
- **Figure 4:** General protein structure of RIFIN and STEVOR protein family A) protein domains and number of copies and B) protein localisation on the IEs. Same colours represent domains with high genetic similarity between the families. The red colour represents domains with high variability.
- **Figure 5:** *stevor* genes position on *P. falciparum* chromosomes, protein expression, domain breakdown, and localisation on the IE within and infection. Chromosome one is used as an example.

Chapter 3: Research Paper 1

- **Figure 1:** General structure of the RIFIN and STEVOR proteins showing the domain architecture of the proteins and the approximate length in amino acids of their variable domains (V2) and semi-conserved domains (SC). Transmembrane regions were identified using the TMHMM v.2.0 transmembrane domain prediction server.
- **Figure 2:** ELISA results of control serum pools (1/200 serum dilution), challenged against four steps six-fold dilution series of each recombinant protein. Brefet: Serum

pool of hyperimmune adults from The Gambia; CP3: LSHTM in house serum pool of hyperimmune Tanzanian adults; DTNMaPa: Serum pool of 40 *Pf* infection confirmed adults, with high antibody responses to *Pf* markers of exposure from Bijagos Archipelago, Guinea-Bissau; PRISM1: Serum pool of 40 Ugandan adults with confirmed *Pf* infection and clinical disease; PRISM2: Serum pool of 80 Ugandan all age individuals with high antibody responses to established *Pf* markers of exposure; WHO: WHO 1st international *Pf* serum from Kenyan adult population. The control serum selected from this experiment is PRISM1, indicated in red. The last plot of the series demonstrates the seroreactivity of all recombinants against PRISM1, with color-coded intensity in descending order.

- **Figure 3:** Population seroprevalence (%) per survey year for A) long-term markers; B) short, medium-term markers C) STEVOR/RIFIN recombinants and D) internal serology control, stratified by age group. Seroprevalence threshold per recombinant was selected using a cut off mean MFI from PHE negative samples + three times standard deviation. Bar plots are paired colour coded, for visual representation of differences in seroprevalence between the two surveyed years. Tet.tox represents the internal positive human serological control. Seroprevalence of 0% is indicated with an asterisk and exact seroprevalence percentages are summarised in Supplementary Table 2.
- **Figure 4:** Forest plot of odds ratio analysis per recombinant antigen of being seropositive in 2016 compared to 2013, calculated in seroprevalence (%), stratified by age: A) 6 months to 5 years old, B) 5 years to 11 years old, and C) more than 18 years old. Squares graphically represent the calculated and summarised in the table's ORs with 95% confidence intervals (CIs) as whiskers. CIs of above 10 are not graphically represented but are displayed in the table on the left. Grey lines separate the antigens into three groups: long-term markers of infection, short-term markers of infection and STEVOR/RIFIN recombinants from top to bottom.

- **Supplementary Figure 1:** Recombinant proteins titrations on MagPlex microsphere beads against two positive controls: pool serum from Tanzania (CP3) and pool serum from Uganda (PRISM1). Optimum coupling concentration (EC50 point of saturation) was calculated using the median EC50 value from all curves subjected to 4-parameter logistic regression.
- **Supplementary Figure 2:** Sampling distribution per time of the year for the two sample time points (2013 and 2016). Sample IDs are displayed on the y-axis Out of all samples selected for the study indicated in the red rectangles, 74% are paired samples between the 2013 and 2016 timepoints.
- **Supplementary Figure 3.1:** Control standard curves of PRISM1 from 1/10 down to 1/31250 serum dilution for plates 1 to 5 of Luminex assay for the following antigens: S1_SC: STEVOR1_SC, S1_V2: STEVOR1_V2, S6_SC: STEVOR6_SC, S6_V2: STEVOR6_V2, S5_SC: STEVOR5_SC, R2_SC: RIFIN3_SC.
- **Supplementary Figures 3.2:** Control standard curves of PRISM1 from 1/10 down to 1/31250 serum dilution for plates 1 to 5 of Luminex assay for the following antigens: R3_V2: RIFIN3_V2, HSP40.Ag1, Hyp2, Etramp5.Ag1, Etramp4.Ag2, Rh2.203.
- **Supplementary Figures 3.3:** Control standard curves of PRISM1 from 1/10 down to 1/31250 serum dilution for plates 1 to 5 of Luminex assay for the following antigens: GEXP18, AMA1, MSP1.19, MSP2. Dd2, MSP2, CH150, GLURP.R2.
- **Supplementary Figures 3.4:** Control standard curves of PRISM1 from 1/10 down to 1/31250 serum dilution for plates 1 to 5 of Luminex assay for the following control recombinants: Tet.tox and GST.

Chapter 4: Methods for sequence characterisation (Linking material)

- **Figure 1:** MAFFT multiple alignment of 42-variable domain 3D7 STEVOR sequences performed using G-INS-I with Smith-Waterman pairwise algorithm and gap extension penalty for group-to-group variation value of 0.123 according to BIC, displayed using AliView open-source alignment viewer and editor software. The alignment is color-coded according to amino acids.
- **Figure 2:** Two-dimensional scaling of dissimilarity matrix from the phylogenetic tree of PlasmoDB STEVOR V2 protein sequences, applied using A) three *k*-means and B) 11 *k*-means.
- **Figure 3:** Neighbour joining phylogenetic tree of PlasmoDB database STEVOR V2 protein sequences, colour coded according to the *k*-mean position from Table 1 according to distance matrix multidimensional scaling applied using A) three *k*-means and B) 11 *k*-means.
- **Figure 4:** First and second principal components from the PCA analysis applied on the PlasmoDB 3D7 STEVIR V2 sequences after digitalisation and matrix transformation using singular value decomposition. Variants are colour coded according to their *k*-mean position from Table 1 when multidimensional scaling of dissimilarity matrix was applied using A) three *k*-means and B) 11 *k*-means. Percentages indicate the proportion diversity explained by each of the two principal components.

Chapter 5: Research Paper 2

- **Figure 1:** Schematic representation of a generic STEVOR protein. The architecture is in the following sequence: signal peptide (SP); short variable domain (V1); PEXEL motif; approximately 120 amino acid long semi-conserved domain (SC); two transmembrane domains (TM) flanking a large hypervariable domain (V2) composed of approximately 70 amino acids, length slightly varying between variants; ending with a 17 amino acid long conserved domain (C).
- **Figure 2:** Shannon Entropy Index plot representing the level of diversity in each amino acid position of the STEVOR domain alignments. The higher the value, the higher the diversity. Double lines on the x-axis represent the end of one and the beginning of the next (position 0) protein domain. Legend values represent the overall mean diversity of the pairwise distances in each of the alignments.
- **Figure 3:** Principal Component Analysis showing the first two principal components (PC1 and PC2) for A) Large variable domain (V2), B) Semi-conserved domain (SC), and C) Conserved domain (C) of the STEVOR protein sequences. Sequences are colour coded according to the isolate they belong to. Percentage values represent the proportion of variation explained by each of the two principal components.
- **Figure 4:** Visualisation of the large variable domain (V2) sequences selected to be expressed as STEVOR recombinant proteins library, shown in red, among all 493 V2 amino acid sequences used in the PCA analysis (shown in black). The numbers indicate the quadrant number (labelled 1-9) in which each of the library sequences is positioned, summarised in Table 2.
- **Supplementary Figure 1:** Output of the TMHMM 2.0 transmembrane domain predicting software performed on the full-length amino acid sequences of each of the selected variants for the STEVOR V2 recombinant library. Transmembrane domains are

predicted and displayed in purple, intercellular portions of the protein in blue and extracellular in orange. Signal peptides, marking the beginning of each protein sequence are also displayed in purple. All 11 graphics are displayed, and the identification name of each sequence is displayed above its representative graph.

- **Supplementary Figure 2:** Coomassie Brilliant Blue R-250 stain of 4-15% BioRad precast Mini-PROTEAN TGX SDS-PAGE gels of **A)** Molecular weight ladder (L), Pf3D7_1300900_SC (1), Pf3D7_0832000_SC (2), Pf3D7_0832600_SC (3), PfGN01_040031100_V2 (4), PfIT_130077600_V2 (5), PfSN01_140006300_V2 (6), PfGN01_100006100_V2 (7), PfKE01_10000580_V2 (8), and **B)** Molecular weight ladder (L), PfSN01_030005600_V2 (9), PfHB3_040028900_V2 (10), PfGA01_070034600_V2 (11), PfGN01_020006800_V2 (12), PfSN01_000011500_V2 (13), PfGN01_130006600_V2 (14).
- **Supplementary Figure 3:** Graphical summary of titration of all 14 recombinants at six-point eight-fold dilution from 1000 µg/ml down to 0.0305 µg/ml against five-point two-fold dilution series of the PRISM positive control serum pool from 1/100 down to 1/1600. The first three graphs, both from A) and first from B) represent the semi-conserved domain recombinants and the rest, down to G) represent the variable domain recombinants.
- **Supplementary Figure 4:** Shannon Entropy Index plot representing the level of diversity in each amino acid position of the STEVOR semi-conserved domain alignment. The higher the value, the higher the diversity. The black portions represent the highly conserved regions of the domain. The x-axis represents the amino acid positions of the domain. Legend values represent the approximate entropy of the different parts of the SC domain.
- **Supplementary Figure 5:** Visualisation of the large variable domain (V2) sequences selected to be expressed as STEVOR recombinant protein library, shown in red, among the 80 available plasmids in LSHTM from all 493 V2 amino acid sequences used in the PCA analysis (shown in

black). The numbers indicate the quadrant number (labelled 1-9) in which each of the library sequences is positioned, summarised in Table 2.

- **Supplementary Figure 6:** A graphical representation of a phylogenetic tree of all 493 STEVOR V2 domain sequences generated on IQtree using JTT+F+G4 model and 1000 bootstrap replicates, optimised using hill-climbing nearest neighbour interchange (NNI) and visualised using iTOL: Tree of Life tool. Sequences names are colour coded according to the strain they belong to, indicated in the legend. Sequences names with larger font represent the 11 selected variants for the recombinant library.

Chapter 6: Description of Study Sites for sample selection for Research Paper 3: “Breadth of responses to a library of STEVOR recombinant antigens in populations characterised with different *P. falciparum* endemicity levels”. (Linking material)

- **Figure 1:** Figure adapted from the official site of Apac District, Government of Uganda. The area in red indicates the borders of the Apac district within Uganda in yellow.
- **Figure 2:** FIGHTMAL longitudinal cohort study profile with participants per time point of sampling, intervention type, study eligibility criteria and type of collected samples.
- **Figure 3:** Map of the demographic division of The Gambia, MASSIV was conducted in the Upper River region. Figure is adapted from Robello M. *et. al* (2015), *PLOS NTD*.
- **Figure 4:** MASSIV clinical trial study profile with participants per trial arm, intervention type, study inclusion criteria and collected samples.
- **Figure 5:** Map of the Bijagos Archipelago of Guinea-Bissau with MATAMAL clusters marked as stars and landmarks of islands separated into different clusters marked as lines. Figure is adapted from Hutchins H. *et al.* (2023), *BMJ Open*.
- **Figure 6:** MATAMAL clinical trial study profile with participants per trial arm, intervention type, study exclusion criteria and collected samples.

Chapter 7: Research Paper 3

- **Figure 1:** Percentage seropositive FIGHTMAL samples for baseline (prior to treatment) (salmon), 6 weeks after ACT treatment (pink) and 16 weeks after ACT treatment (red), stratified by age-group (< 5years, 5-15 years and > 15 years). Recombinant antigens are shown on the x-axis. STEVOR antigens with the _SC suffix refer to semi-conserved antigen targets. STEVOR antigens *PfGNO1_130006600* – *PfGA01_070034600* were all based on the large hypervariable loop.
- **Figure 2:** Percentage seropositive MASSIV/MATAMAL samples for the Control (salmon) and Intervention (red) treatment arms, stratified by age-group (< 5 years, 5-15 years and > 15 years). Recombinant antigens are shown on the x-axis. STEVOR antigens with the _SC suffix refer to semi-conserved antigen targets. STEVOR antigens *PfGN01_130006600* – *PfGA01_070034600* were all based on the large hypervariable loop.
- **Figure 3:** *PfAMA1* raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.
- **Figure 4:** STEVOR1_SC raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity

threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.

- **Figure 5:** *PfGA01_070034600* raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.
- **Figure 6:** FIGHTMAL population (n=1,084) prevalence of the breadth of antibody responses to the STEVOR V2 recombinant antigens according to four categories (1-3, 4-6, 7-9, and 10-11 recombinant antigens). Breadth of responses prevalence is stratified by age group (< 5 years, 5-15 years and > 15 years) and sampling timepoint: Baseline (grey), 6 weeks post treatment (light purple) and 16 weeks post treatment (dark purple), indicated in the legend.
- **Figure 7:** MASSIV/MATAMAL population (n=1,209) prevalence of breadth of antibody responses to the STEVOR V2 recombinant antigens according to four categories (1-3, 4-6, 7-9, and 10-11 recombinant antigens). Breadth of responses prevalence is stratified by age group (< 5 years, 5-15 years and > 15 years) and intervention arms: Control (grey) and Intervention (purple), indicated in the legend.
- **Supplementary Figure 1.1:** *PfMSP1.19* raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity

threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.

- **Supplementary Figure 1.2:** Etramp5.Ag1 raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.
- **Supplementary Figure 1.3:** CSP raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.
- **Supplementary Figure 1.4:** HSP40.Ag1 raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.
- **Supplementary Figure 1.5:** Rh5.1 raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and

the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.

- **Supplementary Figure 1.6:** STEVOR5_SC raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.
- **Supplementary Figure 1.7:** STEVOR6_SC raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.
- **Supplementary Figure 1.8:** *PfGN01_130006600* raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.

- **Supplementary Figure 1.9:** *PfGN01_100006100* raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.
- **Supplementary Figure 1.10:** *PfGN01_020006800* raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.
- **Supplementary Figure 1.11:** *PfGN01_040031100* raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.
- **Supplementary Figure 1.12:** *PfSN01_000011500* raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the

seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.

- **Supplementary Figure 1.13:** *Pf*SN01_140006300 raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.
- **Supplementary Figure 1.14:** *Pf*SN01_030005600 raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.
- **Supplementary Figure 1.15:** *Pf*KE01_100005800 raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.
- **Supplementary Figure 1.16:** *Pf*IT_130077600 raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of

the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.

- **Supplementary Figure 1.17:** *Pf*HB3_040028900 raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.
- **Supplementary Figure 1.18:** Tetanus.toxoid raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.
- **Supplementary Figure 2:** FIGHTMAL population prevalence of the breadth of antibody responses to the STEVOR V2 recombinant antigens according to five categories (0, 1-3, 4-6, 7-9, and 10-11 recombinant antigens). Breadth of responses prevalence is stratified by age group (< 5 years, 5-15 years and > 15 years) and sampling timepoint: Baseline (grey), 6 weeks post treatment (light purple) and 16 weeks post treatment (dark purple), indicated in the legend. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.

- **Supplementary Figure 3:** MASSIV/MATAMAL pool population prevalence of breadth of antibody responses to the STEVOR V2 recombinant antigens according to five categories (0, 1-3, 4-6, 7-9, and 10-11 recombinant antigens). Breadth of responses prevalence is stratified by age group (< 5 years, 5-15 years and > 15 years) and intervention arms: Control (grey) and Intervention (purple), indicated in the legend.
- **Supplementary Figure 4:** Percentage seropositive MASSIV versus MATAMAL samples for the Control (light green and light blue, respectively) and Intervention (green and blue, respectively) treatment arms, stratified by age-group (< 5 years, 5-15 years and > 15 years). Recombinant antigens are shown on the x-axis. STEVOR antigens with the _SC suffix refer to semi-conserved antigen targets. STEVOR antigens *PfGN01_130006600* – *PfGA01_070034600* were all based on the large hypervariable loop. There was weak evidence for significant differences of seropositivity values between MASSIV and MATAMAL with a p-value of 0.03 according to Man-Whitney non-parametric test.
- **Supplementary Figure 5.1:** *PfAMA1* raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.
- **Supplementary Figure 5.2:** *PfMSP1.19* raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI

plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.

- **Supplementary Figure 5.3:** *PfEtramp5.Ag1* raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.
- **Supplementary Figure 5.4:** *PfCSP* raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.
- **Supplementary Figure 5.5:** *PfHSP40.Ag1* raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.
- **Supplementary Figure 5.6:** *PfRh5.1* raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud

represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.

- **Supplementary Figure 5.7:** STEVOR1_SC raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.
- **Supplementary Figure 5.8:** STEVOR5_SC raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.
- **Supplementary Figure 5.9:** STEVOR6_SC raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.

- **Supplementary Figure 5.10:** *PfGA01_070034600* raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.
- **Supplementary Figure 5.11:** *PfGN01_130006600* raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.
- **Supplementary Figure 5.12:** *PfGN01_100006100* raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.
- **Supplementary Figure 5.13:** *PfGN01_020006800* raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative

MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.

- **Supplementary Figure 5.14:** *PfGN01_040031100* raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.
- **Supplementary Figure 5.15:** *PfHB3_040028900* raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.
- **Supplementary Figure 5.16:** *PfIT_130077600* raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.
- **Supplementary Figure 5.17:** *PfKE01_100005800* raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the

cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.

- **Supplementary Figure 5.18:** *Pf*SN01_000011500 raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences,
- **Supplementary Figure 5.19:** *Pf*SN01_140006300 raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.
- **Supplementary Figure 5.20:** *Pf*SN01_030005600 raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.

- **Supplementary Figure 5.21:** Tetanus toxoid raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.

List of Tables

Chapter 3: Research Paper 1

- **Table 1:** Summary of recombinant STEVOR and RIFIN proteins selected.
- **Table 2:** Summary statistics of the longitudinal cohort PRISM study population.
- **Supplementary Table 1:** Summary of previously established *Plasmodium falciparum* markers of infection exposure complementing the studied STEVOR and RIFIN recombinant proteins.
- **Supplementary Table 2:** Summary table of mean MFI values of PHE negative controls and calculated seropositivity thresholds of PHE mean MFI plus three times standard deviation, per antigen.
- **Supplementary Table 3:** Summary of population seroprevalence per survey year for each recombinant antigen.

Chapter 4: Methods for analysis of STEVOR protein sequences. (Linking Material to Chapter 5)

- **Table 1:** *P. falciparum* 3D7 reference strain STEVOR data base IDs, adjusted IDs, and *k*-mean positions when multidimensional scaling is applied with three and eleven *k*-mean clusters.
- **Table 2:** Parameters summary of sequence clustering methods.

Chapter 5: Research Paper 2

- **Table 1:** Summary of *Plasmodium falciparum* STEVOR protein sequences.
- **Table 2:** Summary of STEVOR V2 library recombinant antigens.
- **Table 3:** Summary of SC recombinant antigens.

Chapter 7: Research Paper 3

- **Table 1:** FIGHTMAL samples characteristics.
- **Table 2:** MASSIV/MATAMAL samples characteristics.
- **Table3:** Logistic regression model analysing the effect of age group, sampling time point and their interaction on seropositivity for each antigen for FIGHTMAL samples, in data format: Z-score (p-value)
- **Table4:** Logistic regression model analysing the effect of age group, intervention arm and their interaction on seropositivity for each antigen for MASSIV/MATAMAL samples, in data format: Z-score (p-value)
- **Table 5:** Summary of MASSIV/MATAMAL samples Chi-square* significant differences in seroprevalence among age-groups, per intervention arm.
- **Table 6:** Summary of MASSIV/MATAMAL samples Chi-square* significant differences in seroprevalence among intervention arms, per age-group.
- **Supplementary Table 1:** Sample proportion seropositive to n number of STEVOR V2 recombinants.
- **Supplementary Table 2:** Seropositivity threshold per antigen according to malaria naïve samples mean MFI plus three standard deviations.
- **Supplementary Table 3:** Antigen specific percentage seropositivity of FIGHTMAL samples stratified by age group and sampling time-point.
- **Supplementary Table 4:** Antigen specific percentage seropositivity of MASSIV/MATAMAL samples stratified by age group and intervention arms.

List of Abbreviations

- *Pf*: *Plasmodium falciparum*
- IE: Infected Erythrocytes
- VSA: Variant Surface Antigens
- RIFIN: Repetitive Interspersed Family of proteins
- STEVOR: Sub-Telomeric Variable Open Reading frame family of proteins
- *rif*: Repetitive Interspersed Family of gene
- *stevor*: Sub-Telomeric Variable Open Reading frame family of genes
- CDC: Centres for Disease Control and Prevention
- DDT: Dichlorodiphenyltrichloroethane
- WHO: World Health Organisation
- ACT: Artemisinin Combination Therapy
- RBCs: Red Blood Cells
- GDV1: Gametocyte Development Protein 1
- PCR: Polymerase Chain Reaction
- RDTs: Rapid Diagnostic Tests
- HRP2: Histidine-Rich Protein 2
- LDH: Lactate Dehydrogenase
- IFA: Immunofluorescent Antibody Assay
- ELISA: Enzyme-Linked Immunosorbent Assay
- LLINs: Long-Lasting Insecticidal Nets
- IRS: Indoor Residual Spraying
- IVM: Ivermectin
- MDA: Mass Drug Administration

- CSP: Circumsporozoite Protein
- Igs: Immunoglobulins
- H: Heavy chain
- L: Light chain
- CDR: Complementarity-determining region
- mAbs: Monoclonal Antibodies
- Rh5.1: Reticulocyte-binding protein homologue 5
- CHMI: Controlled Human Malaria Infections
- NAI: Naturally Acquired Immunity
- *PfMSP1.19*: *P. falciparum* Merozoite Surface Protein 1
- *PfAMA1*: *P. falciparum* Apical Membrane Antigen 1
- EBA175.RII.V: Erythrocyte Binding Antigen 175, region II-V
- Etramp5: Early-Transcribed Membrane Protein 5
- *PfEMP1*: *Plasmodium falciparum* Erythrocyte Membrane Protein 1
- *PfMC-2TM*: *Plasmodium falciparum* Maurer's cleft 2 Trans-Membrane domains family of proteins
- SURFIN: Surface-Associated Interspersed family of proteins
- DBL: Buffy-Like Binding domain
- CIDR: Cysteine-rich Inter-Domain Region
- CD36: Cluster of Differentiation 36
- ICAM1: Intercellular Adhesion Molecule 1
- TLR: Toll-like Receptor
- PRR: Pattern Recognition Receptor
- DC: Domain Cassette
- EPCR: Endothelial Protein C Receptor

- SM: Severe Malaria
- UM: Uncomplicated Malaria
- PNG: Papa New Guinea
- APC: Activated Protein C
- InD: Insertion Deletion element (INDEL)
- TM: Trans-Membrane domain
- SP: Signal Peptide
- V1: small Variable domain
- P: PEXEL trafficking motif
- SC: Semi-Conserved domain
- V2: large Variable domain
- C: Conserved domain
- LILRB1: Leukocyte Immunoglobulin-Like Receptor B1
- LILRB2: Leukocyte Immunoglobulin-Like Receptor B2
- LILR: Leukocyte Immunoglobulin-Like Receptor family
- LAIR1: Leukocyte-Associated Immunoglobulin-Like Receptor 1
- GPC: Glycoprotein C
- WGCF: Wheat Germ Cell Free system
- TSs: Trans-sialidases
- SMV: Soybean mosaic virus

Chapter 1: Background

1.1 Malaria: History, Burden, Diagnosis and Prevention

Malaria is a disease caused by an infection with the *Plasmodium* parasites species, transmitted human to human via the bite of an infected pregnant female *Anopheles* mosquito (1). Out of all 172 recognised species of *Plasmodium*, only six are known to regularly cause disease in humans. These include *P. falciparum*, *P. vivax*, *P. knowlesi*, *P. malariae* and two *P. ovale* (*P.o. curtisi* and *P.o.wallikeri*) species (2),(3).

Malaria is an ancient disease, dating back to 3200 BC, as evidenced by the malarial antigens detected in Egyptian mummy remains (4),(5). Following the miasma theory, which associated the disease with swamps, the name “malaria” originates from Italian: 'mal'aria', meaning “bad air” (6),(7). Historically, malaria has been a global problem, suggested that it arrived in Rome during the first century AD and subsequently spread throughout Europe, marking a significant turning point in European history (8). In the United States of America, malaria claimed more lives during World War II than the war itself and led to the foundation of the Centre for Disease Control and Prevention (CDC) in 1946 (9),(10). Thanks to CDC's National Malaria Eradication Program and the extensive use of the pesticide Dichlorodiphenyltrichloroethane (DDT), the USA was declared malaria-free in 1951 (11). In 1975, Europe was declared malaria-free for the first time, as a result of the World Health Organisation (WHO) Global Malaria Eradication programme. However, the reintroduction of the disease due to political and commercial reasons led to Europe being declared malaria-free again in 2015 (12). In India and China, social factors, such as population growth and migration to more tropical, southern zones, have contributed to a significant burden of vector-borne diseases, including malaria (9). Despite malaria affecting people globally, Africa is the geographical region most strongly associated with the disease, shaping the continent's history and evolution (9). The high prevalence of the disease led to the

selection for the sickle cell trait in vast majority of the African population, which has been proven to be protective against malaria (13),(14). Moreover, historically, malaria has been reported as the biggest obstacle in colonisation of African countries, earning the colonial-era term "The White Man's Grave" by the 18th century (15).

The Global Malaria Eradication programme had a huge impact in reducing the incidence and transmission of malaria globally between the 1900s and early 2000s, reducing the malaria distribution from 53% to 27% of the global land surface (16). In terms of population at risk, there has been a substantial decrease in the global population at risk of all-cause malaria from 77% to an estimated 48% in the 21st century (17). Despite the observed reduction of malaria cases between 2000 and 2015 due to malaria control strategies, there has been a significant increase in malaria cases since 2016 due to several factors such as the rise of drug-resistant parasites and insecticide resistant vectors (18). According to the 2023 WHO World Malaria Report, there were approximately 249 million cases of malaria reported in 84 malaria endemic countries and an estimated 608 thousand deaths in 2022, with 94% of those cases reported in Sub-Saharan Africa (19). This marks an increase of five million cases per year since 2021, mostly accounted by artemisinin-combination therapy (ACT) resistance (19). Difficult socio-economic conditions, political conflicts, humanitarian disasters, population stigmatism and climate change in Sub-Saharan Africa have also greatly impacted the success of malaria intervention campaigns (20),(21). Additionally, the disruption of malaria case management due to SARS-CoV-2/COVID-19 pandemic delayed intervention campaigns and reduced access to adequate treatment, estimated to result in an additional 13.4 million cases (19).

More than 95% of malaria mortality is caused by an infection with *Plasmodium falciparum*, which is also the predominant malarial species found in Sub-Saharan Africa (19). The *Plasmodium falciparum* parasite life cycle is rather complex, taking place in two phases; the sexual stage which occurs in the mosquito host, and the asexual stage in the human host (22).

The discovery of the *Plasmodium* parasite began in 1880, when Charles Louis Alphonse Laveran discovered the first protozoan and linked it with the disease, due to the observation of pigmented protozoa found in red blood cells (RBCs) from patients with the disease but absent in healthy individuals (23). These observations were further confirmed by the work of the Italians, Golgi and Marchiafava, who observed the amoeboid movement of the organism and proved that it infects red blood cells, grows, and reproduces, further infecting fresh red blood cells (24). The discovery of the sexual stage of the *Plasmodium* parasite came later, first in 1897 by the Canadian medical student William MacCallum, who observed flagellated bodies fusing with non-motile bodies to form a vermicle, now referred to an ookinete (25). In the same year, these observations were confirmed by Patrick Manson's student Ronald Ross, who found the malaria pigment in *Anopheles* mosquitoes and further observed that the parasite can grow in the mosquitoes (26). Building on these investigations, the work of Ross, Laveran, MacCallum and the Italians led to the discovery of the malarial parasite transmission to humans via a vector in the form of *Anopheles* mosquitoes. The gathered evidence, summarising those experiments and investigations between the years of 1884 and 1897 is known as 'The mosquito-malaria doctrine' (7). Another breakthrough of great importance is the discovery of the *Plasmodium* parasite culture in human erythrocytes, the culturing of sexual stage parasites, and cultivation of the liver stages by W. Trager and J. Jensen in 1976, which provided the platform to study the development of parasites, parasite genetics, and important molecular pathways without using animal models (27). Finally in 2002, the first *P. falciparum* clone 3D7 complete genome was sequenced, marking the final stage of the discovery of *Plasmodium falciparum* (28).

As illustrated in Figure 1, the *P. falciparum* life cycle begins with the entry of small number of sporozoites injected into the human host along with saliva of a female *Anopheles* mosquito during blood feeding. The sporozoites then travel through the bloodstream to the liver, where they invade the hepatocytes, replicate asexually, forming a schizont, known as the liver stage

of the infection. After approximately 7 days, the schizont ruptures, releasing tens of thousands merozoites that enter the peripheral blood system, known as the infectious stage. Merozoites infect red blood cells, initiating a second asexual replicating cycle, known as the blood stage. Over 48 hours, the ring parasites undergo maturation to trophozoites, followed by replication into merozoites and subsequent formation of schizonts, known as schizogony. Further, the RBCs rupture, releasing the merozoites which are then ready to infect uninfected RBCs. During each blood stage cycle, about 0.1% to 5% of the merozoites, termed gametocyte-committed merozoites differentiate into male and female gametocytes which enter the extravascular space of the bone marrow for maturation and gametocytogenesis. Some of the factors contributing to sexual differentiation of merozoites are parasitaemia, human host body temperature and the expression and upregulation of gametocyte development protein 1 (GDV1) (29). When a pregnant female *Anopheles* mosquito takes up a blood meal and ingests the gametocytes, they mature into gametes and travel to the mosquito midgut (30),(31). There, the male and female gametocytes develop into flagellated microgametes and a macrogamete, respectively, further fertilising, and resulting in a zygote, the sexual stage of the parasite cycle. The zygote undergoes meiosis and leaves the gut, becomes an oocyte, and asexually replicates forming sporozoites. Upon rupturing the oocyst, the sporozoites are released and travel to the salivary glands of the mosquito, ready to be injected into the human host again (11),(31).

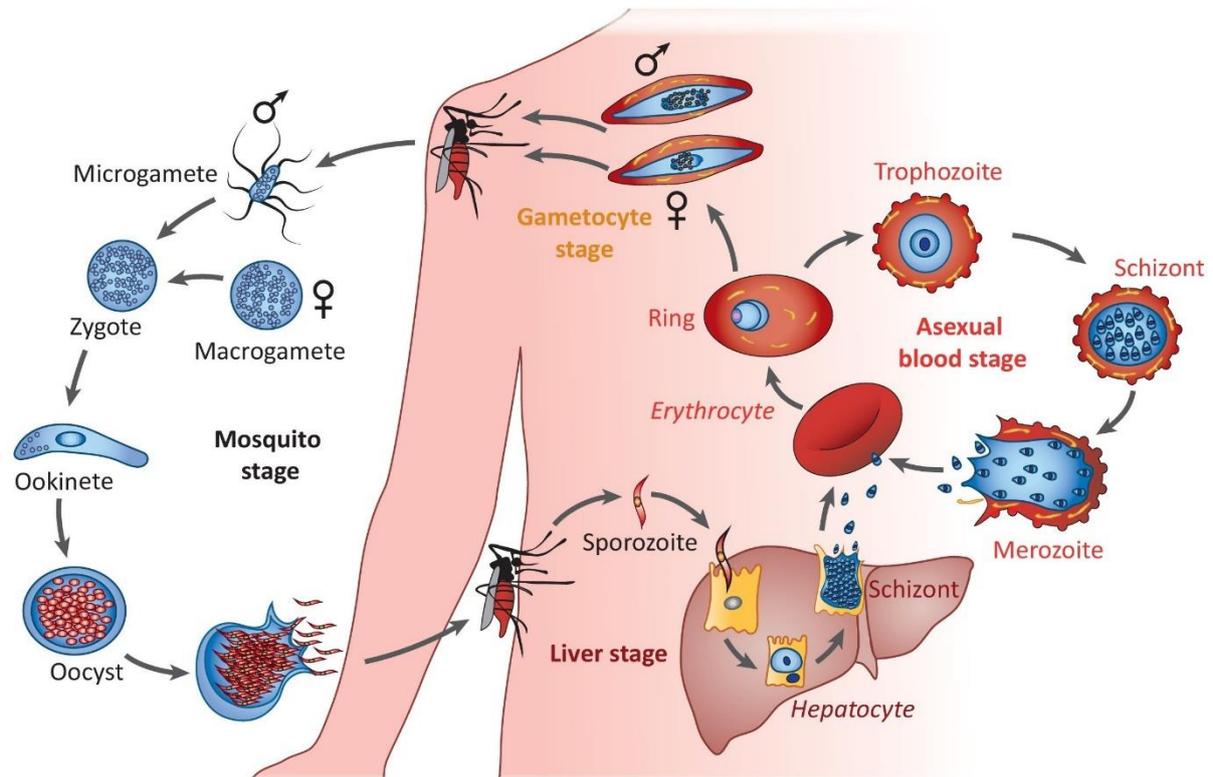


Figure 1: *Plasmodium falciparum* life cycle. Illustration of the complete *P. falciparum* life cycle, highlighting the stages in the human host, and the mosquito vector (31). Figure taken from Maier *et al.* (2019)

The disease clinically presents with a variety of symptoms and degrees of severity, ranging from mild and asymptomatic to severe disease and death (11). In terms of clinical malaria, the most common symptoms include fever, chills, sweats, headaches, body aches, and general malaise (11). Children under 5 years of age, pregnant women, immunosuppressed individuals, and those exposed to the infection for the first time are at higher risk of developing severe malaria (32),(33),(34). School-age children of age five to fifteen years bear the highest burden of asymptomatic malaria irrespective of the infection exposure level, and make up almost 50% of the population at risk of malaria (35). The clinical outcome of the infection is complex and depends not only on host factors such as immunity, genetics, age, and pregnancy, but also on parasitic factors, as well as socio-economic status of the infected individuals. Some pathogenic factors are related to drug resistance, rate of multiplication, alterations of red blood cells, and

antigenic polymorphism (36). Clinical findings must be confirmed using one or more laboratory tests. These include the gold standard method of light microscopy of Giemsa-stained blood slides for detecting the presence of parasites in patients' blood. However, this method is highly dependent on professional expertise, as well as the physical presence of the Giemsa staining chemicals, which are rarely available in Sub-Saharan Africa. Moreover, the sensitivity of the method is not ideal, and a negative blood slide in presence of clinical symptoms does not necessarily indicate the absence of infection (37). Other malaria laboratory measures include direct methods, such as measuring parasitic nuclear acids, a technique also known as Polymerase Chain Reaction (PCR). This method is regarded as more sensitive and specific than all other conventional methods, with detection levels lower than one parasite copy per microlitre (38). However, it is time-consuming and reliant on laboratory expertise and equipment, and is usually used for the confirmation of microscopy results (37),(39). Other typical methods are the indirect measures of infection, such as the detection of antibodies to specific parasitic antigens in the blood of clinical patients and/or asymptomatic individuals (33). Rapid Diagnostic Tests (RDTs) are immunochromatographic tests for the detection of malarial antigens in blood and are currently the WHO-recommended first choice of tests to be used in the field in malaria-endemic areas due to their feasibility, accessibility, and speed (40). Although RDTs can be performed in the field without any specific expertise and equipment, they have shown to have relatively low sensitivity resulting in potential false-negative results due to gene deletion targets found in *P. falciparum* isolates (41). The *P. falciparum* histidine-rich protein 2 (HRP2) is the most common RDT target used, delivering false-negative results due to the fast emergence of *pfhrp2*-deleted parasites. Hence, dual target HRP2 + pan-lactate dehydrogenase (LDH) is now used in the field and offers better results in terms of sensitivity (42). Other indirect serology tests, which measure antibodies against parasitic proteins, include the Immunofluorescent Antibody Assay (IFA) and the Enzyme-Linked Immunosorbent Assay (ELISA), both of which are time-consuming and require significant laboratory expertise and equipment. Consequently,

they are not suitable for large-scale sample testing and are subjective in terms of result evaluation (37),(43).

The choice of malaria treatment also depends on multiple pathogen, host, and socio-economic factors. The first antimalarial drug used was quinine, derived from the bitter bark of the South American Cinchona tree, utilised since the 16th century. In 1820, the quinine compound was isolated from the plant and used worldwide for malaria treatment (1),(9). However, due to poor tolerance, it is not recommended for use in children and pregnant women (44). By the end of World War II, the antimalarial chemicals chloroquine and sulfadoxine-pyrimethamine had been discovered and were widely used to treat malaria, which quickly led to *Plasmodium* drug resistance to both mono-therapies (9),(45). These challenges inspired research into new antimalarial treatments and led to the discovery of another plant-based compound, artemisinin, in China in the early 1970s. Artemisinin, derived from the plant *Artemisia annua*, which has been used against fever and chills since the 2nd century BC, is now part of the artemisinin-based combination therapy (ACT), the WHO-recommended main line of malaria treatment, which ensures a high malaria cure rate while minimising the risk of drug resistance (41).

Another method for reducing the incidence of malaria is through vector control, using pesticides such as DDT, whose anti-mosquito properties were discovered in 1939, although the compound was first synthesised in 1874 (46). However, DDT presents numerous issues, from human and environment toxicity, to mosquito resistance (46),(47). Today, DDT is used in only a few Sub-Saharan African countries in their malaria combat policies (41). Currently, other vector control strategies, such as long-lasting insecticidal nets (LLINs) and indoor residual spraying (IRS), are widely implemented in populations at risk of malaria (48). Another vector control method of a significant interest nowadays is the use of the endectocide Ivermectin (IVM). IVM is a broad-spectrum drug used to control the transmission of roundworms, such as the causative agent of lymphatic filariasis (49). IVM binds to glutamate-gated chloride ion channels in

muscles and nerve cells of invertebrates, resulting in paralysis and death. However, it is safe for humans as it cannot pass the blood-brain barrier (50),(51). IVM is readily absorbed in humans; therefore, if a mosquito feeds on a person treated with IVM, the blood meal is toxic for the mosquito. Thus, mass drug administration (MDA) interventions with IVM have proven to disturb malaria transmission in Sub-Saharan Africa endemic settings (52). Since IVM's half-life is relatively short at 18 hours, the interventions propose IVM MDA of 300µg/kg/day for three consecutive days across three consecutive months. This regiment has been shown to be mosquitocidal for up to 28 days and is predicted to reduce malaria prevalence to as low as 1% in low-endemicity settings, while remaining safe for humans (53),(54). However, recent clinical trials have shown contradictory results regarding the effect of the IVM MDA regiment on reducing *P. falciparum* prevalence, with no significant results in reducing *Anopheles* density between control and intervention arms in a study in the Bijagos Archipelago of Guinea-Bissau (Hutchins H *et. al* unpublished).

In terms of malaria vaccines, RTS,S/AS01 is the first vaccine launched in three Sub-Saharan African countries: Ghana, Kenya, and Malawi. RTS,S has shown an efficacy of about 50% with a single dose administration, reducing to 40% after a booster dose 12 months later (55). The second WHO-approved vaccine in 2023, the saponin-adjuvanted recombinant vaccine R21/MM, has demonstrated a protection of around 75% in children during the first year, following three doses of administration, with detected reduced protection of 71% after booster dose 12 months later, in combination with LLINs interventions (56),(57). Thus, current vaccine candidates show promise, but there is still a need to understand the reasons for the heterogeneity in their protective effects and the waning of immune responses. Both WHO-approved vaccines are based on the central repeat region of the circumsporozoite *P. falciparum* protein (CSP) as the vaccine target containing T-cell epitopes and B-cell epitope (R21/MM) to stimulate an antibody response in recipients to protect them from the infection (56),(55). The main

difference between the vaccines is the adjuvant being AS01 in RTS,S and Matrix-M a smaller saponin adjuvant associated with stimulation of antibody production and cellular immune response in R21 (55),(56). The central repeat region encoded by the *cs* gene, contains regions such as Th2R and Th3R encoding for epitopes recognised by T-cells. These regions are also associated with high polymorphism due to non-synonymous SNPs and the proportion of epitope diversity is found to be higher in populations characterised by high malaria transmission, compared to those with lower (58). R21/MM has shown to have higher initial protection against clinical disease compared to RTS,S in high transmission settings (55),(56). The transmission intensity and therefore level of exposure to the parasite, coupled with host factors, such as age, previous exposure of *P. falciparum*, as well as parasite genetic diversity, are some of the factors contributing to the heterogeneity of efficacy of the malaria vaccines.

Since both vaccines contain parts of a pre-erythrocytic stage *P. falciparum* antigen and no parasite blood stage antigen, there could potentially be no stimulation of immune response to help combat the infection of non-vaccine CSP variants that manage to progress to the blood stage of infection. Thus, due to the complexity of the parasite's life cycle and its antigenic variation, the approach of using one antigen for one vaccine might not be the most appropriate for achieving long-lasting protection (36),(59).

1.2 *P. falciparum* serology as a tool for understanding transmission, protection, and infection exposure.

B-cells are a type of leukocytes that when activated secrete highly specialised molecules known as immunoglobulins (Igs). These molecules recognise pathogenic proteins, or antigens, with high specificity. Upon recognition, the terminally differentiated B-cells produce antibodies, which are the soluble, secreted version of those same antigen-specific Igs. The antibodies, in turn, bind to the antigenic proteins of pathogens circulating in the extracellular space and recruit

phagocytic cells and molecules, targeting the pathogen for destruction. Antibodies can also bind toxins, preventing their action, a process called neutralisation

Antibodies are composed of two key regions: the variable (V) regions, located at the ends of the Y-shaped arms, which are responsible for antigen binding, and the constant (C) region in the stem, which interacts with effector cells and molecules. All antibodies, known as immunoglobulins, are made from paired heavy and light chains. The five classes of immunoglobulins—IgM, IgD, IgG, IgA, and IgE—are distinguished by their C regions, while antigen specificity is governed by the V regions. In this context, the focus is on IgG antibodies in response to *Plasmodium falciparum* infection.

IgG antibodies are large, approximately 150 kDa, composed of two identical heavy (H) chains, each about 50 kDa, and two identical light (L) chains, each 25 kDa. These chains are linked by disulfide bonds, creating two antigen-binding sites. The structure of the immunoglobulin domains in both heavy and light chains consists of two β sheets linked by a disulfide bond, forming a β barrel shape. The main difference between V and C domains is that V domains are slightly larger due to an extra loop that contributes to antigen binding. Stability of these immunoglobulin domains is maintained by amino acids that are common to both V and C regions. The light chains can be either kappa (κ) or lambda (λ), without functional differences between them. IgG, the most abundant immunoglobulin, has four subclasses (IgG1, IgG2, IgG3, IgG4), with each subclass's functional properties determined by the structure of the heavy chain's carboxy-terminal region. Both heavy and light chains show repeating domains of approximately 110 amino acids, forming compactly folded structures called protein domains. The light chain has two such domains, while the heavy chain of IgG has four. The amino-terminal ends of these chains vary significantly, comprising the V domains that are responsible for antigen specificity, while the rest of the chains consist of C domains, which remain conserved among antibodies of the same isotype.

An antibody's overall structure includes three globular regions connected by a flexible hinge, which allows for movement of the two antigen-binding arms. Each arm, or Fragment antigen-binding (Fab), is composed of a light chain and the amino-terminal half of a heavy chain, while the stem, or Fragment crystallizable (Fc), consists of the carboxy-terminal halves of the heavy chains. The Fab region binds antigens, whereas the Fc region interacts with immune effector molecules. The flexibility provided by the hinge region and the junction between V and C domains allows the Fab arms to move independently and bind antigen sites that vary in distance, such as those on bacterial cell walls. The V regions differ between every antibody molecule with sequence variability concentrated in three hypervariable regions, found in the light and heavy chains of the molecule. Those hypervariable regions are flanked by framework regions. The hypervariable regions are found on the edge of the beta barrel juxtaposed in the folded domain, thus the hypervariability is concentrated and localise on the surface of the molecule. The binding site of the antibody molecule is formed when the V regions of the heavy and light chains are brought together creating a single hypervariable site, termed the complementarity-determining regions (CDR) which determine the antigen specificity of the molecule. The immune system generates antibodies with different antigen specificities by combining different light and heavy V regions known as combinatorial diversity. The different CDR are made of regions of different amino acid sequences, resulting in different shapes of the antigen binding surfaces. This adds another level of complementary interaction between antibodies and antigens. Additional to the sequence and shape complementary, antibody-antigen interactions are size dependent, where small antigens and peptides bind at the groove of the CDR and larger antigenic protein molecules sometimes bind more than one, or all of the CDRs, or even other parts of the antibody. The final level of antibody antigen binding are the interaction forces. Electrostatic interactions, hydrogen bonds, van der Waals forces and hydrophobic interactions can all contribute to the binding. The contribution of interaction forces to the antibody-antigen binding depends on the specific antigen and antibody. The strength of this binding is referred

to as antibody avidity. The antibody complementary binding regions of antigens are called antigenic determinant or epitopes. Some epitopes, termed conformational are formed from different parts of a protein via protein folding, and others composed by single segments are called linear. (60)(61).

Upon infection with a foreign agent, IgM antibodies are the first to be generated by the B-cells, in small proportions and with low affinity. However, the pentameric structure of the molecule provides it with high avidity when binding to multivalent antigens. Due to its size, IgM is mainly found in the bloodstream rather than in lymph, hence it is important in controlling blood infections (61).

IgG is the main isotype of antibodies in the blood and extracellular fluid, present in the circulation of newborn babies through placental transfer from the mother. As IgG antibodies can easily diffuse to the extracellular space and bind with high affinity to specific antigens, they are regarded as the principal neutralising antibodies. Moreover, different subclasses of IgG are present in varying abundance, with the IgG4 subclass being the least abundant. Similarly, IgA antibodies, primary found in mucosal surfaces, bind antigens with high affinity and possess neutralising properties (61). Generally, the IgM humoral response is induced early in infection, but decays quickly and is replaced by IgG, a pattern demonstrated to be true for multiple infectious diseases (62),(63).

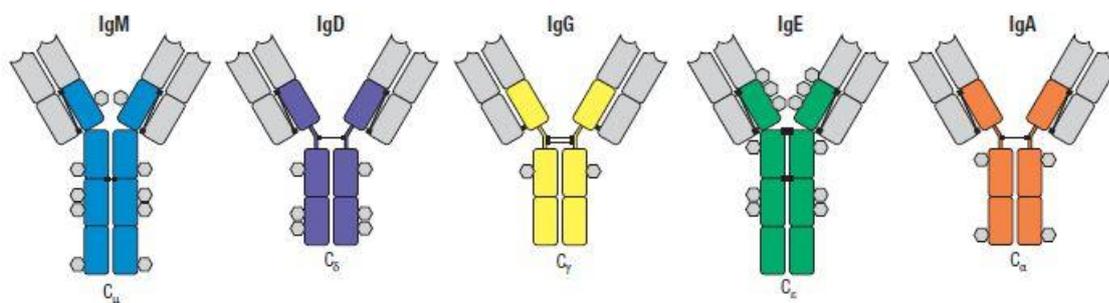


Figure 2: Figure illustrating the general structure of the main immunoglobulin isotypes. Each rectangle represents a domain, and the C-region, which determines the isotype is illustrated as the main colour for each isotype. (Figure taken from Janeway's Immunobiology 7th edition (2017) (61))

Protective IgG antibodies against *Plasmodium falciparum* merozoite proteins play a crucial role in preventing parasite replication and disease progression and are strongly associated with protection from symptomatic malaria. For instance, MSP1 polyclonal antibodies, which target multiple epitopes of the merozoite antigen, primarily inhibit merozoite invasion of erythrocytes and disrupt intracellular parasite development. While these antibodies do not prevent the rupture of 3D7 schizonts, they agglutinate merozoites and arrest young parasites at the early trophozoite stage, thereby exerting both invasion- and growth-inhibitory effects. Furthermore, fluorochrome-labelled anti-MSP1 antibodies demonstrate access to intra-erythrocytic parasites, enabling them to inhibit maturing schizonts and ultimately prevent their rupture and release as infectious merozoites (64).

Similarly, antibodies against AMA1 interfere with erythrocyte invasion by blocking AMA1's proteolytic processing and surface redistribution. Polyclonal antibodies inhibit the secondary processing of AMA1, preventing the conversion of its 66-kDa fragment into smaller functional fragments essential for invasion. These antibodies also induce anomalous processing, resulting in non-functional fragments, and inhibit the circumferential redistribution of AMA1 on merozoites, disrupting their ability to attach to and invade erythrocytes (65).

Overall, antibodies targeting specific merozoite surface antigens, such as MSP-119, MSP-3, AMA1, and GLURP, significantly reduce the risk of symptomatic *P. falciparum* malaria by targeting critical processes in the parasite's life cycle. Meta-analyses reveal that individuals with IgG responses to these antigens have a lower risk of symptomatic malaria, with MSP-119 and MSP-3-Ct providing 18% and 54% reductions in risk, respectively, compared to non-responders. This protection was found to be dose-dependent, with higher antibody levels

correlating with greater reductions in malaria incidence. However, protective effects vary depending on the antigen, allele, and study population, influenced by factors such as antigenic diversity and methodological heterogeneity. These findings highlight the significant contribution of merozoite-targeting antibodies to acquired immunity against malaria, supporting their potential as key targets for vaccine development (66).

Moreover, long-acting monoclonal antibodies (mAbs) against specific *P. falciparum* antigens have been studied in terms of therapeutics and potentially as protective vaccines, with individual targets developed against every stage of the infection. An example of mAbs targeting the merozoites are various identified antibodies that target the reticulocyte-binding protein homolog 5 (Rh5.1), resulting in more than 90% merozoite invasion inhibition *in vitro* (67). Other mAbs designed to tackle the pre-erythrocytic stage, targeting L9LS, a conserved junction epitope in the CSP protein, have shown that low doses of monoclonal anti-L9LS provide protection against the diseases in individuals from controlled human malaria infections (CHMI) (68).

Although specific immunity against malarial infection is possible, and the concept of protective antibodies against malaria has been known since the 1930s, there is still no evidence for sterile immunity in humans (69),(70). Furthermore, studies have shown that an abundance of IgM antibodies against merozoites can also be associated with protection in populations where malaria is endemic. IgM antibodies targeting *Plasmodium falciparum* merozoite antigens, including PfMSRP5, PfSERA9, PfRAMA, PfCyRPA, PfRH5, MSP2, MSP1, AMA1, and EBA175, play a critical and multifaceted role in malaria immunity (71), (72). IgM responses are rapidly induced during acute malaria, particularly in individuals with high parasitaemia, and are significantly more prevalent in children with malaria compared to those without (72). While IgM levels often decline more rapidly than IgG, suggesting production by short-lived plasma cells or B1 cells, its presence in clinically immune adults hints at potential long-term memory

or persistent production (71). Functionally, IgM is highly effective in controlling parasitaemia through complement activation, fixing components like C1q, C3b, and C5b-C9, and promoting merozoite lysis. Notably, IgM demonstrates superior complement-activating efficiency compared to IgG despite its lower avidity (72).

In malaria-endemic regions, high IgM levels are associated with significant reductions in the risk of clinical malaria, comparable to IgG, and both antibodies correlate strongly with complement-fixing activity, a key mechanism of protection (72). While IgM responses are often transient and sometimes weakly associated with higher parasitaemia, their rapid induction may play a pivotal role in early immune defence by reducing parasite replication, facilitating IgG production, and balancing other immune mechanisms such as phagocytosis (72), (71). These findings challenge the traditional view of IgM as a transient, primary infection response and highlight its durability and functional significance in naturally acquired immunity. Together, IgM and IgG form a complementary immune response that effectively reduces parasitaemia and prevents severe disease (72), (71). Finally, IgM and IgG subclasses (IgG1 and IgG3), which are associated with protection, have been shown to exhibit a similar decline in antibody titres over time following acute infection in individuals from *P. falciparum* endemic settings (73),(72).

1.3 Naturally acquired immunity to *Plasmodium falciparum* malaria disease.

Naturally acquired immunity (NAI) to malaria refers to the ability of individuals to develop adaptive immunity against the malaria disease. This immunity can be actively acquired after multiple infections or passively through the transfer of protective antibodies via the placenta or breastfeeding, with levels declining after six months of age (74),(75). NAI develops slowly its mechanisms not entirely understood. NAI is characterised by two types: anti-parasitic NAI, which is the ability to limit parasite growth and maintain parasite density below a specific

threshold, and clinical immunity, which allows individuals to tolerate considerable parasite load without showing disease symptoms (76).

NAI is age and exposure dependent, two interlinked concepts. Thus, individuals from areas with high malaria endemicity, as well as adults, are expected to develop a stronger NAI compared to individuals from low infection exposure settings, and particularly children. (77),(78),(79). Children born in malaria endemic areas from six months to five years of age experience the highest burden of severe malaria prevalence and disease mortality, a heightened vulnerability due to the immaturity of their immune system compounded with the waning of maternal anti-malarial antibodies (80).

However, data from studies on nonimmune populations moving to endemic areas, such as Irian Jaya, Indonesia, highlight the significant role of age in determining the risk of severe malaria. Among nonimmune migrants exposed to *Plasmodium falciparum* for the first time, adults were at a substantially higher risk of severe disease compared to children (81), (82). For instance, during the initial six months of exposure, adults experienced severe events requiring hospital evacuation at an incidence density of 1.34 events/person-year in the third month, compared to 0.25 events/person-year in children, yielding a relative risk of 4.51 (95% CI = 1.94–11). Additionally, 23.2% of adults were hospitalized for severe malaria compared to only 8.6% of children (relative risk = 2.7, 95% CI = 1.9–3.8). This increased susceptibility in adults during primary exposure contrasts with their ability to acquire immunity to chronic exposure more rapidly than children, as evidenced by lower parasitaemia and fewer symptoms after 1–2 years of residence in hyperendemic regions (81).

The historical records from British and French soldiers further reinforce this confounding interplay. Soldiers exposed to *P. falciparum* in regions like West Africa developed immunity that protected them in other malaria-endemic areas, whereas those from non-endemic areas like

the British Isles did not. This suggests that immunity is shaped both by age and the timing of exposure during childhood, complicating attempts to disentangle their effects in endemic populations (83).

These findings underscore the confounding relationship between age and exposure in endemic malaria studies. Age-related immune system changes, such as the shift from naive to memory T cells and thymic involution, may result in adults being more prone to maladaptive responses during acute infections while benefitting more from chronic exposure. This contrasts with children, whose immune systems are more naive and better equipped to handle primary infections but acquire immunity more gradually over time (82). Together, these observations emphasize the need for careful interpretation of age and exposure when studying malaria immunity, as their combined influence can obscure the true dynamics of susceptibility and protection.

Children from five to fifteen years of age are the age group characterised with the most exposure to malaria infection and act as malaria transmission reservoir (74),(84). Although NAI can reduce malaria incidence and risk of mortality, it has not been shown to provide sterile protection due to various factors, including heterogeneity of the disease, the complex plasmodium life cycle, a large repertoire of antigens, diverse *P. falciparum* strains, and intrinsic host characteristics along with environmental factors (76),(85).

Studies investigating protective immunity to *P. falciparum* infection have shown that *P. falciparum* NAI is associated with antibodies against multiple parasitic antigens, rather than a single specific target (86),(87),(88). Furthermore, approximately 30% of all *P. falciparum* proteins are recognised by the human immune system, with some showing higher serorecognition than others (89). Amongst these proteins, a few such as Rh5.1, *P. falciparum* merozoite surface protein 1 (*PfMSP1.19*), *P. falciparum* apical membrane protein 1 (*PfAMA1*),

Erythrocyte binding antigen 174 (EBA175.RIII.V) and CSP, have been prominent candidates in vaccine development due to their immunogenicity and association with protection against clinical disease, with the latter being the target in both WHO-approved vaccines today (56),(70),(89),(90),(91).

In population-level studies, antibody detection platforms that are rapid and have high specificity and sensitivity are essential for the monitoring distribution of pathogens, referred to as transmission, and can potentially be used to date the occurrence of initial infections (92),(93). Combining serological measures with parasite prevalence data, obtained using microscopy or molecular methods, have shown to be effective in estimating malaria transmission in populations over time (94). Furthermore, analysing antibody responses to malaria in children under the age of five provides deeper insights into the transmission dynamics and antibody kinetics within the population, as their naïve immune system responds rapidly to new infections and subsequent clearing of infection (94),(95).

Antibodies to specific *P. falciparum* antigens decline at different rates over time, without the presence of reinfections (94),(96). This variance determines the kinetics and half-life of the antibodies and depends on multiple factors such as the nature of the antigen stimulating antibodies production, the individual's age and genetics, seasonality, the immunoglobulin classes, and levels of exposure to malaria (94),(96),(97). During this project, antigens that stimulate short and long half-life antibody responses will be termed short-term and long-term markers of exposure or seroincidence, respectively. It has been shown that single measures of antibody responses to long-term markers, such as *PfMSP1.19* and *PfAMA1*, can offer insights into parasite transmission over an extended period of time, and can be used to explore short-term variations looking at data from children under five years of age (94). Conversely, short-term markers, such as the Early transcribed membrane protein 5 (Etramp5), can be used to

investigate more gradual changes in *P. falciparum* exposure and assess exposure at an individual level (94),(98),(87).

Protective antibody responses during the blood stage of the *P. falciparum* infection can be characterised into two groups based on the of antigens targeted. The first group includes merozoite antigens such as *PfAMA1* and *PfMSP1.19*, which help prevent merozoite invasion of RBCs through mechanisms like neutralisation, opsonic phagocytosis and complement activation. The second group consist of antibodies to variant surface antigens (VSA) expressed on the surface of the infected erythrocytes (IE); antigens involved in cytoadherence of the IE and characterised by antigenic variation (99).

1.4 Variant surface antigens

During the blood stage of infection with *P. falciparum*, members of several protein families are translocated from the parasite and expressed on the surface of the IEs, as illustrated in Figure 3 (100),(101). These protein families, characterised by hypervariability and antigenic properties, are collectively termed Variant Surface Antigens (102). The VSA protein families include *P. falciparum* erythrocyte membrane protein 1 (*PfEMP1*) family, the Repetitive Interspersed family (RIFIN), the Sub Telomeric Variable Open-Reading frame family (STEVOR), the *P. falciparum* Maurer's cleft Two Transmembrane protein family (*PfMC-2TM*) and the Surface-Associated Interspersed protein family (SURFIN) (103). Members of these families are mostly conserved across their domains within a family but possess variable domains that are associated with their antigenic properties. This variation occurs within members of the same protein family, between different isolates, and across parasites in the same infection; hence they are also known as *P. falciparum* hypervariable surface-expressed protein families (102). The most studied of these is *PfEMP1* (104).

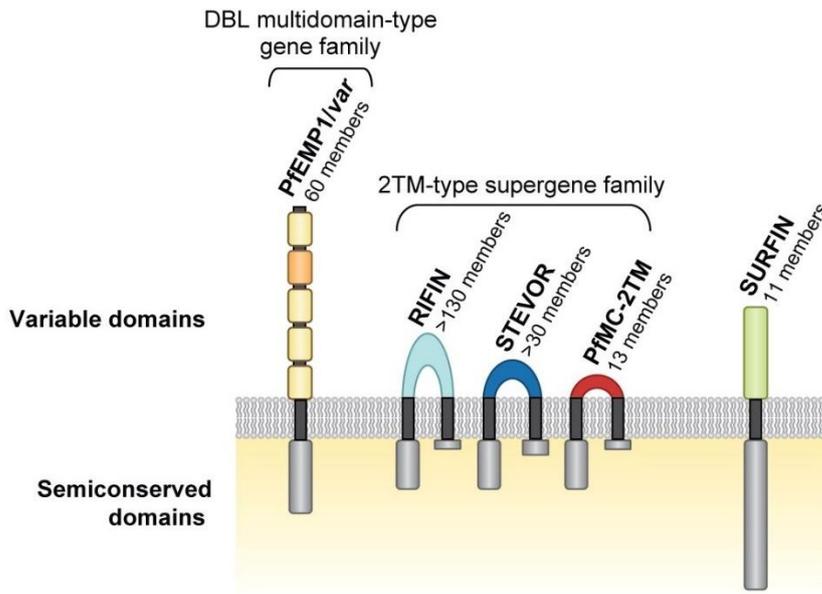


Figure 3: Illustration of the proposed localisation and 2TM topology of hypervariable surface proteins on the membrane of *P. falciparum* infected erythrocytes. Figure taken from Scherf *et. al.* (2008) (102)

PfEMP1 is a family of 200-350 kDa proteins encoded by the multigene family *var*. These genes have two exons, one coding for conserved intracellular domain and another for a large hypervariable extracellular region of two to four Duffy-binding like (DBL) domains and cysteine-rich interdomain regions (CIDR), which function as adhesins, and determine the different variants in the family. Each parasite encodes about 60 *var* genes and expresses one at any given time (105). Despite the large variability, *PfEMP1* proteins are classified in three groups: A, B and C, depending on their gene localisation on the parasite's chromosomes, and are subclassified according to their domain cassettes (DC), which are tandem arrangements of domains. Group B and group C *PfEMP1* CIDR and DBL domains, bind Cluster of Differentiation 36 (CD36) and Intercellular Adhesion Molecule 1 (ICAM1) receptors on endothelial cells, respectively, a process known as cytoadherence (106). Notably, these receptors are also expressed on many human immune cells, including dendritic cells (DCs), implicating the *PfEMP1* involvement in immune modulation (107).

PfEMP1-mediated cytoadherence to DCs inhibits their maturation by preventing the upregulation of key maturation markers such as HLA-DR, CD40, CD80, CD83 and CD86, which are essential for efficient antigen presentation and T-cell activation. This inhibition is specific to cytoadhering parasites. Electron microscope reveals that *PfEMP1*-expressing iRBCs establish close interactions with DCs, disturbing their function without causing cytotoxic damage. This interaction suppresses both primary and secondary immune responses, contributing to the immune dysregulation observed during malaria infection (107).

In addition to impairing DC function, *PfEMP1* variants such as VAR2CSA suppresses key transcription factors, such as NF- κ B, CREB and GAS/ISRE, thereby downregulating early proinflammatory signalling pathways in macrophages. VAR2CSA also modulates innate immune pathways, including Toll-like receptor (TLR) signalling independent of the classical receptors like CD36 and ICAM-1. This suggests *PfEMP1* interacts with other host surface molecules that regulate pattern recognition receptor (PRR) signalling. Additionally, *PfEMP1* has been shown to downregulate the expression of cytokines and chemokines, such as IL-6 and TNF, in monocytes. However, it does not affect the phagocytic activity of monocytes, indicating that its primary role is in regulating inflammatory signalling rather than directly influencing antigen clearance. (108). Specific *PfEMP1* Group A proteins, such as DC8 and DC13 variants, are associated with the binding of the IEs to non-infected erythrocytes, a process known as rosetting (106). These proteins are predominantly expressed in parasites infecting malaria naïve individuals and those presenting with severe malaria, with their domains binding to a different receptor: the endothelial protein C receptor (EPCR). In addition to rosetting, EPCR binding is associated with severe malaria because the receptor's normal function is to promote the activation of protein C (APC), which regulates blood clotting and endothelial barrier properties (109),(110),(111),(112). Thus, changes of EPCR levels are related to blood-brain barrier dysfunction.

In an infection, each parasite expresses one *PfEMP1* variant at a time but is associated with switching of expression of variants over time, selecting for dominant variant to be expressed by most or all of the parasites *in vivo* and *in situ* (113),(114). Controlled human malaria infection studies (CHMI) show controversial results regarding the rapid expansion of and selection for severe diseases-associated variants from Group A *PfEMP1* in malaria naïve individuals and in those with lifelong exposure (115),(116). A CHMI study looking at malaria naïve individuals, showed that an expected severe diseases-associated Group A and DC8 *PfEMP1* variants did not preferentially expand over six replication cycles. Instead, host-intrinsic immune variability shaped the outcomes, with inflammatory responses correlating with symptoms but not parasite replication rates or specific *PfEMP1* expression (115). Conversely, a CHMI study conducted in the malaria endemic Gabon found that in lifelong malaria-exposed individuals, pre-existing immunity significantly shaped *PfEMP1* expression, driving clonal selection of less virulent B- and B/C-type variants while suppressing A-type variants through acquired antibodies. In malaria-naïve individual controls, however, A-type variants were more prominently expressed, potentially reflecting their role in severe disease (116). These studies highlight how host immunity—whether absent or acquired—differentially influences *PfEMP1* expression and disease outcomes, leading to contrasting interpretations of parasite immune evasion and virulence mechanisms.

Antibodies targeting *PfEMP1* play a critical role in protecting against severe malaria (SM), as evidenced by multiple studies. *PfEMP1* is the dominant target of naturally acquired immunity, with antibodies specifically directed at *PfEMP1* being significantly higher in children with uncomplicated malaria (UM) compared to those with SM (117). These antibodies are not only associated with reduced odds of SM but also contribute functionally to immune protection. For instance, *PfEMP1*-specific antibodies mediate opsonic phagocytosis, enhancing the clearance

of infected erythrocytes (IEs) by immune cells, and are particularly effective against virulent *PfEMP1* variants associated with SM (118), (117).

Studies using protein microarrays in Papua New Guinea (PNG) children revealed that antibodies to a limited and conserved subset of *PfEMP1* variants, particularly Group 1 and Group 2 DBL α domains, are strongly linked to protection against SM. These variants are associated with endothelial protein C receptor (EPCR) binding, a mechanism implicated in severe disease pathogenesis, and confer a reduced risk of SM (119). Children who experienced SM initially lacked antibodies to these protective variants but acquired them after exposure to severe-disease episodes. In contrast, children who remained free of SM had higher baseline levels of these antibodies, underscoring the importance of early acquisition of antibodies to specific *PfEMP1* variants in mediating protection (119), (120).

The sequential acquisition of antibodies reflects a structured exposure to *PfEMP1* variants during early infections. This pattern is shaped by the parasite's expression of *PfEMP1* variants with efficient sequestration phenotypes, which confer a selective growth advantage in non-immune hosts. Over time, as immunity develops, subsequent infections involve variants with less efficient adhesion phenotypes, contributing to the broadening of the antibody repertoire. This structured acquisition is crucial for both early protection against SM and the long-term development of immunity (118), (120).

Despite these findings, the association between *PfEMP1*-specific antibodies and protection warrants critical evaluation. While robust evidence links antibodies to Group 1 and Group 2 DBL α domains with reduced SM risk, the relationship between antibody levels and protection against clinical malaria is less definitive. Protection against clinical malaria involves a broader and more diverse range of antibodies, and cumulative exposure (e.g., molecular force of blood-stage infection) is often a stronger predictor of clinical outcomes than specific antibody

responses (119). Furthermore, the protective efficacy of these antibodies varies depending on antigenic diversity and the functional roles of different *PfEMP1* domains, raising questions about their universality across diverse malaria-endemic regions (118), (120).

CHMI studies provide valuable insights into these dynamics, demonstrating that individuals from high malaria transmission areas exhibit a broader and higher pre-existing antibody response to diverse *PfEMP1* domains compared to those from lower transmission regions. This breadth of antibodies, particularly to the variant surface antigens (VSA) expressed on infected red blood cells, is strongly associated with protection. A wide antibody repertoire correlates with reduced parasite growth and a lower likelihood of reaching clinical treatment thresholds. Although specific *PfEMP1* domains, such as CIDR α 3.2 and CIDR α 4(a), have been identified as predictors of protection, their associations lose significance when adjusted for antibody breadth, underscoring the critical role of cross-reactivity across multiple *PfEMP1* variants. This broad immune response likely reflects recognition of a wide range of antigens expressed on IE, potentially capturing both known and unidentified protective targets (121).

Evidenced by these findings, *PfEMP1* is recognized as a crucial target of naturally acquired immunity and a potential candidate for vaccine development. However, the variability in antibody acquisition, the interplay between exposure and immune development, and the immense diversity of *PfEMP1* variants underscore the complexity of translating these observations into universally effective interventions. To address these challenges, future efforts must consider not only the role of individual *PfEMP1* domains but also the broader context of cumulative exposure and cross-reactive immunity, which are essential for sustained protection against malaria.

1.5 STEVOR and RIFIN: structure, function, and localisation

Hypervariability of members of STEVOR and RIFIN protein families is achieved through recombinant events of the genes encoding them, *stevor* and *rif*, respectively, found on the subtelomeric regions of most chromosomes of *P. falciparum*, adjacent to the telomer, while some *rif* genes are found clustering on central regions of the chromosomes. These genes are characterised by highly repetitive heterogenous sequences that are difficult to map and sequence (122),(123),(124),(125). STEVOR and RIFIN proteins appear to be evolutionary related, as they share sequence and structural similarities (125). The STEVOR protein family consists of approximately 30 to 40 protein members, varying from strain to strain, each with a molecular weight of around 40 kDa. These proteins are expressed later in the parasite's life cycle than the products of *var* and *rif* genes, detected approximately 22 hours after infection during the trophozoites and schizonts stages of the parasite's asexual life cycle (126),(127). Some STEVOR proteins localise at the apical end of merozoites, while others are expressed in gametocytes during the sexual stage of the parasite's life cycle (128),(126). The STEVOR proteins expressed on merozoites participate in a secondary transcriptional cascade, producing variants different from those in the schizonts. These proteins are highly expressed in clinical isolates but are not expressed in laboratory-adapted strains (126).

The RIFIN family is considerably larger, comprising of approximately 150 to 200 members per parasite genome, divided into two groups: RIFIN-A and RIFIN-B, distinguished by the presence or absence of a 25 amino acids long insertion-deletion sequence: INDEL (InD), respectively (129). The two RIFIN subtypes also differ in terms of localisation. Similar to STEVORs, type A RIFINs are exported via Maurer's cleft to the membrane of the infected RBCs, whereas type B RIFINs are mainly found inside the parasite (130). Expression of *rif* genes is detected in all stages of the parasite lifecycle including schizonts, merozoites, sporozoites and gametocytes (131),(132).

STEVOR and RIFIN protein families are both part of a larger superfamily known as 2 trans-membrane (2TM) *P. falciparum* protein superfamily (133),(134). Proteins from this superfamily are released from the parasite-derived membranous structures Maurer's cleft into the erythrocyte cytoplasm and localise on the red blood cell membrane (133),(135). As the name suggests, it is structurally proposed that all these proteins possess two trans-membrane domains, exposing a loop of the proteins to the extracellular space, as shown in Figure 4B (102). However, more recent studies have shown that some members of STEVOR and RIFIN protein families possess a single trans-membrane domain topology, where one of the TM domains is referred to as a hydrophobic segment of the protein, and thus not only the variable domain is exposed to the host immune system, but the majority of the protein except of its conserved N-terminus (134),(136),(101). For this work we accept that both protein topologies are possible within members of the protein families, specifically for the STEVOR family, for which the majority of the work in this thesis is focused on (134).

The proposed 2TM protein structure of members from both families, illustrated in Figure 4A, starts with an N-terminal signal peptide (SP), responsible for protein translocation to the endomembrane system for protein folding and post-translational modification, further influencing the mature protein's localisation (137). This is followed by a small variable domain (V1) and a five amino acid long *Plasmodium* export element, or PEXEL motif (P), which codes for protein trafficking from the parasite to the surface of the IE (138),(139),(140). All proteins possess a semi-conserved domain (SC), genetically different between STEVORs and RIFINs and much shorter in RIFIN type A, compared to type B. This is followed by a second large variable domain (V2), wrapped around two trans-membrane domains (TM), and ends with a short conserved domain (C) at the C-terminal of the protein (138).

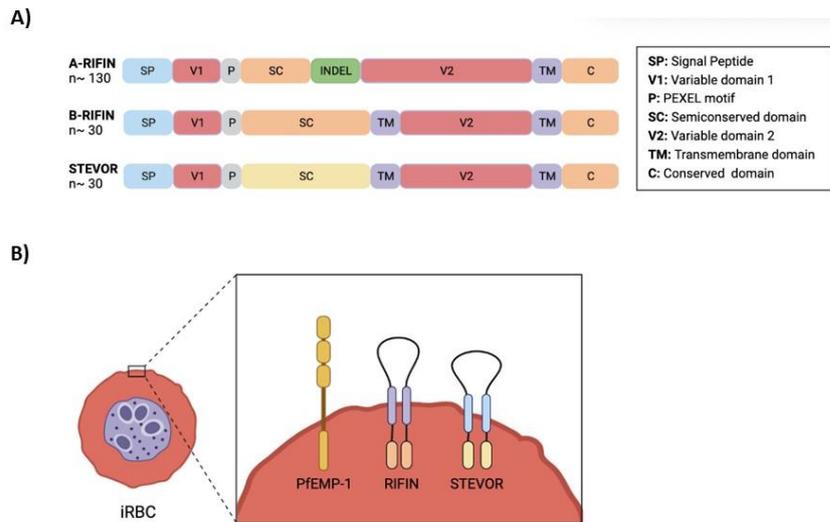


Figure 4: General protein structure of RIFIN and STEVOR protein family A) protein domains and number of copies and B) protein localisation on the IEs. Same colours represent domains with high genetic similarity between the families. The red colour represents domains with high variability (created by Vasileva H. using Biorender.com and based on Zhou *et. al* (2029) (138)).

Functions identified associated with these protein families, important for host cell survival, include immune evasion, rosetting, cytoadherence, alternation of infected host cell rigidity, and some STEVOR proteins are also associated with merozoite invasion (127),(134),(141). Members of the RIFIN protein family, have been shown to interact with host immune inhibitory receptors such as the leukocyte immunoglobulin-like receptor B1 and B2 (LILRB1, LILRB2), , and the leukocyte-associated immunoglobulin-like receptor 1 (LAIR1).

Recent studies reveal that host-derived receptor-based antibodies, incorporating either LAIR1 or LILRB1 domains, target RIFINs as part of an adaptive countermeasure against this immune evasion (145), (146). Broadly reactive antibodies with LAIR1 inserts in their variable heavy chains have been identified in malaria-exposed individuals. These antibodies carry highly mutated LAIR1 domains that retain binding specificity for RIFINs. Structural analyses show

that these antibodies target conserved regions within the V2 apex of certain RIFINs, facilitating immune processes like opsonization and phagocytosis (145). Similarly, antibodies containing LILRB1 domains, particularly the D3 subdomain, bind a distinct subset of RIFINs with high specificity (146). These interactions highlight the polymorphism and redundancy of RIFINs in targeting inhibitory receptors, allowing the parasite to effectively evade host immune responses.

Structural studies further highlight the mechanisms underlying these interactions. For LAIR1-binding RIFINs, specific conserved residues in the V2 domain, such as C254-C265 and R268, are crucial for binding. The interaction involves a highly complementary interface formed by hydrophobic and hydrogen-bonding interactions (147). Comparative analyses demonstrate that LAIR1 and LILRB1 interact with RIFINs through distinct binding footprints, emphasizing the evolutionary diversity of these interactions across multiple *Plasmodium* species. This immune evasion strategy not only suppresses the activation and maturation of immune cells, including natural killer cells, dendritic cells, monocytes, and macrophages, but also downregulates B-cell activity, contributing to immune suppression and severe malaria outcomes (142),(143),(144).

Less is studied and discovered about members of the STEVOR family, although they are found to bind to Glycophorin C (GPC) on uninfected erythrocytes, enabling the process of rosetting, a mechanism which alters the RBC rigidity, and protects parasites from the immune system (148), (149). While direct evidence of rosetting completely blocking phagocytosis remains limited, several lines of experimental and indirect evidence support its role in immune evasion. First, the inclusion of uninfected RBCs in rosettes masks VSAs expressed on iRBCs, thereby reducing antibody and immune cell recognition (150). Second, rosetting through receptors such as CR1 and CD36 may obstruct critical phagocytosis pathways, with CR1-mediated rosette formation potentially inhibiting complement-opsonized clearance of iRBCs (111). Third, the larger size and increased rigidity of rosettes, compared to individual iRBCs, hinder their engulfment by phagocytes, as larger structures impair the attachment phase of phagocytosis

(150), (151). Finally, the reduced deformability of rosettes likely promotes their mechanical sequestration in microvasculature, preventing their filtration and clearance by the spleen (152). Together, these mechanisms highlight the multifaceted role of rosetting in protecting the parasite from immune responses.

1.6 STEVOR: variability and host immune response.

Members of the 2TM superfamily possess antigenic properties and, like the *PfEMP1* family proteins, they are clonally variant contributing to *P. falciparum*'s antigenic variability (102). STEVOR and RIFIN proteins are encoded by a multi-copy gene families, with only one gene expressed at a time by a parasite, following an infection stage-dependent successive expression (134),(135). The chromosome position and expression of 2TM STEVOR variants on the IEs is schematised in Figure 5. Evasion of host immune response is part of the *P. falciparum* virulence. This is primarily achieved through antigenic variation, allowing the parasite to evade immune detection and subsequent destruction in the spleen, prolong its survival, and establish chronic infection (36). The proposed 2TM topology of STEVOR proteins suggests that there is a protein loop exposed to the extracellular space, the V2 domain, the hypervariable region between family members, suggesting antibody stimulating epitopes are found within this domain (102). Furthermore, this variability has been observed both within and between STEVOR family members from different geographically located isolates, with a certain level of clustering association noted in the protein sequences across isolates (125).

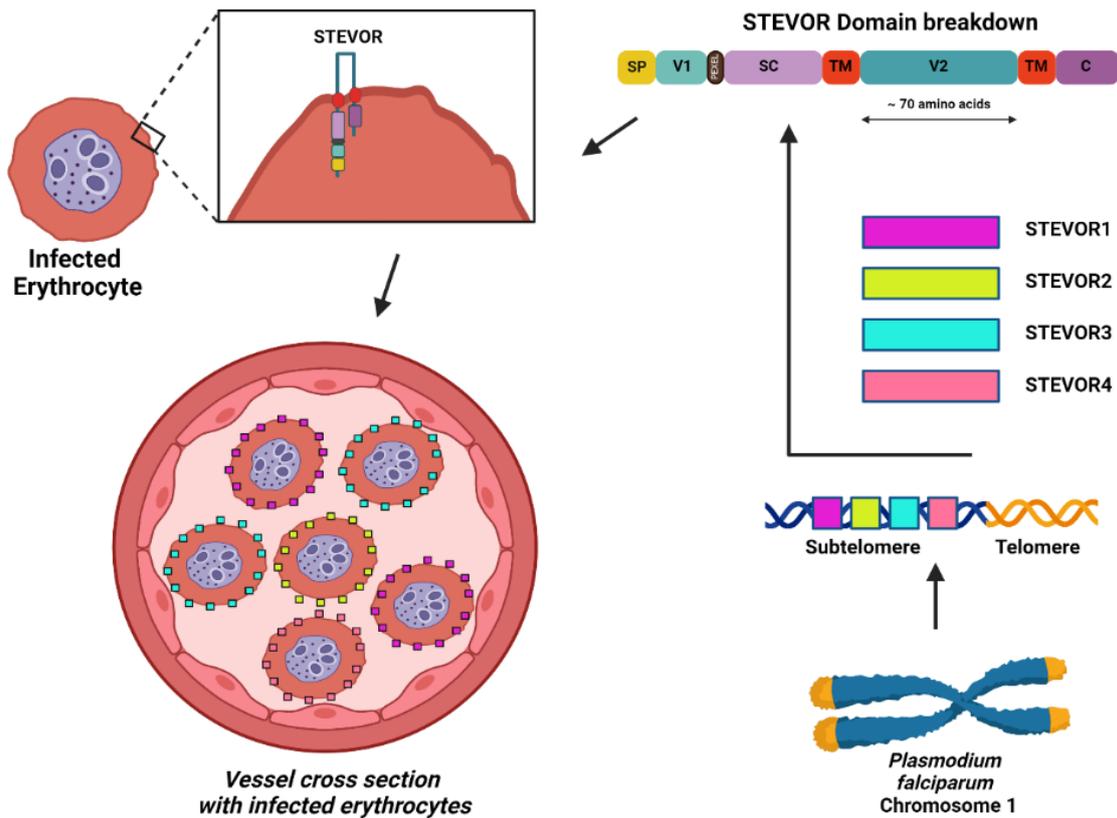


Figure 5: *stevor* genes position on *P. falciparum* chromosomes, protein expression, domain breakdown, and localisation on the IE within and infection. Chromosome one is used as an example. (created by Vasileva H., using Biorender.com)

Studies using peptide microarrays show that children and adults with *P. falciparum* infections exhibit similar levels of serorecognition and seroreactivity to multiple parts of several STEVOR proteins, including SC, V1 and V2. This suggests that other parts of the proteins, besides V2, are also exposed to the host circulation and possess antigenic properties (138). Moreover, it has been demonstrated that seroreactivity to the peptides is age-related, with more intense reactivity detected in adults compared to children, and that serorecognition to STEVORs might be acquired gradually in life, in comparison to RIFINs (138). Additionally, it has been shown that antibodies against STEVOR proteins might be protective against erythrocyte binding, and can inhibit merozoite invasion of the erythrocytes (138),(153). Antibodies against STEVOR proteins have been shown to play a protective role by inhibiting merozoite invasion of

erythrocytes. In merozoites, STEVORs are specifically localized on the apical ends of the parasite, a region critical for interactions with erythrocytes during invasion (153). This apical localization is part of a tightly regulated expression cascade, where distinct STEVOR variants, are predominantly expressed during merozoite maturation. These variants are integrated into the merozoite surface membrane, as demonstrated through Western blot analyses and immunofluorescence microscopy (153). Upon release from schizont-infected erythrocytes, merozoites expose STEVORs on their surface, where they act as mediators of host-parasite interactions. Notably, antibodies such as anti-PFL2610w specifically recognize these STEVOR variants on the merozoite surface but not on intact schizont-infected erythrocytes, underscoring the stage-specific nature of STEVOR expression (153). By binding to these surface-exposed STEVORs, these antibodies can disrupt the interaction between merozoites and erythrocyte receptors, effectively blocking the invasion process and preventing the establishment of new infections. This highlights the potential of anti-STEVOR antibodies as a critical component of protective immunity against malaria.

Study using 3D7 STEVOR recombinant proteins expressed using a wheat germ cell-free (WGCF) system tested against serum samples from Uganda aged 6 to 20 years showed that the accumulation of antibodies to STEVOR variants, breadth of responses, is associated with age but not with clinical malaria. It also identified one variant associated with reduced risk of febrile malaria, thus a target of naturally acquired immunity (154). Similarly, a study using recombinant 3D7 STEVOR proteins expressed in *E. coli* expression system tested against serum samples from Ghana showed high breadth of responses to STEVORs in adults and higher antibody levels in children with parasitaemia irrespective of the presence or absence of disease symptoms. There was an increase of antibody levels against one STEVOR following subsequent infections, indicating an age- and exposure-related acquisition of antibodies against

this STEVOR antigen. However, there were no evidence of protective levels of anti-STEVOR antibodies in children from subsequent infection (155).

1.7 Importance of pathogen protein hypervariability

Hypervariability within protein families in various human, animal and plant pathogens have been linked to adaptation of the pathogen against the host defence responses to ensure prolonged survival. An example of such protein family in pathogenic parasites, apart from the *PfEMP1* in *P. falciparum*, is the trans-sialidases (TSs) protein superfamily which large number of genes are also located near the telomeric regions of the chromosomes of various *Trypanosoma spp.* Similarly to *PfEMP1*, TSs are associated with immune evasion and binding to host cells to facilitate the spread of infection (156). Moreover, bacterial infections, such as group A streptococcus expresses a surface hypervariable protein (M protein) which is also associated with antigenic variation and evasion of immune response, particularly resistance to phagocytosis (157). Further on, even plant viruses such as the Soybean mosaic virus (SMV), the most economically destructive viral pathogen, also expresses a hypervariable protein: P1, associated with virus survival by silencing host defences against the infection (158).

Research on hypervariable pathogenic protein families has shown a common characteristic of antigenic variation and functional importance for host survival with strong associations with diseases pathology. Hence, focusing on understudied surface hypervariable protein families of pathogens associated with big global disease burden and suggested importance in disease pathology, such as the STEVOR protein family of *P. falciparum* should be pursued.

The following thesis has aimed to characterise the protein family antigenic variation and its immunological importance using a multidisciplinary approach employing biochemical, bioinformatics and sero-epidemiological techniques.

1.8 References

1. Talapko J, Škrlec I, Alebić T, Jukić M, Včev A. Malaria: The past and the present. Vol. 7, Microorganisms. MDPI AG; 2019.
2. Antinori S, Galimberti L, Milazzo L, Corbellino M. Biology of human malaria plasmodia including *Plasmodium knowlesi*. Vol. 4, Mediterranean Journal of Hematology and Infectious Diseases. Catholic University in Rome; 2012.
3. Sutherland CJ, Tanomsing N, Nolder D, Oguike M, Jennison C, Pukrittayakamee S, et al. Two Nonrecombining Sympatric Forms of the Human Malaria Parasite *Plasmodium ovale* Occur Globally. *J Infect Dis* [Internet]. 2010 May 15 [cited 2020 Jun 17];201(10):1544–50. Available from: <https://academic.oup.com/jid/article-lookup/doi/10.1086/652240>
4. Miller RL, Ikram S, Armelagos GJ, Walker R, Harers B, Shiffi CJ, et al. Diagnosis of *Plasmodium falciparum* infections in mummies using the rapid manual ParaSight™-F test. *Trans R Soc Trop Med Hyg.* 1994;88(1):31–2.
5. Sherman IW. Malaria: parasite biology, pathogenesis, and protection. Washington, DC: ASM Press; 1998.
6. Jarcho S. Laveran's Discovery in the Retrospect of a Century - PubMed. *Bull Hist Med* [Internet]. 1984 [cited 2020 Jun 22];58(2):215–24. Available from: <https://pubmed.ncbi.nlm.nih.gov/6375771/>
7. Cox FE. History of the discovery of the malaria parasites and their vectors [Internet]. Vol. 3, Parasites and Vectors. BioMed Central; 2010 [cited 2020 Jun 22]. p. 5. Available from: <http://parasitesandvectors.biomedcentral.com/articles/10.1186/1756-3305-3-5>

8. Karen A. Man and microbes: disease and plagues in history and modern times. 1st Touchs. New York, NY: Simon & Schuster; 1996.
9. Arrow KJ, Panosian C, Gelband H. A Brief History of Malaria. 2004;
10. Our History - Our Story | About | CDC [Internet]. [cited 2020 Jun 18]. Available from: <https://www.cdc.gov/about/history/index.html>
11. Prevention C-C for DC and. CDC - Malaria - About Malaria - Disease. 2019;
12. History of malaria elimination in the European Region. 2016.
13. Bayoumi RA. The sickle-cell trait modifies the intensity and specificity of the immune response against *P. falciparum* malaria and leads to acquired protective immunity. *Med Hypotheses* [Internet]. 1987 [cited 2020 Jun 22];22(3):287–98. Available from: <https://pubmed.ncbi.nlm.nih.gov/3295496/>
14. Mulumba LL, Wilson L. Sickle cell disease among children in Africa: An integrative literature review and global recommendations. Vol. 3, *International Journal of Africa Nursing Sciences*. Elsevier Ltd; 2015. p. 56–64.
15. Curtin PD. “The White Man’s Grave:” Image and Reality, 1780-1850. *J Br Stud* [Internet]. 1961 Nov [cited 2020 Jun 22];1(1):94–110. Available from: </core/journals/journal-of-british-studies/article/white-mans-grave-image-and-reality-17801850/3D5A35C18B1B4D2A4956641007FB4572>
16. Simon H. The global distribution and population at risk of malaria: past, present and future. *Lancet Infect Dis*. 2004;4:327–36.
17. Bryan D, Silva N, Rigsby P, Dougall T, Corran P, Mei +, et al. WHO/BS/2014.2235
ENGLISH ONLY EXPERT COMMITTEE ON BIOLOGICAL

- STANDARDIZATION Geneva, 13 to 17 October 2014 International collaborative study to evaluate and establish the 1st WHO Reference Reagent for Anti-Malaria (*Plasmodium falciparum*) human serum. 2014;
18. World malaria report 2022 - World Health Organization - Google Books [Internet]. [cited 2023 Mar 13]. Available from:
https://books.google.com/books/about/World_Malaria_Report_2022.html?id=ST-hEAAAQBAJ&printsec=frontcover&source=kp_read_button&hl=en&redir_esc=y#v=onepage&q&f=false
 19. Health Organization W. World malaria report 2023 -- spread view. 2023 [cited 2024 May 21]; Available from: <https://www.who.int/about/licensing>.
 20. Health Organization W. GLOBAL TECHNICAL STRATEGY FOR MALARIA 2016-2030 Global Technical Strategy for Malaria 2016-2030 Global Malaria Programme World Health Organization.
 21. Le PVV, Kumar P, Ruiz MO, Mbogo C, Muturi EJ. Predicting the direct and indirect impacts of climate change on malaria in coastal Kenya. PLoS One [Internet]. 2019 Feb 1 [cited 2023 Mar 13];14(2). Available from: [/pmc/articles/PMC6364917/](https://pubmed.ncbi.nlm.nih.gov/33011117/)
 22. Soulard V, Bosson-Vanga H, Lorthiois A, Roucher C, Franetich JF, Zanghi G, et al. *Plasmodium falciparum* full life cycle and *Plasmodium ovale* liver stages in humanized mice. Nat Commun. 2015 Jul 24;6.
 23. Laveran A. Un nouveau parasite trouvé dans le sang des malades atteints de fièvre ... - Alphonse Laveran - Google Books [Internet]. 1881 [cited 2020 Jun 24]. 164 p. Available from:
https://books.google.co.uk/books/about/Un_nouveau_parasite_trouvé_dans_le_sang.html?id=_cCDtgAACAAJ&redir_esc=y

24. Marchiafave E. On Summer-Autumn Malaria Fevers. London: New Sydenham Society; 1894.
25. Maccallum WG. ON THE FLAGELLATED FORM OF THE MALARIAL PARASITE. Lancet [Internet]. 1897 Nov 13 [cited 2020 Jun 25];150(3872):1240–1. Available from: <http://www.thelancet.com/article/S0140673600465566/fulltext>
26. Ross R. On some peculiar pigmented cells found in two mosquitos fed on malarial blood. Br Med J [Internet]. 1897 Dec 18 [cited 2020 Jun 25];2(1929):1786–8. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2408186/>
27. Trager W, Jensen JB. Human malaria parasites in continuous culture. Science (80-) [Internet]. 1976 Aug 20 [cited 2020 Jun 29];193(4254):673–5. Available from: <https://science.sciencemag.org/content/193/4254/673>
28. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, et al. Genome sequence of the human malaria parasite Plasmodium falciparum. Nature [Internet]. 2002 [cited 2020 Jun 29];419(6906):498–511. Available from: <https://www.nature.com/articles/nature01097>
29. Usui M, Prajapati SK, Ayanful-Torgby R, Acquah FK, Cudjoe E, Kakaney C, et al. Plasmodium falciparum sexual differentiation in malaria patients is associated with host factors and GDV1-dependent genes. Nat Commun [Internet]. 2019 Dec 1 [cited 2024 Aug 26];10(1). Available from: [/pmc/articles/PMC6514009/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6514009/)
30. Meibalan E, Marti M. Biology of malaria transmission. Cold Spring Harb Perspect Med [Internet]. 2017 Mar 1 [cited 2020 Jun 29];7(3):a025452. Available from: <http://perspectivesinmedicine.cshlp.org/lookup/doi/10.1101/cshperspect.a025452>
31. Maier AG, Matuschewski K, Zhang M, Rug M. Plasmodium falciparum. Vol. 35,

- Trends in Parasitology. Elsevier Ltd; 2019. p. 481–2.
32. Schumacher RF, Spinelli E. Malaria in children [Internet]. Vol. 4, Mediterranean Journal of Hematology and Infectious Diseases. Catholic University in Rome; 2012 [cited 2020 Jul 1]. p. 4. Available from: </pmc/articles/PMC3507524/?report=abstract>
 33. Murphy SC, Shott JP, Parikh S, Etter P, Prescott WR, Stewart VA. Review article: Malaria diagnostics in clinical trials [Internet]. Vol. 89, American Journal of Tropical Medicine and Hygiene. Am J Trop Med Hyg; 2013 [cited 2020 Jul 1]. p. 824–39. Available from: <https://pubmed.ncbi.nlm.nih.gov/24062484/>
 34. Schantz-Dunn J, Nour NM. Malaria and pregnancy: a global health perspective. Rev Obstet Gynecol [Internet]. 2009 [cited 2020 Jul 1];2(3):186–92. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19826576>
 35. Nankabirwa J, Brooker SJ, Clarke SE, Fernando D, Gitonga CW, Schellenberg D, et al. Malaria in school-age children in Africa: an increasingly important challenge. Trop Med Int Heal [Internet]. 2014 Nov 1 [cited 2024 Aug 26];19(11):1294. Available from: </pmc/articles/PMC4285305/>
 36. Miller LH, Baruch DI, Marsh K, Doumbo OK. The pathogenic basis of malaria. Nature [Internet]. 2002 Feb 7 [cited 2021 Nov 17];415(6872):673–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/11832955/>
 37. Tangpukdee N, Duangdee C, Wilairatana P, Krudsood S. Malaria diagnosis: A brief review [Internet]. Vol. 47, Korean Journal of Parasitology. Korean Society for Parasitology; 2009 [cited 2020 Jul 1]. p. 93–102. Available from: </pmc/articles/PMC2688806/?report=abstract>
 38. Hofmann N, Mwingira F, Shekalaghe S, Robinson LJ, Mueller I, Felger I. Ultra-

- Sensitive Detection of Plasmodium falciparum by Amplification of Multi-Copy Subtelomeric Targets. PLOS Med [Internet]. 2015 Mar 1 [cited 2024 Apr 24];12(3):e1001788. Available from:
<https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001788>
39. Mathison BA, Pritt BS. Update on malaria diagnostics and test utilization [Internet]. Vol. 55, Journal of Clinical Microbiology. American Society for Microbiology; 2017 [cited 2020 Jul 1]. p. 2009–17. Available from:
</pmc/articles/PMC5483902/?report=abstract>
40. Kakkilaya BS. Rapid Diagnosis of Malaria. Lab Med [Internet]. 2003 Aug 1 [cited 2020 Jul 1];34(8):602–8. Available from:
<https://academic.oup.com/labmed/article/34/8/602/2504332>
41. WHO Technical Report Series 892 WHO EXPERT COMMITTEE ON MALARIA A Twentieth Report a a World Health Organization Geneva.
42. Gatton, M.L., Chaudhry, A., Glenn J et al. Impact of Plasmodium falciparum gene deletions on malaria rapid diagnostic test performance. Malar J. 2020;19(392).
43. Jun SO, Jang SK, Chang HL, Deok HN, Sun HK, Dae WP, et al. Evaluation of a malaria antibody enzyme immunoassay for use in blood screening. Mem Inst Oswaldo Cruz [Internet]. 2008 Feb [cited 2020 Jul 1];103(1):75–8. Available from:
<https://pubmed.ncbi.nlm.nih.gov/18345458/>
44. Achan J, Talisuna AO, Erhart A, Yeka A, Tibenderana JK, Baliraine FN, et al. Quinine, an old anti-malarial drug in a modern world: Role in the treatment of malaria [Internet]. Vol. 10, Malaria Journal. Malar J; 2011 [cited 2020 Jul 5]. Available from:
<https://pubmed.ncbi.nlm.nih.gov/21609473/>

45. Peters W. *Chemotherapy and Drug Resistance in Malaria*. 2nd edition. London Academic Press; 1987.
46. Cristobal B-CS. DDT and Silent Spring: Fifty years after. *J Mil Veretans' Heal* [Internet]. 2011 [cited 2020 Jul 5];19(4):19–24. Available from: <https://jmvh.org/article/ddt-and-silent-spring-fifty-years-after/>
47. Jarman WM, Ballschmiter K. From coal to DDT: The history of the development of the pesticide DDT from synthetic dyes till Silent Spring. Vol. 36, *Endeavour*. Elsevier Current Trends; 2012. p. 131–42.
48. Health Organization W. *Achieving and maintaining universal coverage with long-lasting insecticidal nets for malaria control*. 2017.
49. Burg RW, Miller BM, Baker EE, Birnbaum J, Currie SA, Hartman R, et al. Avermectins, new family of potent anthelmintic agents: Producing organism and fermentation. *Antimicrob Agents Chemother* [Internet]. 1979 [cited 2020 Oct 20];15(3):361–7. Available from: </pmc/articles/PMC352666/?report=abstract>
50. Meyers JI, Gray M, Kuklinski W, Johnson LB, Snow CD, Black WC, et al. Characterization of the target of ivermectin, the glutamate-gated chloride channel, from *Anopheles gambiae*. *J Exp Biol* [Internet]. 2015 May 15 [cited 2020 Oct 20];218(10):1478–86. Available from: </pmc/articles/PMC4448665/?report=abstract>
51. Ikeda T. Pharmacological effects of ivermectin, an antiparasitic agent for intestinal strongyloidiasis: Its mode of action and clinical efficacy [Internet]. Vol. 122, *Folia Pharmacologica Japonica*. *Nihon Yakurigaku Zasshi*; 2003 [cited 2020 Oct 20]. p. 527–38. Available from: <https://pubmed.ncbi.nlm.nih.gov/14639007/>
52. Kobylinski KC, Sylla M, Chapman PL, Sarr MD, Foy BD. Ivermectin mass drug

- administration to humans disrupts malaria parasite transmission in Senegalese villages. *Am J Trop Med Hyg* [Internet]. 2011 Jul [cited 2020 Oct 20];85(1):3–5. Available from: [/pmc/articles/PMC3122335/?report=abstract](#)
53. Smit MR, Ochomo E, Aljayyousi G, Kwambai T, Abong’o B, Bayoh N, et al. Efficacy and Safety of High-Dose Ivermectin for Reducing Malaria Transmission (IVERMAL): Protocol for a Double-Blind, Randomized, Placebo-Controlled, Dose-Finding Trial in Western Kenya. *JMIR Res Protoc* [Internet]. 2016 Nov 17 [cited 2020 Oct 20];5(4):e213. Available from: [/pmc/articles/PMC5133431/?report=abstract](#)
 54. Smit MR, Ochomo EO, Aljayyousi G, Kwambai TK, Abong’o BO, Chen T, et al. Safety and mosquitocidal efficacy of high-dose ivermectin when co-administered with dihydroartemisinin-piperaquine in Kenyan adults with uncomplicated malaria (IVERMAL): a randomised, double-blind, placebo-controlled trial. *Lancet Infect Dis* [Internet]. 2018 Jun 1 [cited 2020 Oct 20];18(6):615–26. Available from: <https://pubmed.ncbi.nlm.nih.gov/29602751/>
 55. RTS, Partnership SCT. Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: Final results of a phase 3, individually randomised, controlled trial. *Lancet* [Internet]. 2015 Jul 4 [cited 2020 Sep 16];386(9988):31–45. Available from: [/pmc/articles/PMC5626001/?report=abstract](#)
 56. Dattoo MS, Natama MH, Somé A, Traoré O, Rouamba T, Bellamy D, et al. Efficacy of a low-dose candidate malaria vaccine, R21 in adjuvant Matrix-M, with seasonal administration to children in Burkina Faso: a randomised controlled trial. *www.thelancet.com* [Internet]. 2021 [cited 2023 Jun 10];397. Available from: <https://doi.org/10.1016/>
 57. Mahase E. Re: WHO recommends second vaccine R21/Matrix-M vaccine for malaria

- prevention in children. *BMJ*. 2024 Mar 19;383:2291.
58. Gandhi K, Thera MA, Coulibaly D, Traoré K, Guindo AB, Ouattara A, et al. Variation in the Circumsporozoite Protein of *Plasmodium falciparum*: Vaccine Development Implications. *PLoS One* [Internet]. 2014 Jul 3 [cited 2024 Sep 10];9(7):101783. Available from: [/pmc/articles/PMC4081809/](https://pubmed.ncbi.nlm.nih.gov/2481809/)
 59. Arnold KB, Chung AW. Prospects from systems serology research. *Immunology*. 2018 Mar 1;153(3):279–89.
 60. Charles A Janeway J, Travers P, Walport M, Shlomchik MJ. The structure of a typical antibody molecule. 2001 [cited 2024 Oct 9]; Available from: <https://www.ncbi.nlm.nih.gov/books/NBK27144/>
 61. Janeway C. *Janeway's Immunobiology*. 9th ed. Murphy K, Weaver C, editors. Garland Science; 2017.
 62. Guzman MG, Halstead SB, Artsob H, Buchy P, Farrar J, Gubler DJ, et al. Dengue: A continuing global threat. *Nat Rev Microbiol* [Internet]. 2010 [cited 2020 Sep 22];8(12):S7–16. Available from: <https://pubmed.ncbi.nlm.nih.gov/21079655/>
 63. Tomaras GD, Yates NL, Liu P, Qin L, Fouda GG, Chavez LL, et al. Initial B-Cell Responses to Transmitted Human Immunodeficiency Virus Type 1: Virion-Binding Immunoglobulin M (IgM) and IgG Antibodies Followed by Plasma Anti-gp41 Antibodies with Ineffective Control of Initial Viremia. *J Virol* [Internet]. 2008 Dec 15 [cited 2020 Sep 22];82(24):12449–63. Available from: <https://pubmed.ncbi.nlm.nih.gov/18842730/>
 64. Bergmann-Leitner ES, Duncan EH, Angov E. MSP-1p42-specific antibodies affect growth and development of intra-erythrocytic parasites of *Plasmodium falciparum*.

- Malar J [Internet]. 2009 Aug 3 [cited 2024 Nov 25];8(1):1–12. Available from:
<https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-8-183>
65. Dutta S, Haynes JD, Barbosa A, Ware LA, Snaveley JD, Moch JK, et al. Mode of Action of Invasion-Inhibitory Antibodies Directed against Apical Membrane Antigen 1 of *Plasmodium falciparum*. *Infect Immun* [Internet]. 2005 Apr [cited 2024 Nov 25];73(4):2116. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC1087451/>
 66. Fowkes FJI, Richards JS, Simpson JA, Beeson JG. The Relationship between Anti-merozoite Antibodies and Incidence of *Plasmodium falciparum* Malaria: A Systematic Review and Meta-analysis. Bejon PA, editor. *PLoS Med* [Internet]. 2010 Jan 19 [cited 2020 Sep 22];7(1):e1000218. Available from:
<https://dx.plos.org/10.1371/journal.pmed.1000218>
 67. Ord RL, Caldeira JC, Rodriguez M, Noe A, Chackerian B, Peabody DS, et al. A malaria vaccine candidate based on an epitope of the *Plasmodium falciparum* RH5 protein. *Malar J* [Internet]. 2014 Aug 18 [cited 2024 May 24];13(1). Available from:
<https://pubmed.ncbi.nlm.nih.gov/25135070/>
 68. Wu RL, Idris AH, Berkowitz NM, Happe M, Gaudinski MR, Buettner C, et al. Low-Dose Subcutaneous or Intravenous Monoclonal Antibody to Prevent Malaria. *N Engl J Med* [Internet]. 2022 Aug 8 [cited 2024 May 24];387(5):397. Available from:
[/pmc/articles/PMC9806693/](https://pmc/articles/PMC9806693/)
 69. Coggeshall LT, Kumm HW. Demonstration of passive immunity in experimental monkey malaria. *J Exp Med* [Internet]. 1937 Aug 1 [cited 2020 Sep 16];66(2):177–89. Available from: <https://pubmed.ncbi.nlm.nih.gov/19870655/>
 70. Teo A, Feng G, Brown G V., Beeson JG, Rogerson SJ. Functional Antibodies and Protection against Blood-stage Malaria [Internet]. Vol. 32, *Trends in Parasitology*.

Elsevier Ltd; 2016 [cited 2020 Sep 16]. p. 887–98. Available from:

<https://pubmed.ncbi.nlm.nih.gov/27546781/>

71. Walker MR, Knudsen AS, Partey FD, Bassi MR, Frank AM, Castberg FC, et al. Acquisition and decay of IgM and IgG responses to merozoite antigens after *Plasmodium falciparum* malaria in Ghanaian children. *PLoS One* [Internet]. 2020 Dec 1 [cited 2024 Nov 25];15(12):e0243943. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC7746192/>
72. Boyle MJ, Chan JA, Handayani I, Reiling L, Feng G, Hilton A, et al. IgM in human immunity to *Plasmodium falciparum* malaria. *Sci Adv* [Internet]. 2019 Sep 25 [cited 2020 Sep 16];5(9). Available from: [/pmc/articles/PMC6760923/?report=abstract](https://pmc.ncbi.nlm.nih.gov/articles/PMC6760923/?report=abstract)
73. Beeson JG, Drew DR, Boyle MJ, Feng G, Fowkes FJI, Richards JS. Merozoite surface proteins in red blood cell invasion, immunity and vaccines against malaria [Internet]. Vol. 40, *FEMS Microbiology Reviews*. Oxford University Press; 2016 [cited 2020 Sep 22]. p. 343–72. Available from: <https://pubmed.ncbi.nlm.nih.gov/26833236/>
74. Rodriguez-Barraquer I, Arinaitwe E, Jagannathan P, Kanya MR, Rosenthal PJ, Rek J, et al. Quantification of anti-parasite and anti-disease immunity to malaria as a function of age and exposure. *Elife* [Internet]. 2018 Jul 25 [cited 2024 Apr 24];7. Available from: <https://pubmed.ncbi.nlm.nih.gov/30044224/>
75. Dechavanne C, Dechavanne S, Sadissou I, Lokossou AG, Alvarado F, Dambrun M, et al. Associations between an IgG3 polymorphism in the binding domain for FcRn, transplacental transfer of malaria-specific IgG3, and protection against *Plasmodium falciparum* malaria during infancy: A birth cohort study in Benin. *PLoS Med* [Internet]. 2017 Oct 1 [cited 2024 Apr 24];14(10). Available from: <https://pubmed.ncbi.nlm.nih.gov/28991911/>

76. Guinovart C, Dobaño C, Bassat Q, Nhabomba A, Quintó L, Manaca MN, et al. The role of age and exposure to *Plasmodium falciparum* in the rate of acquisition of naturally acquired immunity: a randomized controlled trial. *PLoS One* [Internet]. 2012 Mar 7 [cited 2024 Apr 24];7(3). Available from: <https://pubmed.ncbi.nlm.nih.gov/22412865/>
77. Griffin JT, Déirdre Hollingsworth T, Reyburn H, Drakeley CJ, Riley EM, Ghani AC. Gradual acquisition of immunity to severe malaria with increasing exposure. *Proceedings Biol Sci* [Internet]. 2015 [cited 2024 Apr 24];282(1801). Available from: <https://pubmed.ncbi.nlm.nih.gov/25567652/>
78. Barua P, Beeson JG, Maleta K, Ashorn P, Rogerson SJ. The impact of early life exposure to *Plasmodium falciparum* on the development of naturally acquired immunity to malaria in young Malawian children. *Malar J* [Internet]. 2019 Jan 18 [cited 2024 Apr 24];18(1):1–12. Available from: <https://malariajournal.biomedcentral.com/articles/10.1186/s12936-019-2647-8>
79. Idro R, Aloyo J, Mayende L, Bitarakwate E, John CC, Kivumbi GW. Severe malaria in children in areas with low, moderate and high transmission intensity in Uganda. *Trop Med Int Health* [Internet]. 2006 Jan [cited 2024 Apr 24];11(1):115–24. Available from: <https://pubmed.ncbi.nlm.nih.gov/16398762/>
80. Aponte JJ, Menendez C, Schellenberg D, Kahigwa E, Mshinda H, Vountasou P, et al. Age interactions in the development of naturally acquired immunity to *Plasmodium falciparum* and its clinical presentation. *PLoS Med* [Internet]. 2007 [cited 2024 Apr 24];4(7):1259–67. Available from: <https://pubmed.ncbi.nlm.nih.gov/17676985/>
81. Kevin Baird J, Masbar S, Basri H, Tirtokusumo S, Subianto B, Hoffman SL. Age-Dependent Susceptibility to Severe Disease with Primary Exposure to *Plasmodium*

- falciparum. *J Infect Dis* [Internet]. 1998 Aug 1 [cited 2024 Nov 25];178(2):592–5.
Available from: <https://dx.doi.org/10.1086/517482>
82. Baird JK. Age-dependent characteristics of protection v. susceptibility to *Plasmodium falciparum*. *Ann Trop Med Parasitol*. 1998;92(4):367–90.
 83. Curtin PD. Malarial immunities in nineteenth-century West Africa and the Caribbean. *Parasitologia*. 1994;36(1):69–82.
 84. Ranjha R, Singh K, Baharia RK, Mohan M, Anvikar AR, Bharti PK. Age-specific malaria vulnerability and transmission reservoir among children. *Glob Pediatr*. 2023 Dec 1;6:100085.
 85. Pinkevych M, Petravic J, Chelimo K, Kazura JW, Moormann AM, Davenport MP. The Dynamics of Naturally Acquired Immunity to *Plasmodium falciparum* Infection. *PLOS Comput Biol* [Internet]. 2012 [cited 2024 Apr 24];8(10):e1002729. Available from: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002729>
 86. Proietti C, Krause L, Trieu A, Dodoo D, Gyan B, Koram KA, et al. Immune signature against *Plasmodium falciparum* antigens predicts clinical immunity in distinct malaria endemic communities [Internet]. Vol. 24. [cited 2020 Sep 24]. Available from: <https://www.mcponline.org>
 87. Wu L, Mwesigwa J, Affara M, Bah M, Correa S, Hall T, et al. Antibody responses to a suite of novel serological markers for malaria surveillance demonstrate strong correlation with clinical and parasitological infection across seasons and transmission settings in The Gambia. *BMC Med* [Internet]. 2020 Sep 25 [cited 2023 Mar 13];18(1):304. Available from: [/pmc/articles/PMC7517687/](https://pubmed.ncbi.nlm.nih.gov/35117687/)
 88. Rono J, Osier FHA, Olsson D, Montgomery S, Mhoja L, Rooth I, et al. Breadth of

- Anti-Merozoite Antibody Responses Is Associated With the Genetic Diversity of Asymptomatic *Plasmodium falciparum* Infections and Protection Against Clinical Malaria. *Clin Infect Dis* [Internet]. 2013 Nov 15 [cited 2024 Apr 4];57(10):1409–16. Available from: <https://dx.doi.org/10.1093/cid/cit556>
89. Proietti C, Doolan DL. The case for a rational genome-based vaccine against malaria [Internet]. Vol. 5, *Frontiers in Microbiology*. Frontiers Media S.A.; 2014 [cited 2020 Sep 24]. p. 741. Available from: <http://www.malariavaccine.org/files/MVI-GSK-RTSSfactsheetFINAL-web.pdf>
90. Malkin EM, Diemert DJ, McArthur JH, Perreault JR, Miles AP, Giersing BK, et al. Phase 1 Clinical Trial of Apical Membrane Antigen 1: an Asexual Blood-Stage Vaccine for *Plasmodium falciparum* Malaria. *Infect Immun* [Internet]. 2005 Jun [cited 2023 May 5];73(6):3677. Available from: </pmc/articles/PMC1111886/>
91. Blank A, Fürle K, Jäschke A, Mikus G, Lehmann M, Hüsing J, et al. Immunization with full-length *Plasmodium falciparum* merozoite surface protein 1 is safe and elicits functional cytophilic antibodies in a randomized first-in-human trial. *npj Vaccines* 2020 51 [Internet]. 2020 Jan 31 [cited 2023 May 4];5(1):1–15. Available from: <https://www.nature.com/articles/s41541-020-0160-2>
92. Byrne B, Stack E, Gilmartin N, O’Kennedy R. Antibody-Based Sensors: Principles, Problems and Potential for Detection of Pathogens and Associated Toxins. *Sensors (Basel)* [Internet]. 2009 Jun [cited 2023 Mar 13];9(6):4407. Available from: </pmc/articles/PMC3291918/>
93. Wu L, Hall T, Ssewanyana I, Oulton T, Patterson C, Vasileva H, et al. Optimisation and standardisation of a multiplex immunoassay of diverse *Plasmodium falciparum* antigens to assess changes in malaria transmission using sero-epidemiology. *Wellcome*

- Open Res [Internet]. 2020 Apr 23 [cited 2020 Oct 15];4:26. Available from:
<https://doi.org/10.12688/wellcomeopenres.14950.1>
94. Drakeley CJ, Corran PH, Coleman PG, Tongren JE, McDonald SLR, Carneiro I, et al. Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. *Proc Natl Acad Sci U S A* [Internet]. 2005 Apr 5 [cited 2020 Sep 29];102(14):5108–13. Available from:
www.pnas.org/cgi/doi/10.1073/pnas.0408725102
 95. White MT, Griffin JT, Akpogheneta O, Conway DJ, Koram KA, Riley EM, et al. Dynamics of the antibody response to *Plasmodium falciparum* infection in African children. *J Infect Dis* [Internet]. 2014 Oct 1 [cited 2020 Mar 24];210(7):1115–22. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24719471>
 96. Akpogheneta OJ, Duah NO, Tetteh KKA, Dunyo S, Lanar DE, Pinder M, et al. Duration of Naturally Acquired Antibody Responses to Blood-Stage *Plasmodium falciparum* Is Age Dependent and Antigen Specific. *Infect Immun* [Internet]. 2008 Apr [cited 2023 Mar 13];76(4):1748. Available from: [/pmc/articles/PMC2292892/](https://pubmed.ncbi.nlm.nih.gov/12599061/)
 97. Boutlis CS, Fagan PK, Gowda DC, Lagog M, Mgone CS, Bockarie MJ, et al. Immunoglobulin G (IgG) responses to *Plasmodium falciparum* glycosylphosphatidylinositols are short-lived and predominantly of the IgG3 subclass. *J Infect Dis* [Internet]. 2003 Mar 1 [cited 2023 Mar 13];187(5):862–5. Available from:
<https://pubmed.ncbi.nlm.nih.gov/12599061/>
 98. Helb DA, Tetteh KKA, Felgner PL, Skinner J, Hubbard A, Arinaitwe E, et al. Novel serologic biomarkers provide accurate estimates of recent *Plasmodium falciparum* exposure for individuals and communities. 2015;
 99. Chan J-A, Fowkes FJI, Beeson JG. Surface antigens of *Plasmodium falciparum*-

- infected erythrocytes as immune targets and malaria vaccine candidates. *Cell Mol Life Sci* [Internet]. 2014 Oct 2 [cited 2020 Mar 24];71(19):3633–57. Available from: <http://link.springer.com/10.1007/s00018-014-1614-3>
100. Baruch DI. Adhesive receptors on malaria-parasitized red cells. *Baillieres Best Pract Res Clin Haematol* [Internet]. 1999 [cited 2024 Apr 25];12(4):747–61. Available from: <https://pubmed.ncbi.nlm.nih.gov/10895262/>
 101. Bachmann A, Scholz JAM, Janßen M, Klinkert MQ, Tannich E, Bruchhaus I, et al. A comparative study of the localization and membrane topology of members of the RIFIN, STEVOR and PfMC-2TM protein families in *Plasmodium falciparum*-infected erythrocytes. *Malar J*. 2015 Jul 16;14(1).
 102. Scherf A, Lopez-Rubio JJ, Riviere L. Antigenic Variation in *Plasmodium falciparum*. *Annu Rev Microbiol* [Internet]. 2008 Oct 11 [cited 2019 May 20];62(1):445–70. Available from: <http://www.annualreviews.org/doi/10.1146/annurev.micro.61.080706.093134>
 103. Recker M, Buckee CO, Serazin A, Kyes S, Pinches R, Christodoulou Z, et al. Antigenic variation in *Plasmodium falciparum* malaria involves a highly structured switching pattern. *PLoS Pathog* [Internet]. 2011 Mar [cited 2019 May 20];7(3):e1001306. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21408201>
 104. Baruch DI, Pasloske BL, Singh HB, Bi X, Ma XC, Feldman M, et al. Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. *Cell* [Internet]. 1995 Jul 14 [cited 2020 Oct 5];82(1):77–87. Available from: <https://pubmed.ncbi.nlm.nih.gov/7541722/>
 105. Su X-R, Heatwoie VM, Wertheimer SP, Guinet F, Hertfeldt JA, Peterson DS, et al. The Large Diverse Gene Family w Encodes Proteins Involved in Cytoadherence and

- Antigenic Variation of Plasmodium falciparum-Infected Erythrocytes. *Cell*. 1995;82:89–100.
106. Smith JD. The role of PfEMP1 adhesion domain classification in Plasmodium falciparum pathogenesis research. *Mol Biochem Parasitol* [Internet]. 2014 [cited 2023 Nov 7];195(2):82. Available from: </pmc/articles/PMC4159067/>
107. Urban BC, Ferguson DJP, Pain A, Willcox N, Plebanski M, Austyn JM, et al. Plasmodium falciparum-infected erythrocytes modulate the maturation of dendritic cells. *Nature* [Internet]. 1999 Jul 1 [cited 2024 Apr 26];400(6739):73–7. Available from: <https://pubmed.ncbi.nlm.nih.gov/10403251/>
108. Sampaio NG, Eriksson EM, Schofield L. Plasmodium falciparum PfEMP1 modulates monocyte/macrophage transcription factor activation and cytokine and chemokine responses. *Infect Immun* [Internet]. 2018 Jan 1 [cited 2021 Nov 24];86(1). Available from: <https://journals.asm.org/doi/abs/10.1128/IAI.00447-17>
109. Bouwens EAM, Stavenuiter F, Mosnier LO. Mechanisms of anticoagulant and cytoprotective actions of the protein C pathway. *J Thromb Haemost* [Internet]. 2013 Jun [cited 2024 Apr 26];11(0 1):242. Available from: </pmc/articles/PMC3713536/>
110. Lavstsen T, Turner L, Saguti F, Magistrado P, Rask TS, Jespersen JS, et al. Plasmodium falciparum erythrocyte membrane protein 1 domain cassettes 8 and 13 are associated with severe malaria in children. *Proc Natl Acad Sci U S A* [Internet]. 2012 Jun 26 [cited 2024 Jan 5];109(26):E1791. Available from: </pmc/articles/PMC3387094/>
111. Alexandra Rowe J, Moulds JM, Newbold CI, Miller LH. P. falciparum rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. *Nature* [Internet]. 1997 [cited 2024 Apr 26];388(6639):292–5. Available from: <https://pubmed.ncbi.nlm.nih.gov/9230440/>

112. Rottmann M, Lavstsen T, Mugasa JP, Kaestli M, Jensen ATR, Müller D, et al. Differential expression of var gene groups is associated with morbidity caused by *Plasmodium falciparum* infection in Tanzanian children. *Infect Immun* [Internet]. 2006 Jul [cited 2024 Apr 26];74(7):3904–11. Available from: <https://pubmed.ncbi.nlm.nih.gov/16790763/>
113. Bachmann A, Predehl S, May J, Harder S, Burchard GD, Gilberger TW, et al. Highly co-ordinated var gene expression and switching in clinical *Plasmodium falciparum* isolates from non-immune malaria patients. *Cell Microbiol* [Internet]. 2011 Sep 1 [cited 2024 Apr 26];13(9):1397–409. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1462-5822.2011.01629.x>
114. Scherf A, Hernandez-Rivas R, Buffet P, Bottius E, Benatar C, Pouvelle B, et al. Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in *Plasmodium falciparum*. *EMBO J*. 1998;17(18):5418–26.
115. Milne K, Ivens A, Reid AJ, Lotkowska ME, O’toole A, Sankaranarayanan G, et al. Mapping immune variation and var gene switching in naive hosts infected with *Plasmodium falciparum*. *Elife* [Internet]. 2021 Mar 1 [cited 2024 Apr 22];10. Available from: [/pmc/articles/PMC7924948/](https://pmc/articles/PMC7924948/)
116. Bachmann A, Bruske E, Krumkamp R, Turner L, Wichers JS, Petter M, et al. Controlled human malaria infection with *Plasmodium falciparum* demonstrates impact of naturally acquired immunity on virulence gene expression. *PLoS Pathog* [Internet]. 2019 Jul 1 [cited 2024 Apr 26];15(7). Available from: [/pmc/articles/PMC6650087/](https://pmc/articles/PMC6650087/)
117. Chan JA, Boyle MJ, Moore KA, Reiling L, Lin Z, Hasang W, et al. Antibody Targets on the Surface of *Plasmodium falciparum*–Infected Erythrocytes That Are Associated

- With Immunity to Severe Malaria in Young Children. *J Infect Dis* [Internet]. 2019 Mar 3 [cited 2023 Nov 7];219(5):819. Available from: [/pmc/articles/PMC6376912/](#)
118. Barry AE, Trieu A, Fowkes FJI, Pablo J, Kalantari-Dehaghi M, Jasinskas A, et al. The Stability and Complexity of Antibody Responses to the Major Surface Antigen of *Plasmodium falciparum* Are Associated with Age in a Malaria Endemic Area. *Mol Cell Proteomics* [Internet]. 2011 [cited 2024 Apr 3];10(11). Available from: [/pmc/articles/PMC3226400/](#)
119. Tessema SK, Nakajima R, Jasinskas A, Monk SL, Lekieffre L, Lin E, et al. Protective Immunity against Severe Malaria in Children Is Associated with a Limited Repertoire of Antibodies to Conserved PfEMP1 Variants. *Cell Host Microbe* [Internet]. 2019 Nov 13 [cited 2024 Apr 18];26(5):579-590.e5. Available from: <https://pubmed.ncbi.nlm.nih.gov/31726028/>
120. Cham GKK, Turner L, Kurtis JD, Mutabingwa T, Fried M, Jensen ATR, et al. Hierarchical, domain type-specific acquisition of antibodies to *Plasmodium falciparum* erythrocyte membrane protein 1 in Tanzanian children. *Infect Immun* [Internet]. 2010 Nov [cited 2022 Mar 6];78(11):4653–9. Available from: <https://journals.asm.org/doi/abs/10.1128/IAI.00593-10>
121. Kinyua AW, Turner L, Kimingi HW, Mwai K, Mwikali K, Andisi C, et al. Antibodies to PfEMP1 and variant surface antigens: Protection after controlled human malaria infection in semi-immune Kenyan adults. *J Infect*. 2024 Oct 1;89(4):106252.
122. Kyes SA, Rowe JA, Kriek N, Newbold CI. Rifins: A second family of clonally variant proteins expressed on the surface of red cells infected with *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* [Internet]. 1999 Aug 3 [cited 2020 Oct 5];96(16):9333–8. Available from: [/pmc/articles/PMC17783/?report=abstract](#)

123. Blythe JE, Xue YY, Kuss C, Bozdech Z, Holder AA, Marsh K, et al. Plasmodium falciparum STEVOR proteins are highly expressed in patient isolates and located in the surface membranes of infected red blood cells and the apical tips of merozoites. *Infect Immun*. 2008 Jul;76(7):3329–36.
124. Kwapisz M, Morillon A. Subtelomeric Transcription and its Regulation. *J Mol Biol* [Internet]. 2020 Jul 7 [cited 2023 Aug 29];432(15):4199. Available from: [/pmc/articles/PMC7374410/](https://pubmed.ncbi.nlm.nih.gov/39111110/)
125. Blythe JE, Niang M, Marsh K, Holder AA, Langhorne J, Preiser PR. Characterization of the repertoire diversity of the Plasmodium falciparum stevor multigene family in laboratory and field isolates. *Malar J* [Internet]. 2009 [cited 2023 Apr 11];8(1):140. Available from: [/pmc/articles/PMC2706845/](https://pubmed.ncbi.nlm.nih.gov/18474651/)
126. Blythe JE, Xue YY, Kuss C, Bozdech Z, Holder AA, Marsh K, et al. Plasmodium falciparum STEVOR proteins are highly expressed in patient isolates and located in the surface membranes of infected red blood cells and the apical tips of merozoites. *Infect Immun* [Internet]. 2008 Jul [cited 2020 Oct 5];76(7):3329–36. Available from: <https://pubmed.ncbi.nlm.nih.gov/18474651/>
127. Singh H, Madnani K, Lim YB, Cao J, Preiser PR, Lim CT. Expression dynamics and physiologically relevant functional study of STEVOR in asexual stages of Plasmodium falciparum infection. *Cell Microbiol* [Internet]. 2017 Jun 1 [cited 2023 Nov 8];19(6). Available from: <https://pubmed.ncbi.nlm.nih.gov/28030753/>
128. McRobert L, Preiser P, Sharp S, Jarra W, Kaviratne M, Taylor MC, et al. Distinct trafficking and localization of STEVOR proteins in three stages of the Plasmodium falciparum life cycle. *Infect Immun* [Internet]. 2004 Nov 1 [cited 2020 Jun 29];72(11):6597–602. Available from: <https://iaa.asm.org/content/72/11/6597>

129. Cheng Q, Cloonan N, Fischer K, Thompson J, Waine G, Lanzer M, et al. Stevor and Rif are Plasmodium falciparum multicopy gene families which potentially encode variant antigens. *Mol Biochem Parasitol* [Internet]. 1998 Nov 30 [cited 2020 Oct 7];97(1–2):161–76. Available from: <https://pubmed.ncbi.nlm.nih.gov/9879895/>
130. Petter M, Haeggström M, Khattab A, Fernandez V, Klinkert MQ, Wahlgren M. Variant proteins of the Plasmodium falciparum RIFIN family show distinct subcellular localization and developmental expression patterns. *Mol Biochem Parasitol* [Internet]. 2007 Nov [cited 2020 Oct 7];156(1):51–61. Available from: <https://pubmed.ncbi.nlm.nih.gov/17719658/>
131. Mwakalinga SB, Wang CW, Bengtsson DC, Turner L, Dinko B, Lusingu JP, et al. Expression of a type B RIFIN in Plasmodium falciparum merozoites and gametes. *Malar J* [Internet]. 2012 Dec 21 [cited 2024 Apr 28];11(1):1–12. Available from: <https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-11-429>
132. Petter M, Bonow I, Klinkert MQ. Diverse expression patterns of subgroups of the rif multigene family during Plasmodium falciparum gametocytogenesis. *PLoS One* [Internet]. 2008 Nov 20 [cited 2024 Apr 28];3(11). Available from: <https://pubmed.ncbi.nlm.nih.gov/19020666/>
133. Hypervariability within the Rifin, Stevor and Pfmc-2TM superfamilies in Plasmodium falciparum [Internet]. [cited 2020 Mar 9]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1751529/>
134. Kaur J, Hora R. “2TM proteins”: An antigenically diverse superfamily with variable functions and export pathways. *PeerJ*. 2018;2018(MAY).
135. Bachmann A, Petter M, Tilly A-K, Biller L, Uliczka KA, Duffy MF, et al. Temporal Expression and Localization Patterns of Variant Surface Antigens in Clinical

- Plasmodium falciparum Isolates during Erythrocyte Schizogony. Hviid L, editor. PLoS One [Internet]. 2012 Nov 15 [cited 2020 Oct 6];7(11):e49540. Available from: <https://dx.plos.org/10.1371/journal.pone.0049540>
136. Niang M, Xue YY, Preiser PR. The Plasmodium falciparum STEVOR Multigene Family Mediates Antigenic Variation of the Infected Erythrocyte. PLoS Pathog [Internet]. 2009 Feb [cited 2024 Apr 28];5(2):1000307. Available from: </pmc/articles/PMC2637975/>
 137. Burdukiewicz M, Sobczyk P, Chilimoniuk J, Gagat P, Mackiewicz P. Prediction of Signal Peptides in Proteins from Malaria Parasites. Int J Mol Sci [Internet]. 2018 Dec 1 [cited 2023 Sep 4];19(12). Available from: </pmc/articles/PMC6321056/>
 138. Zhou AE, Berry AA, Bailey JA, Pike A, Dara A, Agrawal S, et al. Antibodies to Peptides in Semiconserved Domains of RIFINs and STEVORs Correlate with Malaria Exposure. mSphere. 2019 Mar 20;4(2).
 139. Boddey JA, Moritz RL, Simpson RJ, Cowman AF. Role of the Plasmodium export element in trafficking parasite proteins to the infected erythrocyte. Traffic [Internet]. 2009 [cited 2020 Oct 7];10(3):285–99. Available from: </pmc/articles/PMC2682620/?report=abstract>
 140. Soni R, Sharma D, Bhatt TK. Plasmodium falciparum secretome in erythrocyte and beyond [Internet]. Vol. 7, Frontiers in Microbiology. Frontiers Media S.A.; 2016 [cited 2020 Oct 7]. p. 194. Available from: www.frontiersin.org
 141. Wichers JS, Scholz JAM, Strauss J, Witt S, Lill A, Ehnold LI, et al. Dissecting the gene expression, localization, membrane topology, and function of the plasmodium falciparum STEVOR protein family. MBio. 2019 Aug 27;10(4).

142. Saito F, Hirayasu K, Satoh T, Wang CW, Lusingu J, Arimori T, et al. Immune evasion of *Plasmodium falciparum* by RIFIN via inhibitory receptors. *Nature* [Internet]. 2017 Dec 12 [cited 2023 Jul 18];552(7683):101. Available from: [/pmc/articles/PMC5748893/](https://pubmed.ncbi.nlm.nih.gov/2748893/)
143. Brown D, Trowsdale J, Allen R. The LILR family: modulators of innate and adaptive immune pathways in health and disease. *Tissue Antigens* [Internet]. 2004 Sep [cited 2024 Apr 28];64(3):215–25. Available from: <https://pubmed.ncbi.nlm.nih.gov/15304001/>
144. Sakoguchi A, Saito F, Hirayasu K, Shida K, Matsuoka S, Itagaki S, et al. *Plasmodium falciparum* RIFIN is a novel ligand for inhibitory immune receptor LILRB2. *Biochem Biophys Res Commun*. 2021 Apr 9;548:167–73.
145. Tan J, Pieper K, Piccoli L, Abdi A, Foglierini M, Geiger R, et al. A LAIR-1 insertion generates broadly reactive antibodies against malaria variant antigens. *Nature* [Internet]. 2015 Jan 7 [cited 2024 Nov 27];529(7584):105. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC4869849/>
146. Chen Y, Xu K, Piccoli L, Foglierini M, Tan J, Jin W, et al. Structural basis of malaria RIFIN binding by LILRB1-containing antibodies. *Nature* [Internet]. 2021 Apr 22 [cited 2024 Nov 27];592(7855):639. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8068667/>
147. Xu K, Wang Y, Shen CH, Chen Y, Zhang B, Liu K, et al. Structural basis of LAIR1 targeting by polymorphic *Plasmodium* RIFINs. *Nat Commun* 2021 121 [Internet]. 2021 Jul 9 [cited 2024 Nov 27];12(1):1–8. Available from: <https://www.nature.com/articles/s41467-021-24291-6>
148. Niang M, Bei AK, Madnani KG, Pelly S, Dankwa S, Kanjee U, et al. STEVOR is a

- plasmodium falciparum erythrocyte binding protein that mediates merozoite invasion and rosetting. *Cell Host Microbe* [Internet]. 2014 Jul 9 [cited 2022 Feb 22];16(1):81–93. Available from: <http://www.cell.com/article/S1931312814002182/fulltext>
149. Sanyal S, Egeé S, Bouyer G, Perrot S, Safeukui I, Bischoff E, et al. Plasmodium falciparum STEVOR proteins impact erythrocyte mechanical properties. *Blood* [Internet]. 2012 Jan 1 [cited 2024 Apr 28];119(2):e1. Available from: </pmc/articles/PMC3257022/>
150. Moll K, Palmkvist M, Ch'ng J, Kiwuwa MS, Wahlgren M. Evasion of Immunity to Plasmodium falciparum: Rosettes of Blood Group A Impair Recognition of PfEMP1. *PLoS One* [Internet]. 2015 Dec 1 [cited 2024 Nov 27];10(12):e0145120. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC4694710/>
151. Champion JA, Walker A, Mitragotri S. Role of particle size in phagocytosis of polymeric microspheres. *Pharm Res* [Internet]. 2008 Aug [cited 2024 Nov 27];25(8):1815–21. Available from: <https://pubmed.ncbi.nlm.nih.gov/18373181/>
152. Nash GB, Cooke BM, Carlson J, Wahlgren M. Rheological properties of rosettes formed by red blood cells parasitized by Plasmodium falciparum. *Br J Haematol* [Internet]. 1992 [cited 2024 Nov 27];82(4):757–63. Available from: <https://pubmed.ncbi.nlm.nih.gov/1482664/>
153. Khattab A, Meri S. Exposure of the Plasmodium falciparum clonally variant STEVOR proteins on the merozoite surface. *Malar J* [Internet]. 2011 [cited 2024 Apr 29];10:58. Available from: </pmc/articles/PMC3062604/>
154. Kanoi BN, Nagaoka H, White MT, Morita M, Palacpac NMQ, Ntege EH, et al. Global Repertoire of Human Antibodies Against Plasmodium falciparum RIFINs, SURFINs, and STEVORs in a Malaria Exposed Population. *Front Immunol* [Internet]. 2020 May

- 12 [cited 2023 Aug 24];11. Available from:
<https://pubmed.ncbi.nlm.nih.gov/32477363/>
155. Schreiber N, Khattab A, Petter M, Marks F, Adjei S, Kobbe R, et al. Expression of *Plasmodium falciparum* 3D7 STEVOR proteins for evaluation of antibody responses following malaria infections in naïve infants. *Parasitology* [Internet]. 2008 Feb [cited 2020 Oct 16];135(2):155–67. Available from:
<https://pubmed.ncbi.nlm.nih.gov/17931459/>
156. De Pablos LM, Osuna A. Multigene Families in *Trypanosoma cruzi* and Their Role in Infectivity. *Infect Immun* [Internet]. 2012 Jul [cited 2024 May 24];80(7):2258. Available from: </pmc/articles/PMC3416482/>
157. Fischetti VA. Streptococcal M protein: molecular design and biological behavior. *Clin Microbiol Rev* [Internet]. 1989 [cited 2024 May 24];2(3):285. Available from: </pmc/articles/PMC358122/?report=abstract>
158. Mao C, Shan S, Huang Y, Jiang C, Zhang H, Li Y, et al. The hypervariable N-terminal of soybean mosaic virus P1 protein influences its pathogenicity and host defense responses. *Phytopathol Res* [Internet]. 2022 Dec 1 [cited 2024 May 24];4(1):1–12. Available from: <https://phytopatholres.biomedcentral.com/articles/10.1186/s42483-022-00115-3>

Chapter 2: Study Aim and Objectives

2.1 Aim

This study aims to characterise antibody responses to two hypervariable *Plasmodium falciparum* protein families, RIFIN and STEVOR, with more focus on STEVOR, and to determine their role in the development of immunity to the infection.

i. Hypothesis

Antibodies against the STEVOR and RIFIN hypervariable proteins are important markers of immunity to *Plasmodium falciparum* infection.

2.2 Objectives

2.2.1 Objective one

To develop an optimised pipeline for the expression of STEVOR and RIFIN recombinant proteins and evaluate their antigenicity using human malaria endemic serum from Sub-Saharan African countries, as a proof of concept.

i. Hypothesis

Hypervariable protein targets, represent antigenic recombinants which are recognised by *Plasmodium falciparum* endemic samples, irrespective of the recombinant expression platform used.

2.2.2 Objective two

To develop a library of recombinant antigens representing the hypervariable domain of the STEVOR protein family.

i. Hypothesis

The variability in the protein sequences of STEVOR variants, along with the specific strains they originate from, can be systematically analysed to select specific variants for inclusion in a recombinant antigen library.

2.2.3 Objective three

To investigate the development of immunity to STEVOR variants in an age- and exposure-dependent manner, using the developed recombinant antigen library against serological samples from Sub-Saharan African populations characterised by various levels of malaria endemicity, stratified by age groups.

i. Hypothesis

Increased *P. falciparum* exposure increases the breadth of antibody responses to STEVOR protein members in an age- and exposure-dependent manner.

Chapter 3: Research Paper 1

Novel *Plasmodium falciparum* hypervariable erythrocyte surface expressed recombinant proteins and their potential as serological markers of infection exposure.

Hristina Vasileva^{1,2}, Ana Copo-Pizarro¹, Michael Ooko³, Elin Dumont¹, Lindsey Wu¹, Isaac Ssewanyana⁴, Anna Last^{2*}, Kevin K.A. Tetteh^{1,5*}

¹Department of Infectious Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

²Clinical Research Department, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

³Medical Research Council International Statistics and Epidemiology Group, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

⁴Infectious Diseases Research Collaboration, Uganda National Health Laboratory Services, Kampala, Uganda

⁵Medical Affairs Department, FIND, Geneva, Switzerland

*These authors share last authorship.

Key words: malaria, hypervariable, recombinant proteins, markers of infection, *Plasmodium falciparum*, STEVOR, RIFIN

Abstract

Malaria, caused by *Plasmodium spp.*, leads to significant morbidity and mortality, particularly in Sub-Saharan African countries where *Plasmodium falciparum* is the predominant infectious species. The pathogenesis of *P. falciparum* depends on multiple host and parasitic factors, one of which is the evasion of host immune response, due to antigenic variability, during the blood stage of infection. The understudied infected erythrocyte-expressed protein families, STEVOR and RIFIN, characterised by antigenic hypervariability, are associated with clinical outcomes of the infection and protective acquired immunity, based on their topology and localisation.

In this study, two molecular tag methods were used for the successful expression of members of the STEVOR and RIFIN protein families as recombinant proteins in *E. coli* expression system. Further, the antigenicity of these recombinants was established, and Ugandan cohort samples with various *P. falciparum* infectious statuses were analysed for the detection of antibody levels to the STEVOR and RIFIN recombinants. The seropositivity rates to these recombinants in different age groups were compared against already established short- and long-term markers of infection.

This study demonstrated age-dependent immunity acquisition against the tested recombinants, suggesting the potential use of STEVOR and RIFIN recombinants as novel markers of *P. falciparum* exposure in serosurveillance.

Due to the hypervariable nature of members of these protein families, it is proposed that further, more extensive research using a library of recombinant variants is needed to strengthen the conclusions made.

3.1 Introduction

Malaria is a vector-borne disease caused by infection with parasites from the genus *Plasmodium* and transmitted by *Anopheles* mosquito species, responsible for approximately 608,000 deaths and 249 million infectious cases in 2022, with the greatest burden of the disease felt in Sub-Saharan Africa (1),(2). There are five main parasite species which cause human malaria, of which *Plasmodium falciparum* (*Pf*) is the most virulent, responsible for over 99% of malaria cases, the majority of which occur in children under five years of age and pregnant women (1),(3). Despite the observed reduction of malaria cases between 2000 and 2015 due to malaria control strategies, cases have been increasing significantly since 2016 due to several factors such as the rise of drug-resistant parasites and vector insecticide resistance, coupled with issues of regional political instability (1),(4),(5),(6). Additionally, the World Health Organisation (WHO) estimated that interruptions of malaria control efforts and case management caused by the SARS-CoV-2 pandemic resulted in an additional 13.4 million cases (1).

Pathogenesis of malaria depends on multiple parasite and host factors, causing different clinical outcomes and disease severity (7). Some of the parasite's virulence is due to evasion of the human host immune system during the blood stage of the infection. *Pf* infected erythrocytes (IE) adhere to the endothelial wall of blood vessels, via cytoadherence, preventing the clearance of the IE from the blood stream. Another pathogenic mechanism of *P. falciparum* is the formation of rosettes, clumping uninfected erythrocytes around the IEs. (7). Sequestration and rosetting are characteristic *Pf* virulence factors, enabled by parasite derived proteins expressed on the surface of IEs. These proteins possess antigenic properties and antibodies produced following exposure are associated with acquired protective immunity to *Pf* malaria (8). Moreover, IE surface expressed antigens are associated with antigenic variability, termed

Variant Surface Antigens (VSA) (8) . These antigens are the protein products of multi-copy gene families, grouped into several protein families: *P. falciparum* Erythrocyte Membrane Protein 1 (*PfEMP1*), Repetitive Interspersed protein family (RIFIN), Sub-Telomeric Variable Open Reading Frame family (STEVAR), *P. falciparum* Maurer's cleft Two Transmembrane domain family (*PfMC-2TM*) and Surface-Associated Interspersed Protein family (SURFIN) (9). These proteins are translocated from the blood stage *Pf* parasites (trophozoites and schizonts) to the surface of the IE via protein trafficking through the Maurer's cleft, a parasite derived membranous structure, and are expressed on the IE membrane, protruding into the extracellular space (9).

RIFIN and STEVAR protein families are encoded by approximately 160 *rif* and 40 *stevor* gene copies per parasite, respectively, where one gene from each family is expressed at a time per individual parasite, following an infection stage depended successive expression (9),(11). Members of each family differ mostly in their hypervariable region, believed to be the only domain exposed to the circulation that possesses antigenic epitopes, until a study using peptide microarrays showed that in *Pf* infected populations, there were similar levels of serorecognition and seroreactivity to both the semi-conserved (SC) and the large hypervariable (V2) protein domains of STEVARs and RIFINs (12),(13). Moreover, both SC and V2 domains were found to be associated with *Pf* exposure and potentially with clinical outcomes of the infection (12). Seroreactivity and serorecognition to both protein families have been demonstrated to be age and exposure dependent, with high reactivity in adults compared to children and higher domain recognition in individuals with clinical malaria as opposed to sub symptomatic infections (12).

The kinetics and half-life of antibodies to specific *Pf* antigens varies depending on multiple factors, such as protein antigenic properties, human host age and genetics, seasonality, immunoglobulin class, and endemicity of the setting (14),(15). The antibody responses to

certain antigens have a short half-life and thus indicate recent exposure to infection (short-term markers), such as HSP40.Ag1 (Heat Shock Protein 40 Antigen 1), Hyp2 (Hypothetical protein 2), Etramp5.Ag1 (Early Transcribed Membrane Protein 5 Antigen 1), Etramp4.Ag2 (Early Transcribed Membrane Protein 4 Antigen 2) and GEXP18 (Gametocyte Export Protein 18). Some already established markers, such as Rh2.2030 (Reticulocyte-binding protein homologue) and EBAs (Erythrocyte-binding antigens), indicate *Pf* infection during the past six months and therefore are considered as moderate-exposure markers (16). Antibody responses to AMA1 (Apical Membrane Antigen 1), MSP1.19 (Merozoite Surface Protein 1), MSP2.Dd2 (Merozoite Surface Protein 2, Dd2 allele), and GLURP.R2 (Glutamate Rich Protein Region 2) indicate long-term exposure with antibodies persisting for years after infection, termed as long-term markers (16),(26),(31). The measurement of antibody levels against different types of immune markers is a useful tool to assess the time of infection in individuals or in a population. This level of information is of a high importance for the design of control programs to interrupt transmission in a specific setting (26),(41). In concordance with other molecular techniques, measurement of antibody levels can also inform about the immune profiles in individuals protected from severe malaria disease, essential information for new vaccine approaches (26),(42). High throughput multiplex platforms, like quantitative suspension array technologies (qSAT) used in this study, have been developed to allow sensitive and accurate estimates of time of exposure via serosurveillance, testing antibody titres against multiple targets simultaneously (30).

This study is demonstrating successful expression of isolated domains of members of the STEVOR and RIFIN protein families as recombinant proteins in *E. coli* expression system. Antibody reactivity to previously established long- and short-term markers of exposure have been used as juxtaposition to classify the nature of antigenicity of these new targets

(14),(15),(16). This study aims to fill in the gaps in *Plasmodium falciparum* proteome knowledge and to investigate possible new markers for measuring exposure to infection.

3.2 Methods

3.2.1 Ethics

This study has received ethical approval by the Ethics Committee of the London School of Hygiene and Tropical Medicine (LSHTM) (reference number: 21505). The serum samples used in this study come from the Program for Resistance, Immunology, Surveillance, and Modelling of Malaria in Uganda longitudinal cohort (PRISM), approved by the Ethics Committee of LSHTM (reference number: 15823) and the Research and Ethics Committee of the Makerere University School of Medicine in Kampala, Uganda (reference number: 2011-167), and the Mapping Malaria and Neglected Tropical Diseases on the Bijagos Archipelago of Guinea Bissau (DTNMaPa), approved by the Ethics committee of LSHTM (reference number: 22899) and the Comite Nacional de Eticas de Saude (CNES) in Bissau, Guinea Bissau (reference number: 076/CNES/INASA/2017).

3.2.2 Design and expression of STEVOR and RIFIN recombinant proteins.

Six STEVOR (*Pf3D7_1300900*, *Pf3D7_1254100*, *Pf3D7_0115400*, *Pf3D7_0300400*, *Pf3D7_0832000*, *Pf3D7_0832600*) and three RIFIN (*Pf3D7_0833200*, *Pf3D7_1041100*, *Pf3D7_0732400*) protein sequences from the 3D7 *Plasmodium falciparum* reference strain were initially selected, based on previously published work on peptide arrays (12). Protein sequences were downloaded from the PlasmoDB database (Plasmodium Genomic Database, RRID:SCR_013331), and multiple alignment of the sequences was performed for each protein

family using Clustal2X (17),(18). Additional inspection of the alignments was conducted manually to identify misalignments. Both protein families presented a similar structure containing a signal peptide (SP), a semiconserved domain (SC), genetically different between the families conserved domain (C), two hypervariable domains (V1 and V2) and two transmembrane domains (2TM), as well as a translocator signalling element (PEXEL motif) (Figure 1) (19). The SC and V2 domains of the amino acid sequences were isolated for both the STEVORs and RIFINs, using published data on domain architecture, and a transmembrane domain prediction server (TMHMM v.2.0) (20). The edited sequences were then realigned using the 'msa' package on R computational platform (v3.6.3; R Core team 2020) (21). The V2 and SC portions of three STEVORs (*Pf3D7_1300900*, *Pf3D7_0832000* and *Pf3D7_0832600*) and one RIFIN (*Pf3D7_0732400*) proteins were selected as representative sequences, based on divergence using protein sequence phylogenetic trees (IQtree) (22).

The selected SC and V2 domains were expressed as recombinant antigens in *E. coli*, as described elsewhere (23),(24). Briefly, the SC and V2 domains were cloned into the pET15b expression vector as dual His (N-terminal) and Myc (C-terminal) tagged recombinant constructs (Biomatik, USA). A duplicate set of expression constructs were also cloned into the pGEX-4T-1 expression vector as single-tagged GST (N-terminal) recombinant constructs (Biomatik, USA), resulting in a total of fourteen recombinant proteins. BL21(DE3) chemically competent *E. coli* (Trans, China) was transformed with either the pET-15b or pGEX-4T-1 plasmid constructs, each containing an Ampicillin (Amp) resistance cassette. The transformed bacteria colonies were cultured in ZY auto-induction media supplemented with 100 µg/ml Amp at 37°C, 150 rpm overnight (~16 hr) (24). The cells were pelleted at 7500 rfc, resuspended in 1xPBS and lysed using an LM20 microfluidizer (Analytik LTD, UK) under 18,000 psi pressure units. The dual His-Myc tagged recombinant proteins were affinity purified using an ÄKTA Pure purification system (Cytiva, USA) nickel chromatography. Bound proteins were eluted

using an imidazole gradient. The eluted His-Myc tagged recombinants were further concentrated using 3kDa ultra concentration units (Merck, Germany). The GST-tagged recombinant proteins were affinity purified using Glutathione Sepharose 4B beads and purified in batch mode. No further concentration of the GST tagged proteins was required. Both sets of recombinant proteins were quantified using the Bradford protein quantification assay (Bio-Rad, UK).

3.2.3 Serum pools

An ELISA assay was developed to evaluate the reactogenicity of the STEVOR and RIFIN recombinant proteins (25). A two-fold dilution series was prepared for each recombinant protein in coating buffer (1.59g/l Na₂CO₃ and 2.93g/l NaHCO₃) starting at 4 µg/ml, down to 0.5 µg/ml. The dilution series for each recombinant was assayed against six positive control serum pools in addition to a malaria naïve negative control pool (PHE: Public Health England), now known as the UK Health Security Agency (26). All sera were assayed at a 1/200 dilution. Following incubation, plates were developed using 3,3',5,5'-tetramethylbenzidine (TMB; Tebubio: #TMBW-1000-01) and read at 450 nm. Generating a specific serological control for this study was imperative since it was practically unknown which variants of the STEVOR and RIFIN families were serorecognised by the tested population, as well as the control pooled populations. The control serum pools tested for control selection were as follows. Two in-house control pools together with an international reference standard were: *P. falciparum* hyperimmune serum pool based on Tanzanian adults positive for *Pf* infection (CP3), Gambian adults hyperimmune control pool (Brefet) and the 10/198 WHO 1st international *Pf* serum standard based on Kenyan adult samples (WHO) (27). Three additional pooled controls added to the panel were: a pool of 40 Ugandan serum samples from individuals aged above 18 years

with confirmed infection using Loop-mediated isothermal amplification (LAMP) and confirmed clinical disease (PRISM1), the highest quartile of tested individuals from Uganda of variable ages with calculated percentile antibody reactivity against in-house established markers of *Pf* infection (PRISM2, n=80), and the highest quartile of tested individuals from the Bijagos Archipelago, Guinea-Bissau of above 18 years with confirmed infection according to *Pf* 18S qPCR, with calculated percentile antibody reactivity against already tested in-house established markers of *Pf* infection (DTNMaPa, n=40) (28). Highest population quartiles of antibody reactivity for PRISM2 and DTNMaPa were calculated based on median fluorescence intensity (MFI) data of antibodies against the recombinant proteins *Pf*AMA1, *Pf*MSP1.19, GLURP.R2, Etramp5.Ag1 and HSP40, obtained from MagPix multiplex bead-based assay (29), (Vasileva *et. al.*; DTNMaPa unpublished data).

3.2.4 Detection of optimum recombinant protein antigenicity

After confirming protein seroreactivity using ELISA, the recombinant proteins were chemically coupled to MagPlex microsphere beads (Luminex) using an eight-fold six-point dilution series according to established methods (30). Briefly, titration of the coupled beads was performed against five-point two-fold dilutions of two positive control pools down selected from the ELISA data (CP3 and PRISM1) and a negative (PHE) control serum pool starting from 1/200 down to 1/1600 serum dilutions. MFI data was used to calculate the 50% maximum effect of antibody titres, also known as the EC50 point on a sigmoidal curve. Four parameter logistic regression was used to calculate the EC50 point per dilution curve, and the median EC50 point across all dilutions of both controls for each recombinant was selected as an optimum bead coupling concentration.

3.2.5 Study samples

Serum samples used for this study were collected during the first series of the cohort studies from the Program for Resistance, Immunology, Surveillance and Modelling of Malaria in Uganda (PRISM), conducted between October 2011 and June 2016. The samples come from Nagongera, Tororo district in Uganda, a rural setting with the highest transmission intensity district in Uganda. Malaria control interventions were applied from 2014 following the Ugandan National Malaria Control Program policy, consisting of insecticide treated bed nets, artemisinin therapies, sulfadoxine-pyrimethamine treatment during pregnancy and the introduction of indoor residual spraying with bendiocarb every six months between December 2014 and February 2015. For this study samples from two timepoints were selected: 2013 (T1) pre-interventions samples and 2016 (T4) post-interventions samples reflecting paired samples coming from a timepoint of high endemicity and a lower endemicity timepoint as a result of the applied interventions. A total of 505 samples, collected from April to July for both timepoints were selected, to include the highest peak of malaria transmission per year (June-July). Samples were stratified in three age groups: 6 months to 5 years, 5 years to 11 years and over 18 years old. Study participants were closely monitored to identify malaria cases via passive (participants seeking free healthcare when feeling ill), and active surveillance (routine visits every 90 days) tested by both microscopy slides and loop-mediated isothermal amplification (LAMP).

3.2.6 Multiplex bead-based serological assay

MagPlex magnetic beads were coupled with the STEVOR or RIFIN recombinant proteins, with one expression system per recombinant selected, summarised in Table 1. The recombinants were complemented in the Luminex assay with additional markers of seroprevalence, summarised in Supplementary Table 1 (31),(26),(16). Serum samples were diluted at 1/400 in

antibody elution buffer and incubated overnight (30). The Luminex serology assay was performed following an optimised standard operating procedure, as described previously (30). Briefly, serum samples were incubated with the antigen coupled beads, washed with 1xPBS/T, then incubated with Goat anti-human IgG R-phycoerythrin (R-PE) labelled antibody (Jackson ImmunoResearch©) for signal detection. A six point five-fold dilution series, starting from 1/10 down to 1/31250 for CP3 and PRISM1 positive controls, were run per plate for obtaining standard curve values used for plate-to-plate variation quality control of the assay. Additional controls included two wells of PHE malaria naïve negative controls diluted 1/400 and two wells of background control containing antibody elution buffer. The plates were read using a MagPix© bioanalyzer and data was obtained in the form of MFI, a proxy measure of antibody titres. MFI data was background adjusted and data quality checked by comparing the control standard curves and using Levy Jennings plots of mean MFI data for assessing plate-to-plate variations (Supplementary Figures 3.1-3.4). Any plate with data falling outside of the accepted variation of mean plus/minus 3 standard deviations was repeated (30).

3.2.7 Data analysis

Antigen-specific seropositivity was calculated using mean MFI data plus three times the standard deviation of the PHE negative serum control per antigen. Samples above the threshold are termed seropositive and percentage seropositivity per antigen per age group was displayed for both sample time points, 2013 and 2016 as a figure using R version 4.2.3 statistical software. The three samples age groups selected were as follow: “6 months – 5 years”; “5 years - 11 years” and “above 18 years”. The proportion percentage seropositivity, accounting for the differences in samples size between the two time points, was used for calculating the odds ratio (OR) of being seropositive in 2016 compared to 2013, briefly as follows. Logistic regression for binomial data distribution model was built to calculate the log odds of being seropositive in 2016 over being seropositive in 2013, calculated using year and age as covariate, including

the variables age and year interaction. The logistic regression was adjusted for clustering at the individual level, using robust standard error method, to account for repeated samples from the same individuals across multiple years. Data was exponentiated to obtain OR and confidence intervals (CI). The logistic regression model was built using “glm” function from the “stats” package and displayed as a forest plot using “forestplot” package in R version 4.2.3 statistical software.

3.3 Results

3.3.1 Design and expression of STEVOR and RIFIN recombinant proteins

Dissecting the SC and V2 domains (Figure 1) was critical for successful expression of the recombinants in an *E. coli* system, as incorporating malarial proteins specific domains such as signal peptides and transmembrane domains was likely going to result in protein aggregation into bacterial inclusion bodies, resulting in insoluble and dysfunctional recombinants (32). All targets were expressed using both types of plasmids (dual His-Myc tag and single GST tag). For each of the proteins, one type of recombinant was selected based on multiple factors such as volume of successful protein expression, concentration of protein after purification and seroreactivity according to the titrations against the controls, summarised in Table 1. GST tagged recombinants showed higher reactivity compared to the His-Myc tagged recombinants (shown in the titration plots in Supplementary Figure 1). The nature of the recombinants could explain this in terms of semi-conserved region versus large variable domain, where the SC is expected to be more serorecognised as this region is more conserved between variants, thus higher reactivity is expected. Additionally, to correct for any potential reactivity to the GST tag itself, MFI values against the GST tag alone per sample could be subtracted from the total sample MFI data. Although, this was not necessary as the seropositivity analysis to GST showed that none of the samples have MFI values above the seropositive threshold (shown on Supplementary Table 2). Furthermore, the His-Myc dual tag recombinants show some level of precipitation after protein purification, which is nullified in the GST-tag expression system since the GST molecule acts as a solubility factor (33). Despite the molecular differences of their weight, 4.5-8.8 kDa for His-Myc and 31.7-40.3 kDa for GST, both tag systems were comparable in terms of recombinant concentrations (Table 1).

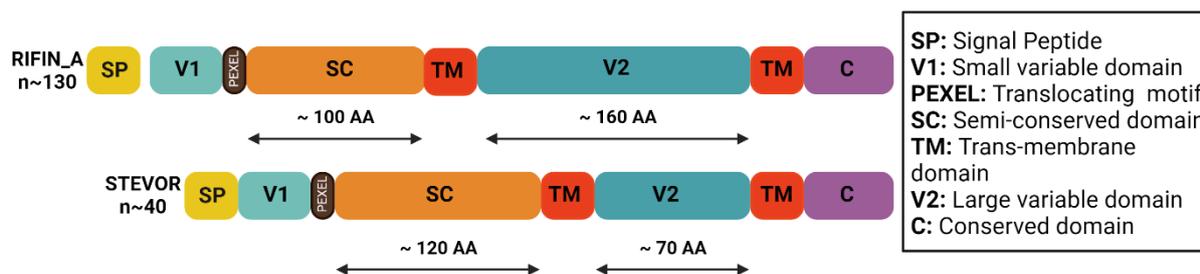


Figure 1: General structure of the RIFIN and STEVOR proteins showing the domain architecture of the proteins and the approximate length in amino acids of their variable domains (V2) and semi-conserved domains (SC). Transmembrane regions were identified using the TMHMM v.2.0 transmembrane domain prediction server. (Figure adapted by Zhoe E. Albert *et. al* (2019), *mSphere*; Created by Vasileva H. with BioRender.com)

Table 1: Summary of recombinant STEVOR and RIFIN proteins selected.

Reference Name	Recombinant Name ^a	Molecular Tag	Mol. weight (kDa)	Concentration (µg/mL)	EC50 ^b (µg/mL)
Pf3D7_1300900	STEVOR1_SC	GST	38.02	593.30	3.70
Pf3D7_0832600	STEVOR6_SC	GST	38.55	1049.20	12.85
Pf3D7_0353200	STEVOR5_SC	GST	38.71	1109.10	2.95
Pf3D7_0732400	RIFIN3_V2	GST	40.26	1193.30	7.05
Pf3D7_1300900	STEVOR1_V2	His-Myc	4.44	842.30	3.30
Pf3D7_0832600	STEVOR6_V2	His-Myc	4.66	448.20	2.90
Pf3D7_0732400	RIFIN3_SC	His-Myc	4.90	504.30	14.00

^a Semi-conserved domain recombinant (SC); Large variable domain recombinants (V2)

^b Protein titration plots for the selection of optimum protein concentration (EC50 point of saturation) are displayed as Supplementary Figure 1.

PRISM1 is a pool of 40 LAMP positive Ugandan samples that had the highest reactivity against all recombinants compared to the rest of the tested serum pools, as shown in Figure 2. Moreover, the tested samples in this study were from the same geographical population as the selected control pool, making it the best control candidate for the study. As per the

recombinants, there is an overall higher reactivity to the SC domains compared to the V2 domains, more prominently observed in STEVOR1 and RIFIN3, shown in the last plot in Figure 2.

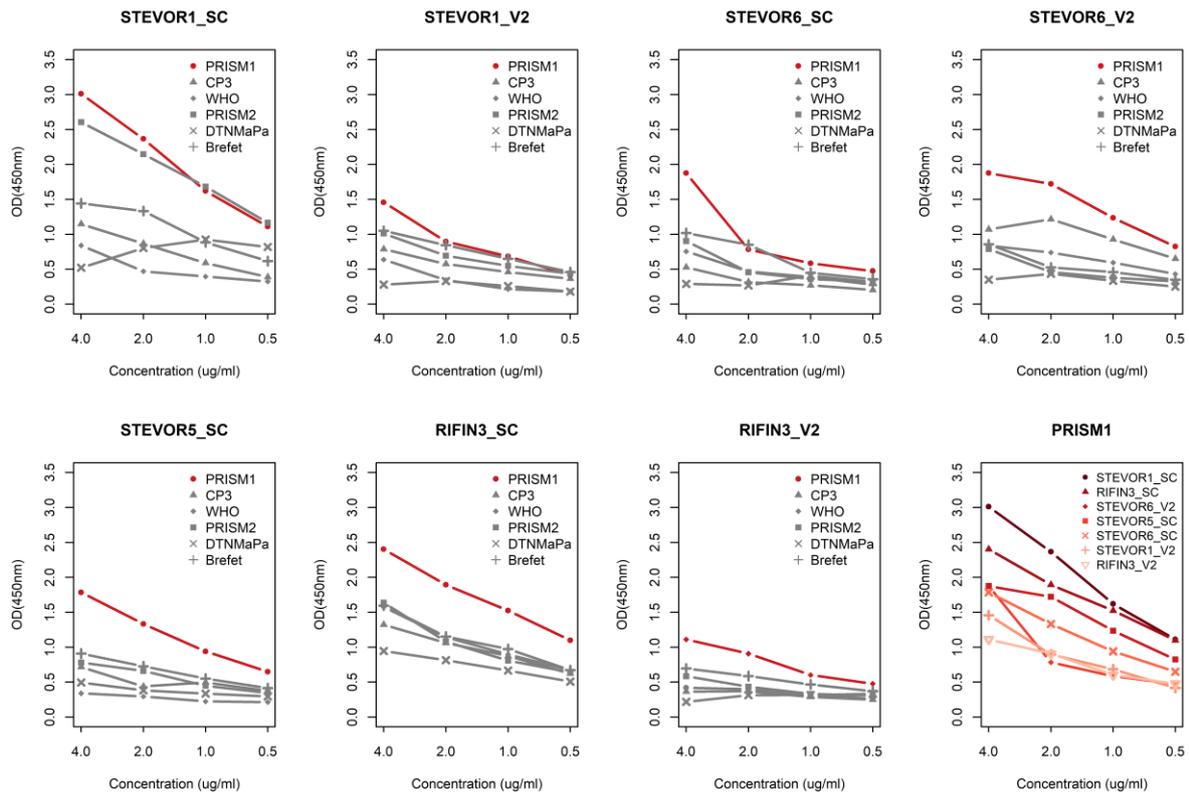


Figure 2: ELISA results of control serum pools (1/200 serum dilution), challenged against four steps six-fold dilution series of each recombinant protein. Brefet: Serum pool of hyperimmune adults from The Gambia; CP3: LSHTM in house serum pool of hyperimmune Tanzanian adults; DTNMaPa: Serum pool of 40 *Pf* infection confirmed adults, with high antibody responses to *Pf* markers of exposure from Bijagos Archipelago, Guinea-Bissau; PRISM1: Serum pool of 40 Ugandan adults with confirmed *Pf* infection and clinical disease; PRISM2: Serum pool of 80 Ugandan all age individuals with high antibody responses to established *Pf* markers of exposure; WHO: WHO 1st international *Pf* serum from Kenyan adult population. The control serum selected from this experiment is PRISM1, indicated in red.

The last plot of the series demonstrates the seroreactivity of all recombinants against PRISM1, with color-coded intensity in descending order.

3.3.2 Sample summary statistics.

There was a decrease of 37% of samples used in the study between 2013 and 2016 from 310 to 195, respectively, due to a change of the regime in sampling from all samples collected at two time points per year in 2013, to collection of samples over a period of 3 months in 2016 from the same study participants, indicated in Supplementary Figure 2. The decrease in malaria cases confirmed using microscopy and LAMP from 217 to 40 cases, accounting for more than 50% decrease, summarised in Table 2, is explained by the introduction of the control interventions in the population in 2014, as demonstrated in previous studies (4),(34). A paired sample t-test indicated a significant difference of malaria prevalence per age group between the time points with a p-value of < 0.001. The uneven distribution of sample size between the different age groups observed for both time points was due to limited availability of PRISM samples, particularly samples belonging to the above 18 years age group with only one *Pf* positive sample for 2016.

Table 2: Summary statistics of the longitudinal cohort PRISM study population.

	2013, n (%)	2016, n (%)	Total, n (%)
Sample size	310 (61.39)	195 (38.61)	505 (100.00)
Age strata			
6 months – 5 years	71 (22.90)	45 (23.08)	116 (22.97)
5 years – 11 years	152 (49.03)	139 (71.28)	291 (57.62)
> 18 years	87 (28.06)	11 (5.64)	98 (19.41)
Parasitemia positive by age group *			
6 months – 5 years	49 (69.01)	6 (13.33)	55 (10.89)
5 years – 11 years	116 (76.32)	33 (23.74)	149 (29.50)
> 18 years	52 (59.77)	1 (9.09)	53 (10.50)
Total	217 (70.00)	40 (20.51)	257 (50.89)

***Parasitaemia positive refers to participants positive for *P. falciparum* measured by LAMP or microscopy, or both, with significant difference of malaria prevalence (%) across age groups between the two sampling time points. A large proportion of samples (74%) in 2013 are paired to those in 2016, graphically displayed in Supplementary Figure 2.**

3.3.3 Age-related seropositivity

There was an overall reduction in seroprevalence for the majority of the panel antigens tested, as well as the STEVOR and RIFIN recombinants, between 2013 and 2016, most prominently observed in the “6 months to 5 years” and “5 years to 11 years” age groups, as shown in Figure 3. The exception was RIFIN3_SC recombinant with higher seroprevalence in 2016 samples compared to 2013. However, this trend was not reflected in the “above 18 years” age group, potentially confounded by the large samples size difference between the two time points with only one confirmed malaria positive sample in 2016, compared to 52 samples in 2013. For the 2013 samples (pre-intervention), seroprevalence for all short-term markers increased with age. For all the long-term markers of exposure, seroprevalence increased with age at both timepoints. Responses to MSP2.Dd2 in 2013 followed the same tendency but by contrast, in 2016, there was a decline of seroprevalence with increasing age, although not significant. Higher seroprevalence was observed against the tested recombinants semiconserved domains, particularly for STEVOR1_SC, except for RIFIN3_SC, which presented with lower seroprevalence compared to the variable region. For the youngest age group there was little to no seroprevalence against all STEVOR and RIFIN variable domains compared to their semiconserved regions, possibly reflecting a lack of exposure to the variant types. Seropositivity thresholds are summarised in Supplementary Table 2, and absolute values to the relative percentage values are summarised in Supplementary Table 3. Extremely low to no seroprevalence to tetanus toxoid (Tet.tox) after calculating the threshold serves as a control for the assay. Given that 97% of the global population is vaccinated against *Clostridium tetani*, all samples, including those from malaria-naïve individuals (PHE samples) are expected to show

seropositivity to the toxoid, thus applying a threshold of mean PHE MFI + 3SD is high considering expected high seroreactivity will present with a threshold high enough to encompass all samples within.

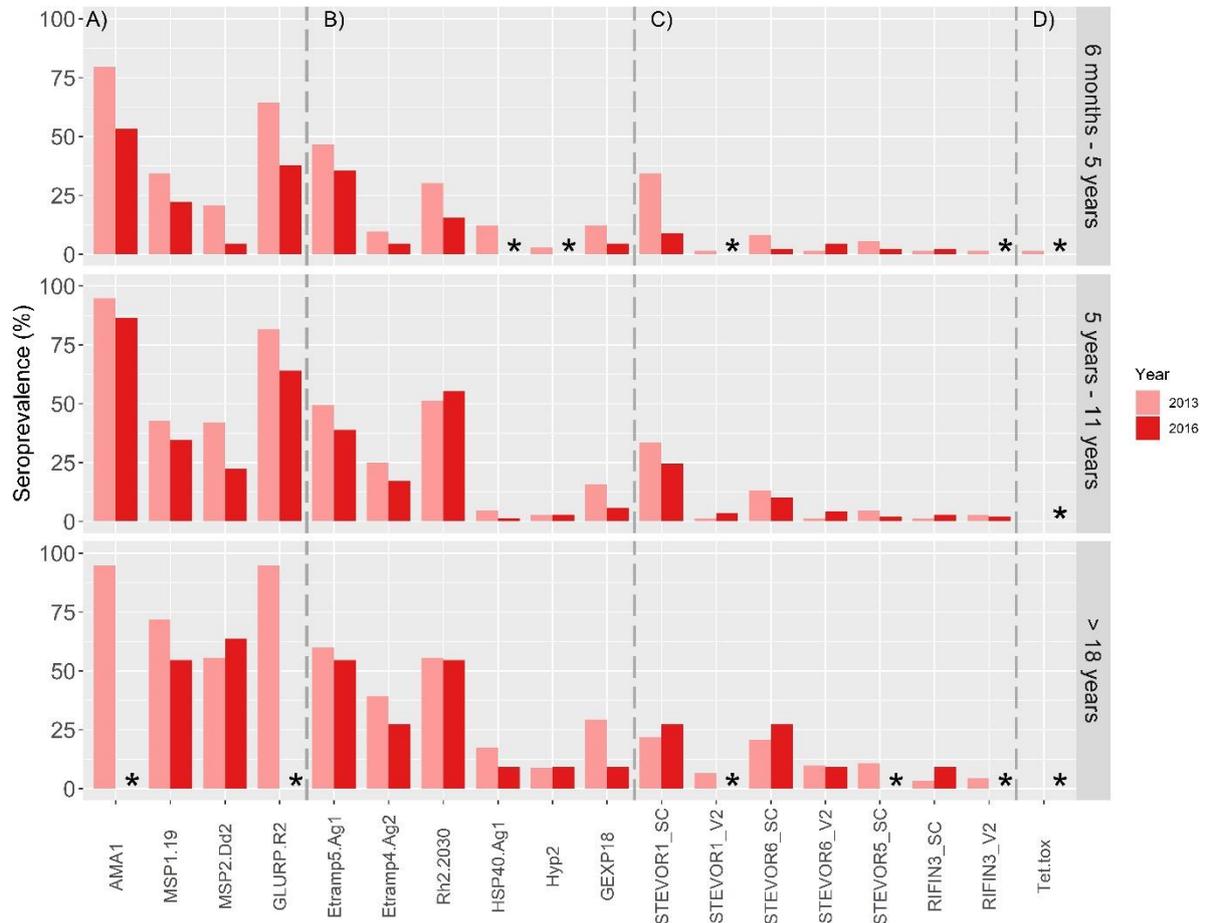
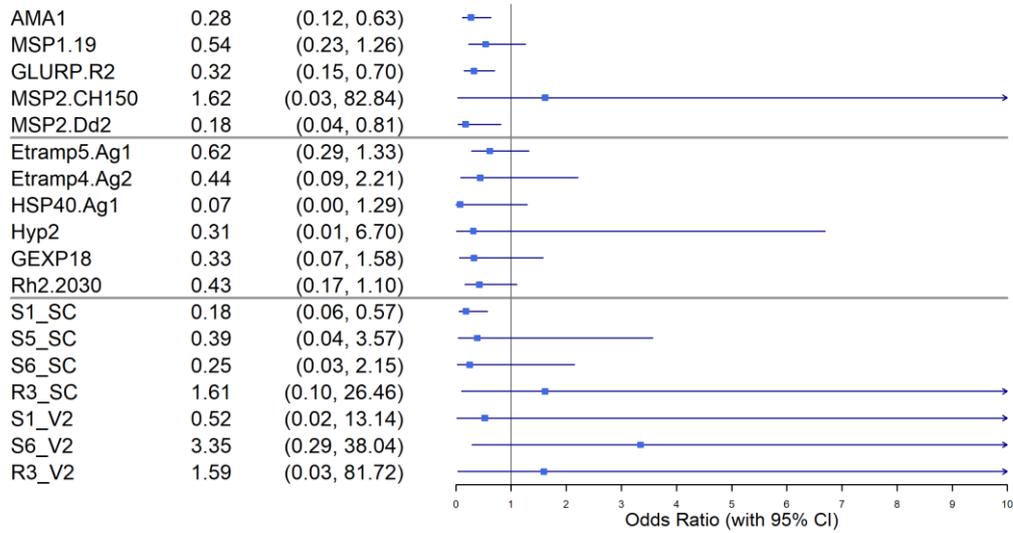


Figure 3: Population seroprevalence (%) per survey year for A) long-term markers; B) short, medium-term markers C) STEVOR/RIFIN recombinants and D) internal serology control, stratified by age group. Seroprevalence threshold per recombinant was selected using a cut off mean MFI from PHE negative samples plus three times the standard deviation. Bar plots are paired colour coded, for visual representation of differences in seroprevalence between the two surveyed years. Tet.tox represents the internal positive human serological control. Seroprevalence of 0% is indicated with an asterisk and exact seroprevalence percentages are summarised in Supplementary Table 2.

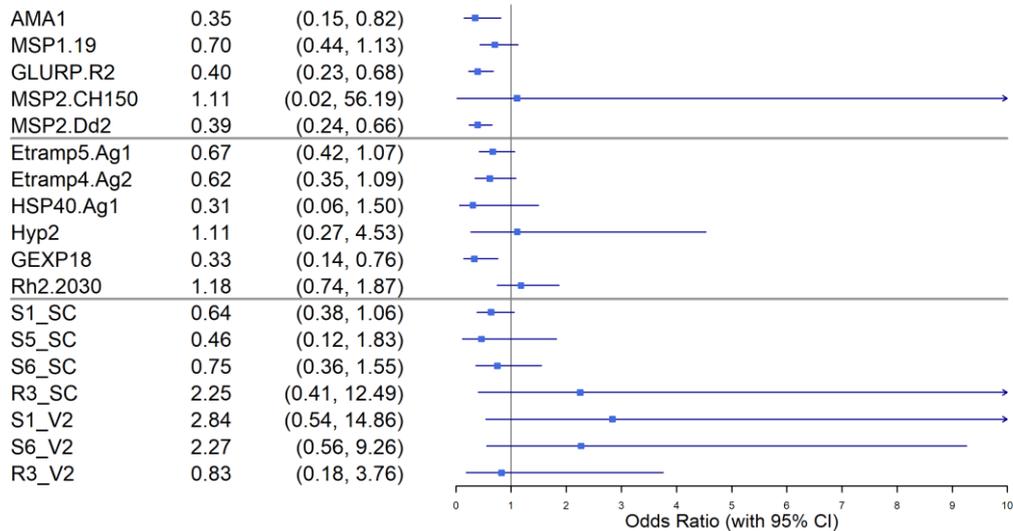
The odds of being seropositive in 2016 over 2013 increased with age for all markers of infection as well as the STEVOR/RIFIN recombinants with the exception of GEXP18 (6 months – 5 years: OR = 0.33, CI = 0.07 - 1.58; 5 years – 11 years: OR = 0.33, CI = 0.14 - 0.76; above 18 years: OR = 0.24, CI = 0.03 - 1.94) and R3_V2 (6 months – 5 years: OR = 1.59, CI = 0.03 - 81.72; 5 years – 11 years: OR = 0.83, CI = 0.18 - 3.76), as summarised in Figure 4. This trend is consistent with the understanding that acquired immunity to malaria is both age- and exposure-dependent. Adults, having been exposed to *P. falciparum* infection for longer periods, are more likely to have developed immunity, resulting in higher seropositivity rates (having detectable antibodies against specific infection markers). In contrast, children, due to their lower cumulative exposure to the parasite, tend to have lower antibody levels and, consequently, lower seropositivity rates.

For the tested STEVOR and RIFIN recombinants, the odds of seropositivity also increased with age, although this trend was less pronounced compared to the short- and long-term markers of exposure. It can be noted that the precision of the OR estimates for the STEVORs and RIFINs is generally lower, as illustrated by the large confidence intervals. This trend is also observed for the majority of markers of exposure in the above 18 age group. These trends can be attributed to the overall lower number of seropositive individuals for STEVORs and RIFINs, as well as the small sample size of the above- 18 age group, outlined as a limitation of the study.

A) 6 months - 5 years



B) 5 years - 11 years



C) > 18 years

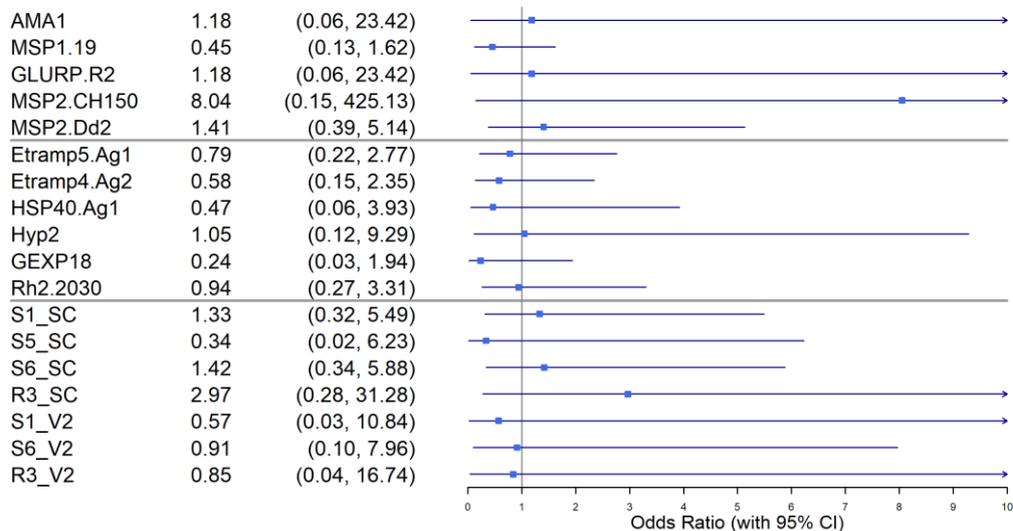


Figure 4: Forest plot of odds ratio analysis per recombinant antigen of being seropositive in 2016 compared to 2013, calculated in seroprevalence (%), stratified by age: A) 6 months to 5 years old, B) 5 years to 11 years old, and C) more than 18 years old. Squares graphically represent the calculated and summarised in the table's ORs with 95% confidence intervals (CIs) as whiskers. CIs of above 10 are not graphically represented but are displayed in the table on the left. Grey lines separate the antigens into three groups: long-term markers of infection, short-term markers of infection and STEVOR/RIFIN recombinants from top to bottom.

3.4 Discussion

This study demonstrates methods for expression of STEVOR and RIFIN protein domains as recombinant proteins, exploring their potential use as serosurveillance tools by comparing them to previously confirmed markers of *P. falciparum* exposure, using the multiplex suspension technology: Luminex (16),(26),(31).

The study is comparing two methods of recombinant expression in *E. coli* bacterial systems: a single GST tag plasmid and a dual His-Myc tag plasmid technology. The SC recombinants tend to express better with the single tag GST system, and because of their higher conservation between variants, it is expected higher serorecognition compared to the V2 recombinants (35). The dual tag recombinants represent the V2 domains of selected STEVOR members which are expected to generate comparatively lower antibody response due to antigenic variation and the lack of knowledge if the samples tested in this study have been exposed to these specific variants. This trend is also observed in the ELISA results of the tested control pools. Variations in protein expression even within a single tag system is an established factor, clearly demonstrated in this study where the STEVOR SC and RIFIN V2 domains expressed better with a GST tag system, and the His-Myc system was preferred for the STEVOR V2 and RIFIN SC domains (36),(37). Some of the reactivity detected could potentially be against the GST tag molecule alone, a naturally occurring enzyme molecule in the parasite *Schistosoma japonicum* (38). There was no data available about past infections of *Schistosoma* for the participants used in this study, hence it was impossible to determine if there was residual antibody reactivity to the GST tag. To eliminate this as a possibility, a GST recombinant alone was included in the recombinant antigen panel for potentially subtracting reactivity values detected against the tag alone. Moreover, using UK control serum for calculating the threshold can overestimate the seroprevalence, since it is expected higher background reactivity of African endemic samples to GST, compared to UK control samples of individuals who never travelled to Africa. To

mitigate for this overestimation a stringent threshold of mean plus three standard deviations was used which was above the highest read out of endemic samples against the GST tag recombinant antigen alone. Therefore, a decision not to subtract the GST seroreactivity from the detected MFI values against the GST tagged recombinants was made as after performing the seropositivity threshold analysis there were no antibody responses against GST detected above the threshold for any of the samples (data available in Supplementary Table 2). The GST tag is used as a solubility factor which perhaps could influence the performance of the recombinants (33). This argument is supported by the detected precipitation of the His-Myc dual tag recombinants. Nonetheless, there was no significant difference in protein yield when comparing both systems, thus a decision that the tag expression system did not significantly affect the production of recombinants was made and the recombinants with the system resulting in higher yield was individually selected per recombinant. Using two different protein expression tag systems can be regarded as a limitation of the study. However, as these were not conformation proteins, there was little need to explore other more disparate expression systems such as yeast or wheat germ. The proposed construct design was based on expertise within the laboratory group and lent itself well to expression of the targets irrespective of the tag system used (39),(40). Moving forward, the optimum solution will be to focus on only one expression system.

Tested samples selected from the two time points of the PRISM longitudinal cohort, 2013 and 2016, were predominantly paired samples, with 74% of the 2016 samples paired with those from 2013. Thus, observing differences in antibody responses to the recombinants before and after intervention is done on the same population, where some of the samples would move from one age group to another between 2013 and 2016. This sample selection allows for an evaluation of the reduction in seroreactivity and serorecognition over time due to the implementation of interventions, leading to a statistically significant (paired t-test p-value <

0.001) reduction in *P. falciparum* malaria prevalence in 2016. It is important to note that following the implementation of interventions, a decrease in seroprevalence is expected due to reduced exposure to *Plasmodium falciparum*. This trend is evident in the seropositivity analysis, where a decline in seroprevalence to all markers of exposure across all age groups was observed (43). However, the seropositivity analysis shows higher population seroprevalence to a few of the variant recombinants (STEVOR1_SC, STEVOR6_SC, RIFIN3_SC) in 2016 as opposed to 2013 for the “above 18 years” age group. There are a few possible explanations for this observation. These recombinants represent the semi-conserved regions of the variants to which higher serorecognition is expected due to high domain conservation, leading to cross reactivity between variants (35). In addition, the low sample size of the adult age group in 2016 can significantly impact the calculation of seroprevalence and introduce bias to the analysis and interpretation of the data (44). Consequently, the focus of further discussion is directed mainly on the results from the other two age groups, which in turn is with a higher importance to the posed question since particularly children under 5 years of age have a naïve immune system and changes in antibody responses due to external or internal pressure can be detected with higher precision and sensitivity. As well as for the 5 years to 11 years of age group it is known to be the age group with the most exposure to infection (45). Another potential limitation was the fact that despite the selected high reactivity serum as a control, it is still possible that some of the recombinant’s reactivity is not captured due to their hypervariability nature.

Calculated seropositivity on the antigen-specific responses between time points and across age groups were directly compared without adjusting for total IgG. Adjusting for total IgG was not necessary because seropositivity thresholds inherently account for baseline variability in the dataset, making the additional IgG adjustment redundant. The additional adjustment would have been necessary if there was a significant variability in total IgG levels in the populations

due to factors such as age, nutritional status, or immunosuppression, as these could confound antigen-specific measurements. It is also important when comparing populations with differing baseline immune profiles or studying immunocompromised groups to ensure fair and unbiased comparisons. In the case of this study the comparison of seropositivity was done on paired samples coming from the same individuals for both time points.

While few studies have been conducted analyzing the humoral response of STEVORs and RIFINs, similar to *PfEMP1*, IgG antibody responses have been correlated with age and it has been observed that children with high anti-RIFIN and anti-STEVOR antibody titers had a reduced risk of febrile malaria (46). Consistent with these findings, there is an observed increase of seroprevalence to STEVORs and RIFINs with age, most pronounced in the case of STEVOR6_SC. Furthermore, all age groups have some seroprevalence calculated against the semi conserved recombinants, where for the variable domain recombinants the younger age groups present with 0% seroprevalence, perhaps reflecting the lack of exposure to these variants. The fact that infants lose their maternal antibodies after 6 months of age could explain these results, as the variable domain forms a surface-exposed loop that is highly variable due to a high recombination rate, a mechanism that the parasite uses for immune system evasion (13)(48). Their location in the sub-telomeric region of the parasite's chromosomes facilitates the mechanisms that induce hypervariability within genes, playing a role in *P. falciparum*'s antigenic variation (13)(49). Apart from their high variability, they are under the host's immune pressure which amplifies the variable repertoire of these proteins (49)(50).

Since the SC domain has higher likelihood of recognition compared to V2, it is possible that contribution to the acquisition of NAI is restricted to the semiconserved region. However, the current data does not suggest that anti-STEVOR and anti-RIFIN antibodies play a major role in immunity, further studies comprised of a larger library of variants from the STEVOR and

RIFIN protein families are needed. Additionally, the role of anti-STEVEOR antibodies in immunity can be tested through functional assays, such as invasion or growth inhibition assays, and passive transfer studies in animal models or humans to evaluate their protective potential. Furthermore, longitudinal or case-control studies in endemic populations, antibody depletion experiments, and vaccination models can help determine their functional significance and correlation with protection against malaria.

The odds ratio (OR) analysis showed that odds of increased seropositivity in 2016 compared to 2013 is greater in the older age groups. These observations are all expected due to the known age and exposure dependent acquired immunity to *P. falciparum* exposure (14)(16). Nevertheless, all markers of exposure as well as the tested novel recombinants show higher odds of seropositivity with increasing age. However, a key limitation of the study is the overall low precision of odds ratio estimates for STEVEOR and RIFIN recombinants and for all markers of exposure for the above 18 years age group, driven by the overall low number of seropositive individuals for the tested recombinants and the small sample size in the adults age group. To mitigate this limitation, more STEVEOR and RIFIN recombinants should be tested, and the sample size should be increased. However, this was unavoidable due to the limited availability of residual samples from the PRISM study. This study reports STEVEOR and RIFIN variants expressed as antigen recombinants using various *in silico* tools and an *E. coli* bacterial expression system. Our data shows that tested STEVEOR and RIFIN recombinants follow the same trends of seropositivity to the already established panel of antigen markers of exposure, thus we propose that members of these protein families could potentially be regarded as novel markers of *P. falciparum* exposure in serosurveillance (16)(26). However, due to the hypervariability between variants, further studies comprised of a larger library of variants representing the STEVEOR and RIFIN families should be compiled to strengthen the findings of this study.

3.5 Funding

This project is nested within the clinical trial “Adjunctive Ivermectin Mass Drug Administration for Malaria Control (MATAMAL)” (ClinicalTrials.gov Identifier: NCT04844905), funded by the Joint Global Health Trials Scheme (Grant Code MR/S005013/1). The scheme is jointly funded by the Medical Research Council (MRC), Foreign, Commonwealth and Development Office (FCDO), the National Institute for Health Research (NIHR) and the Wellcome Trust.

3.6 Acknowledgements

Special thanks to Dr Christian Kositz from Swiss TPH for analytical advice, as well as the data science department in the MRC Unit The Gambia @ LSHTM.

3.7 Conflict of interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

3.8 List of non-standard abbreviations

- *Pf. Plasmodium falciparum*
- IE: Infected Erythrocyte
- VSA: Variant Surface Antigen
- *PfEMP1: Plasmodium falciparum* Erythrocyte Membrane Protein 1
- RIFIN: Repetitive Interspersed Family
- STEVOR: Subtelomeric Variable Open Reading Frame family
- *PfMC-2TM: Plasmodium falciparum* Maurer's Cleft 2 Transmembrane family
- SURFIN: Surface Associated Interspersed family.
- SP: Signal Peptide
- SC: Semi Conserved domain
- V1, V2: Variable domain 1, Variable domain 2
- HSP40.Ag1: Heat-Shock Protein 40
- Hyp2: Hypothetical Protein 2
- Etramp5.Ag1: Early Transcribed Membrane Protein 5
- Etramp4.Ag2: Early Transcribed Membrane Protein 4
- GEXP18: Gametocyte Export Protein 18
- Rh2.2030: Reticulocyte Binding Protein Homologue 2
- EBAs: Erythrocyte Binding Antigens
- AMA1: Apical Membrane Antigen 1
- MSP1.19: Merozoite Surface Antigen 1
- MSP2.Dd2: Merozoite Surface Antigen 2, Dd2 allele
- MSP2.CH150: Merozoite Surface Antigen 2, CH150 allele
- GLURP.R2: Glutamate-Rich Protein, Region 2

- PRISM: Program for Resistance, Immunology, Surveillance and Modeling of Malaria in Uganda
- DNTMaPa: Malaria and Neglected Tropical Diseases mapping survey, Guinea-Bissau.
- GST: Glutathione S-transferase protein
- LAMP: Loop-mediated Isothermal Amplification
- MFI: Median Fluorescence Intensity
- ITN: Insecticides Treated Nets
- IRS: Indoor Residual Spraying
- OR: Odds Ratio
- CI: Confidence Interval
- LILRB1: Leucocyte Immunoglobulin-Like Receptor B1
- LAIR1: Leucocyte-associated Immunoglobulin Receptor 1
- NK: Natural Killer cell

3.9 References

1. World Health Organization. World Malaria Report 2023 [Internet]. 2023. Available from: https://cdn.who.int/media/docs/default-source/malaria/world-malaria-reports/world-malaria-report-2023-spreadview.pdf?sfvrsn=bb24c9f0_4
2. Sutherland CJ, Tanomsing N, Nolder D, Oguike M, Jennison C, Pukrittayakamee S, et al. Two Nonrecombining Sympatric Forms of the Human Malaria Parasite *Plasmodium ovale* Occur Globally. *J Infect Dis* [Internet]. 2010 May 15 [cited 2020 Jun 17];201(10):1544–50. Available from: <https://academic.oup.com/jid/article-lookup/doi/10.1086/652240>
3. Wassmer SC, Taylor TE, Rathod PK, Mishra SK, Mohanty S, Arevalo-Herrera M, et al. Investigating the Pathogenesis of Severe Malaria: A Multidisciplinary and Cross-Geographical Approach. *Am J Trop Med Hyg* [Internet]. 2015 Sep 9 [cited 2023 Mar 13];93(3 Suppl):42. Available from: </pmc/articles/PMC4574273/>
4. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U, et al. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature* [Internet]. 2015 Oct 10 [cited 2023 Mar 13];526(7572):207. Available from: </pmc/articles/PMC4820050/>
5. Health Organization W. GLOBAL TECHNICAL STRATEGY FOR MALARIA 2016-2030 Global Technical Strategy for Malaria 2016-2030 Global Malaria Programme World Health Organization.
6. Le PVV, Kumar P, Ruiz MO, Mbogo C, Muturi EJ. Predicting the direct and indirect impacts of climate change on malaria in coastal Kenya. *PLoS One* [Internet]. 2019 Feb 1 [cited 2023 Mar 13];14(2). Available from: </pmc/articles/PMC6364917/>

7. Miller LH, Baruch DI, Marsh K, Doumbo OK. The pathogenic basis of malaria. *Nature* [Internet]. 2002 Feb 7 [cited 2021 Nov 17];415(6872):673–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/11832955/>
8. Chan J-A, Fowkes FJI, Beeson JG. Surface antigens of *Plasmodium falciparum*-infected erythrocytes as immune targets and malaria vaccine candidates. *Cell Mol Life Sci* [Internet]. 2014 Oct 2 [cited 2020 Mar 24];71(19):3633–57. Available from: <http://link.springer.com/10.1007/s00018-014-1614-3>
9. Scherf A, Lopez-Rubio JJ, Riviere L. Antigenic Variation in *Plasmodium falciparum*. *Annu Rev Microbiol* [Internet]. 2008 Oct 11 [cited 2019 May 20];62(1):445–70. Available from: <http://www.annualreviews.org/doi/10.1146/annurev.micro.61.080706.093134>
10. Baruch DI, Pasloske BL, Singh HB, Bi X, Ma XC, Feldman M, et al. Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. *Cell* [Internet]. 1995 Jul 14 [cited 2020 Oct 5];82(1):77–87. Available from: <https://pubmed.ncbi.nlm.nih.gov/7541722/>
11. Kyes SA, Rowe JA, Kriek N, Newbold CI. Rifins: A second family of clonally variant proteins expressed on the surface of red cells infected with *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* [Internet]. 1999 Aug 3 [cited 2020 Oct 5];96(16):9333–8. Available from: </pmc/articles/PMC17783/?report=abstract>
12. Zhou AE, Berry AA, Bailey JA, Pike A, Dara A, Agrawal S, et al. Antibodies to Peptides in Semiconserved Domains of RIFINs and STEVORs Correlate with Malaria Exposure. *mSphere*. 2019 Mar 20;4(2).
13. Catherine Lavazec, Sohini Sanyal and TJT. Hypervariability within the Rifin, Stevor and Pfmc-2TM superfamilies in *Plasmodium falciparum*. *Nucleic Acids Res* [Internet].

- 2006;34(22):6696–707. Available from:
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1751529/>
14. Akpogheneta OJ, Duah NO, Tetteh KKA, Dunyo S, Lanar DE, Pinder M, et al. Duration of naturally acquired antibody responses to blood-stage *Plasmodium falciparum* is age dependent and antigen specific. *Infect Immun* [Internet]. 2008 Apr 1 [cited 2020 Sep 29];76(4):1748–55. Available from: <http://iai.asm.org/>
 15. Boutlis CS, Fagan PK, Gowda DC, Lagog M, Mgone CS, Bockarie MJ, et al. Immunoglobulin G (IgG) responses to *Plasmodium falciparum* glycosylphosphatidylinositols are short-lived and predominantly of the IgG3 subclass. *J Infect Dis* [Internet]. 2003 Mar 1 [cited 2020 Sep 29];187(5):862–5. Available from: <https://pubmed.ncbi.nlm.nih.gov/12599061/>
 16. Drakeley CJ, Corran PH, Coleman PG, Tongren JE, McDonald SLR, Carneiro I, et al. Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. *Proc Natl Acad Sci U S A* [Internet]. 2005 Apr 5 [cited 2020 Sep 29];102(14):5108–13. Available from:
www.pnas.org/cgi/doi/10.1073/pnas.0408725102
 17. Amos B, Aurrecochea C, Barba M, Barreto A, Basenko EY, Bazant W, et al. VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. *Nucleic Acids Res* [Internet]. 2022 Jan 7 [cited 2023 Jan 5];50(D1):D898–911. Available from: <https://academic.oup.com/nar/article/50/D1/D898/6413610>
 18. Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics* [Internet]. 2007 Nov [cited 2020 Oct 9];23(21):2947–8. Available from:
<https://pubmed.ncbi.nlm.nih.gov/17846036/>

19. Andersson A, Kudva R, Magoulopoulou A, Lejarre Q, Lara P, Xu P, et al. Membrane integration and topology of RIFIN and STEVOR proteins of the *Plasmodium falciparum* parasite. *FEBS J* [Internet]. 2020 Jul 1 [cited 2023 Mar 13];287(13):2744–62. Available from: <https://pubmed.ncbi.nlm.nih.gov/31821735/>
20. Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* [Internet]. 2001 Jan 19 [cited 2023 Jan 5];305(3):567–80. Available from: <https://pubmed.ncbi.nlm.nih.gov/11152613/>
21. Bodenhofer U, Bonatesta E, Horejš-Kainrath C, Hochreiter S. msa: an R package for multiple sequence alignment. *Bioinformatics* [Internet]. 2015 Jul 3 [cited 2023 Jan 5];31(24):3997–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/26315911/>
22. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* [Internet]. 2020 May 1 [cited 2023 Jan 5];37(5):1530–4. Available from: <https://academic.oup.com/mbe/article/37/5/1530/5721363>
23. Flick K, Ahuja S, Chene A, Bejarano MT, Chen Q. Optimized expression of *Plasmodium falciparum* erythrocyte membrane protein 1 domains in *Escherichia coli*. *Malar J* [Internet]. 2004 Dec 15 [cited 2023 Jan 9];3:50. Available from: </pmc/articles/PMC544839/>
24. Studier FW. Protein Expression and purification 41 [Internet]. 2005. Available from: <https://tetramer.yerkes.emory.edu/support/protocols#2>
25. Wirtz RA, Duncan JF, Njelesani EK, Schneider I, Brown AE, Oster CN, et al. ELISA method for detecting *Plasmodium falciparum* circumsporozoite antibody*. 1989;

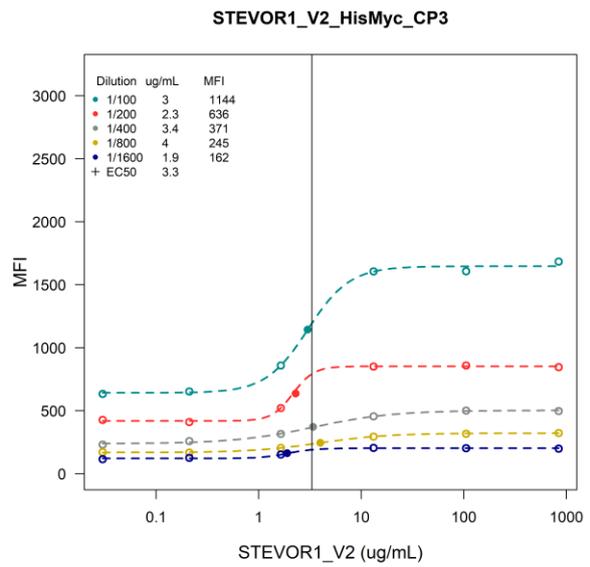
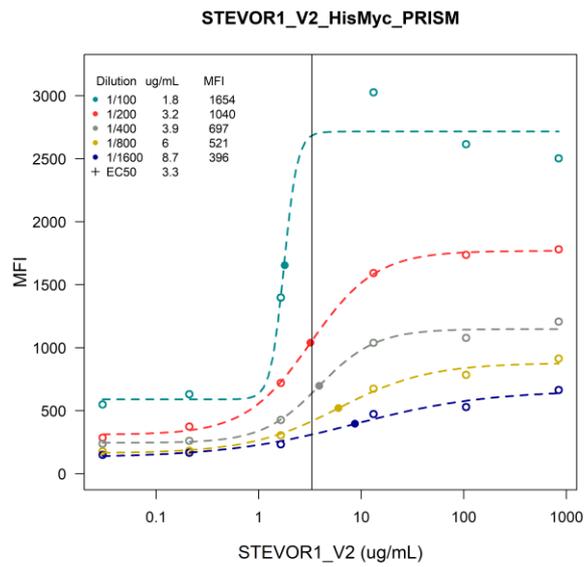
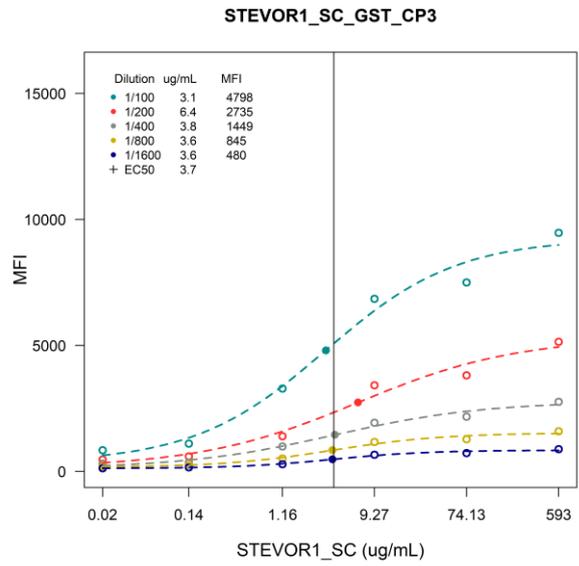
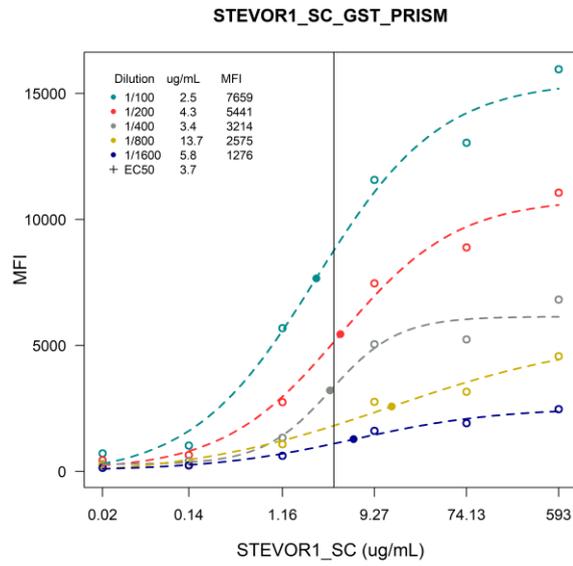
26. Wu L, Mwesigwa J, Affara M, Bah M, Correa S, Hall T, et al. Antibody responses to a suite of novel serological markers for malaria surveillance demonstrate strong correlation with clinical and parasitological infection across seasons and transmission settings in The Gambia. *BMC Med* [Internet]. 2020 Sep 25 [cited 2023 Mar 13];18(1):304. Available from: [/pmc/articles/PMC7517687/](https://pubmed.ncbi.nlm.nih.gov/34911111/)
27. Bryan D, Silva N, Rigsby P, Dougall T, Corran P, Mei +, et al. WHO/BS/2014.2235 ENGLISH ONLY EXPERT COMMITTEE ON BIOLOGICAL STANDARDIZATION Geneva, 13 to 17 October 2014 International collaborative study to evaluate and establish the 1 st WHO Reference Reagent for Anti-Malaria (*Plasmodium falciparum*) human serum. 2014;
28. Program for Resistance, Immunology, Surveillance & Modeling of Malaria in Uganda (PRISM) Renewal | Global Research Projects [Internet]. [cited 2023 Mar 13]. Available from: <https://globalprojects.ucsf.edu/project/program-resistance-immunology-surveillance-modeling-malaria-uganda-prism-renewal>
29. Ssewanyana I, Rek J, Rodriguez I, Wu L, Arinaitwe E, Nankabirwa JI, et al. Impact of a Rapid Decline in Malaria Transmission on Antimalarial IgG Subclasses and Avidity. *Front Immunol*. 2021 Jan 27;11:3308.
30. Wu L, Hall T, Ssewanyana I, Oulton T, Patterson C, Vasileva H, et al. Optimisation and standardisation of a multiplex immunoassay of diverse *Plasmodium falciparum* antigens to assess changes in malaria transmission using sero-epidemiology. *Wellcome Open Res* [Internet]. 2020 Apr 23 [cited 2020 Oct 15];4:26. Available from: <https://doi.org/10.12688/wellcomeopenres.14950.1>
31. Helb DA, Tetteh KKA, Felgner PL, Skinner J, Hubbard A, Arinaitwe E, et al. Novel serologic biomarkers provide accurate estimates of recent *Plasmodium falciparum*

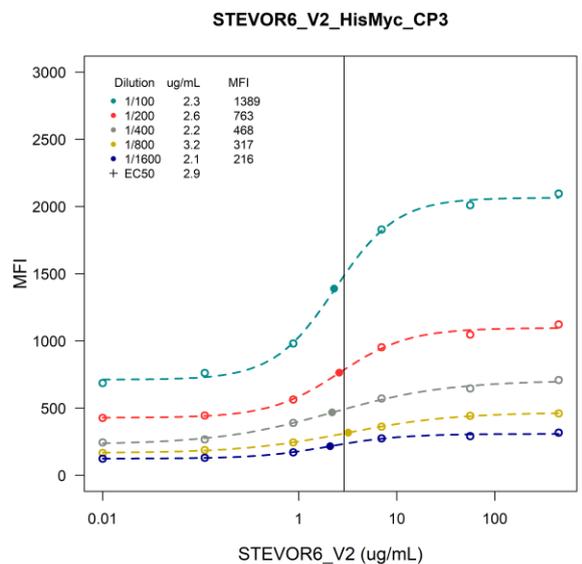
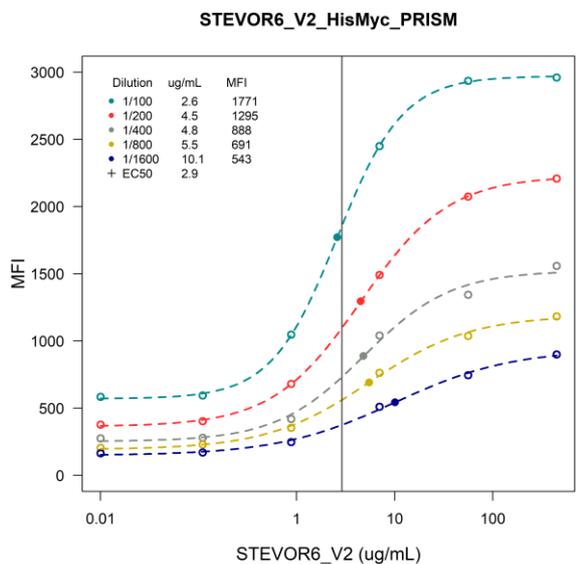
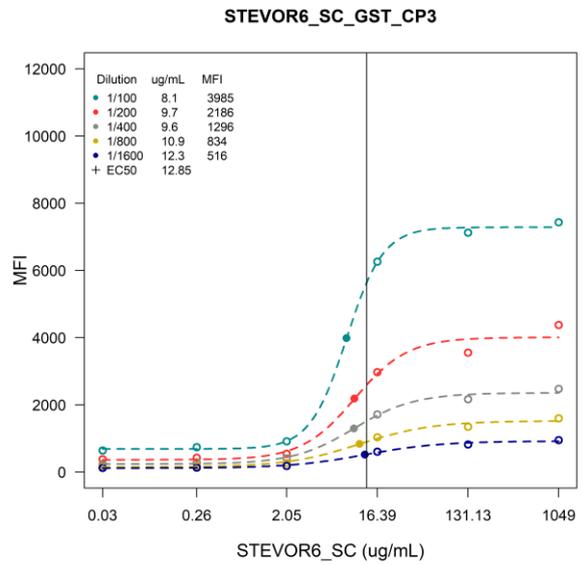
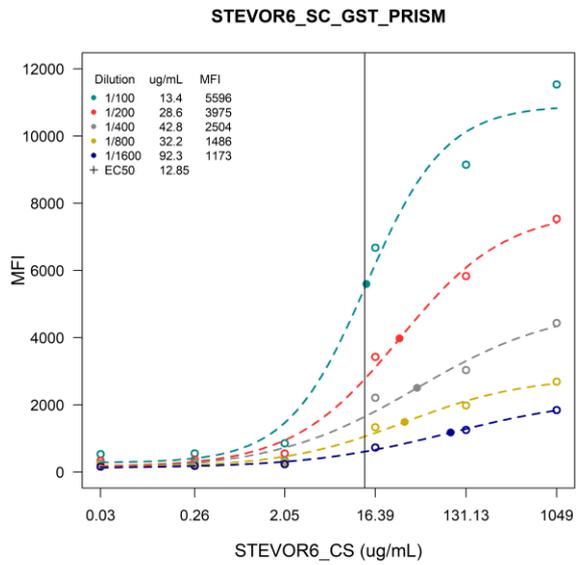
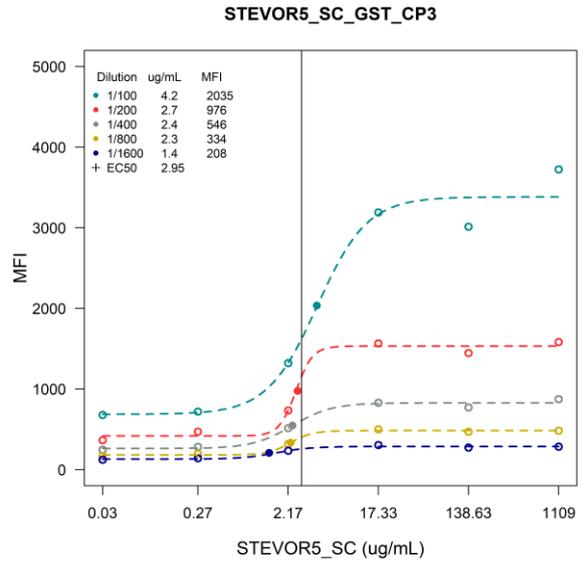
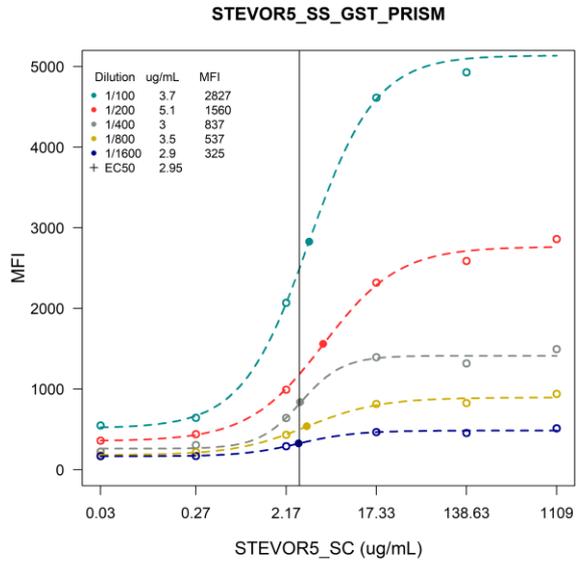
- exposure for individuals and communities. 2015;
32. Bhatwa A, Wang W, Hassan YI, Abraham N, Li XZ, Zhou T. Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in *Escherichia coli* and Strategies to Address Them for Industrial Applications. *Front Bioeng Biotechnol* [Internet]. 2021 Feb 10 [cited 2023 Jul 18];9. Available from: [/pmc/articles/PMC7902521/](#)
 33. Harper S, Speicher DW. Purification of proteins fused to glutathione S-transferase. *Methods Mol Biol* [Internet]. 2011 [cited 2023 Mar 14];681:259. Available from: [/pmc/articles/PMC3584333/](#)
 34. Kanya MR, Arinaitwe E, Wanzira H, Katureebe A, Barusya C, Kigozi SP, et al. Malaria Transmission, Infection, and Disease at Three Sites with Varied Transmission Intensity in Uganda: Implications for Malaria Control. *Am J Trop Med Hyg*. 2015;92(5):903–12.
 35. Singh S, Soe S, Weisman S, Barnwell JW, Pérignon JL, Druilhe P. A Conserved Multi-Gene Family Induces Cross-Reactive Antibodies Effective in Defense against *Plasmodium falciparum*. *PLoS One* [Internet]. 2009 Apr 30 [cited 2023 Apr 11];4(4). Available from: [/pmc/articles/PMC2671155/](#)
 36. Mehlin C, Boni E, Buckner FS, Engel L, Feist T, Gelb MH, et al. Heterologous expression of proteins from *Plasmodium falciparum*: Results from 1000 genes. *Mol Biochem Parasitol*. 2006 Aug 1;148(2):144–60.
 37. Vedadi M, Lew J, Artz J, Amani M, Zhao Y, Dong A, et al. Genome-scale protein expression and structural biology of *Plasmodium falciparum* and related Apicomplexan organisms. *Mol Biochem Parasitol*. 2007 Jan 1;151(1):100–10.

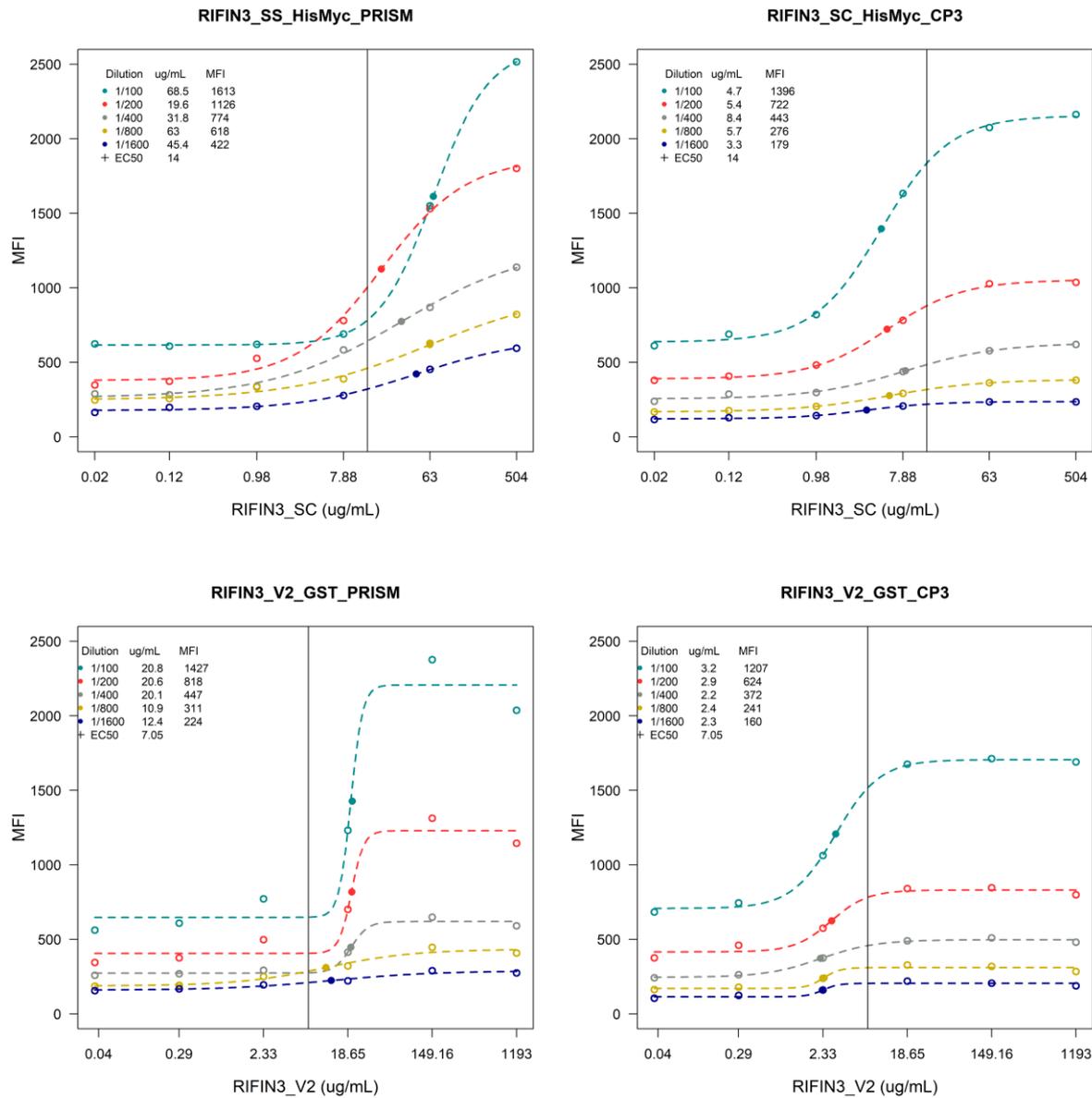
38. Tang C lian, Zhou H hua, Zhu Y wen, Huang J, Wang G bo. Glutathione S-transferase influences the fecundity of *Schistosoma japonicum*. *Acta Trop*. 2019 Mar 1;191:8–12.
39. Herman LS, Fornace K, Phelan J, Grigg MJ, Anstey NM, William T, et al. Identification and validation of a novel panel of *Plasmodium knowlesi* biomarkers of serological exposure. *PLoS Negl Trop Dis* [Internet]. 2018 Jun 1 [cited 2023 Jul 18];12(6):e0006457. Available from:
<https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0006457>
40. Tetteh KKA, Osier FHA, Salanti A, Kamuyu G, Drought L, Faily M, et al. Analysis of Antibodies to Newly Described *Plasmodium falciparum* Merozoite Antigens Supports MSPDBL2 as a Predicted Target of Naturally Acquired Immunity. *Infect Immun* [Internet]. 2013 [cited 2023 Jul 18];81(10):3835. Available from:
</pmc/articles/PMC3811751/>
41. Byrne B, Stack E, Gilmartin N, O’Kennedy R. Antibody-Based Sensors: Principles, Problems and Potential for Detection of Pathogens and Associated Toxins. *Sensors* [Internet]. 2009 Jun 5 [cited 2020 Jul 9];9(6):4407–45. Available from:
<http://www.mdpi.com/1424-8220/9/6/4407>
42. Wu L, Mwesigwa J, Affara M, Bah M, Correa S, Hall T, et al. Sero-epidemiological evaluation of malaria transmission in The Gambia before and after mass drug administration. *BMC Med* [Internet]. 2020 Dec 1 [cited 2023 Mar 13];18(1). Available from: </pmc/articles/PMC7664049/>
43. Doolan DL, Dobaño C, Baird JK. Acquired immunity to Malaria [Internet]. Vol. 22, *Clinical Microbiology Reviews*. American Society for Microbiology (ASM); 2009 [cited 2020 Oct 20]. p. 13–36. Available from:
</pmc/articles/PMC2620631/?report=abstract>

44. Faber J, Fonseca LM. How sample size influences research outcomes. *Dental Press J Orthod* [Internet]. 2014 Jul 1 [cited 2023 Apr 11];19(4):27. Available from: [/pmc/articles/PMC4296634/](#)
45. Greenwood B, Marsh K, Snow R. Why do some African children develop severe malaria? *Parasitol Today*. 1991 Jan 1;7(10):277–81.
46. Kanoi BN, Nagaoka H, White MT, Morita M, Palacpac NMQ, Ntege EH, et al. Global Repertoire of Human Antibodies Against *Plasmodium falciparum* RIFINs, SURFINs, and STEVORs in a Malaria Exposed Population. *Front Immunol* [Internet]. 2020 May 12 [cited 2023 Apr 11];11. Available from: <https://pubmed.ncbi.nlm.nih.gov/32477363/>
47. Saito F, Hirayasu K, Satoh T, Wang CW, Lusingu J, Arimori T, et al. Immune evasion of *Plasmodium falciparum* by RIFIN via inhibitory receptors. *Nature* [Internet]. 2017 Dec 12 [cited 2023 Jul 18];552(7683):101. Available from: [/pmc/articles/PMC5748893/](#)
48. Dechavanne C, Cottrell G, Garcia A, Migot-Nabias F. Placental Malaria: Decreased Transfer of Maternal Antibodies Directed to *Plasmodium falciparum* and Impact on the Incidence of Febrile Infections in Infants. *PLoS One* [Internet]. 2015 Dec 1 [cited 2023 Apr 11];10(12). Available from: [/pmc/articles/PMC4689360/](#)
49. Lavazec C, Sanyal S, Templeton TJ. Expression switching in the stevor and Pfmc-2TM superfamilies in *Plasmodium falciparum*. *Mol Microbiol*. 2007;64(6):1621–34.
50. Blythe JE, Niang M, Marsh K, Holder AA, Langhorne J, Preiser PR. Characterization of the repertoire diversity of the *Plasmodium falciparum* stevor multigene family in laboratory and field isolates. *Malar J* [Internet]. 2009 [cited 2023 Apr 11];8(1):140. Available from: [/pmc/articles/PMC2706845/](#)

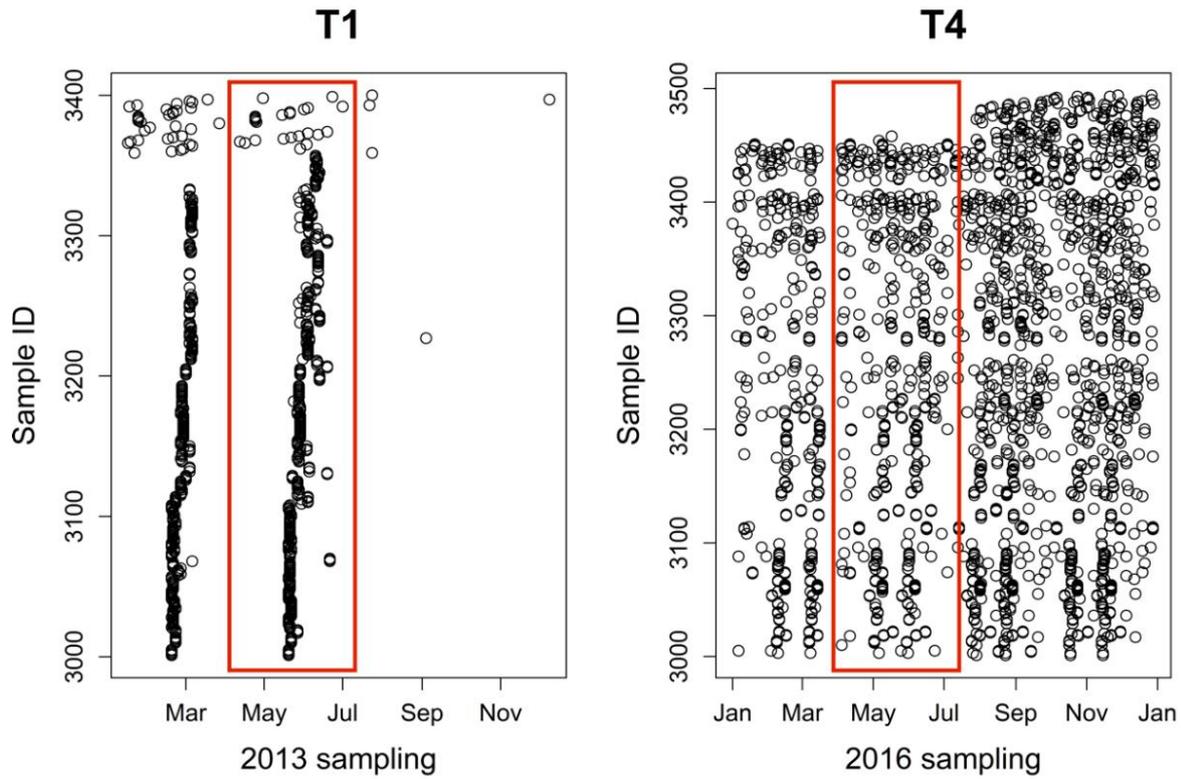
3.10 Supplementary materials



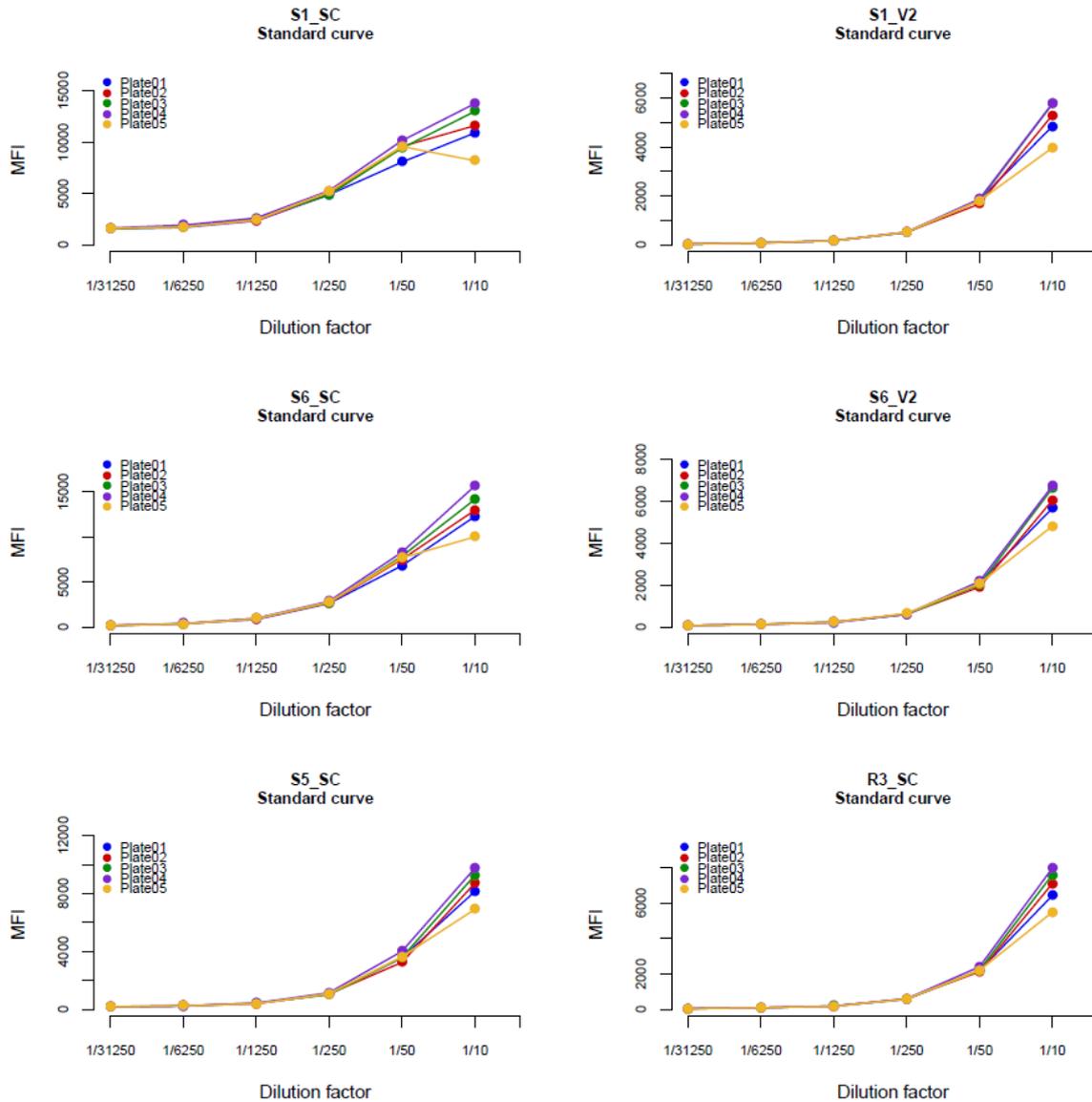




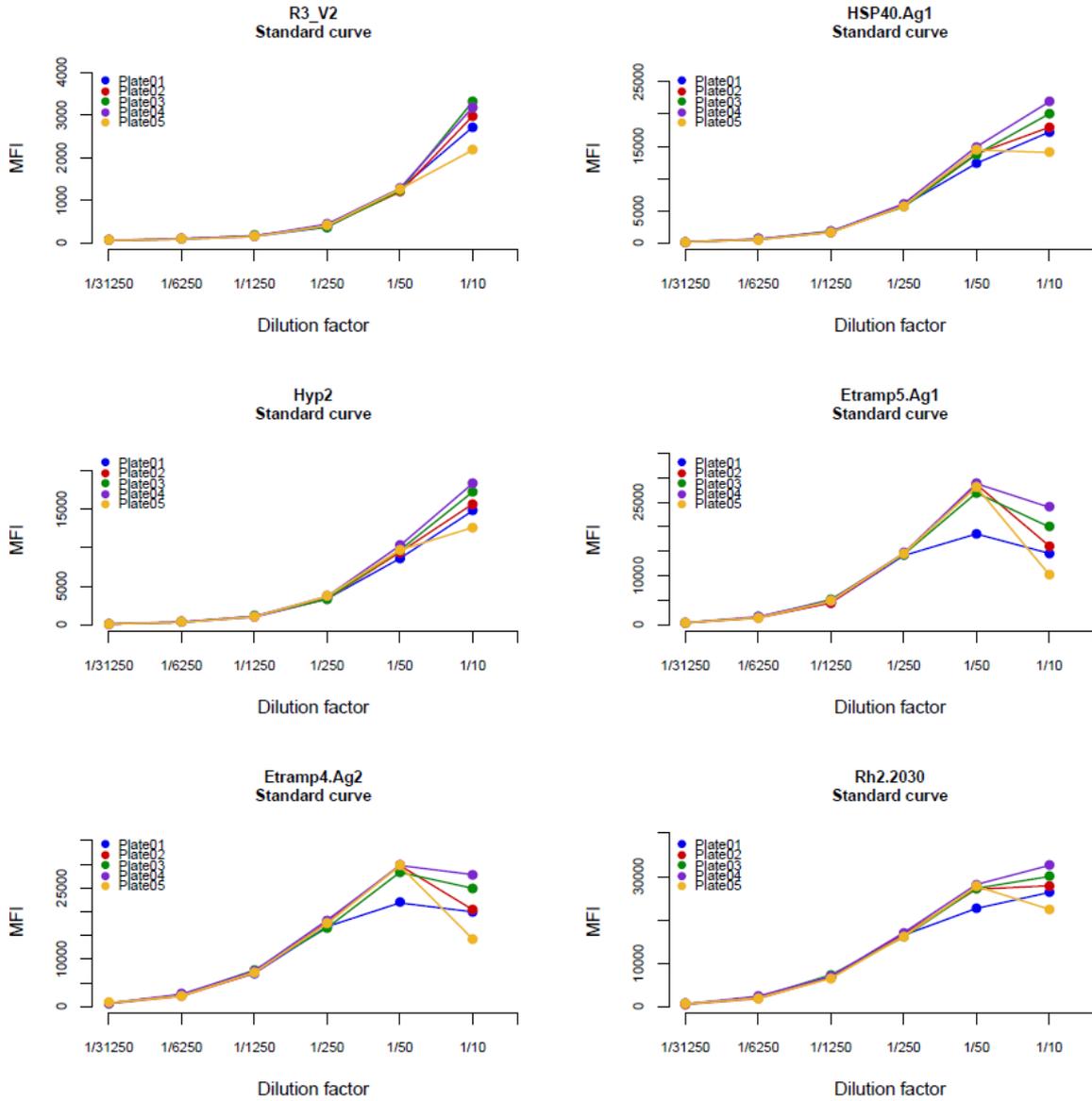
Supplementary Figure 1: Recombinant proteins titrations on MagPlex microspheres against two positive controls: pool serum from Tanzania (CP3) and pool serum from Uganda (PRISM1). Optimum coupling concentration (EC50 point of saturation) was calculated using the median EC50 value obtained from all curves subjected to 4-parameter logistic regression.



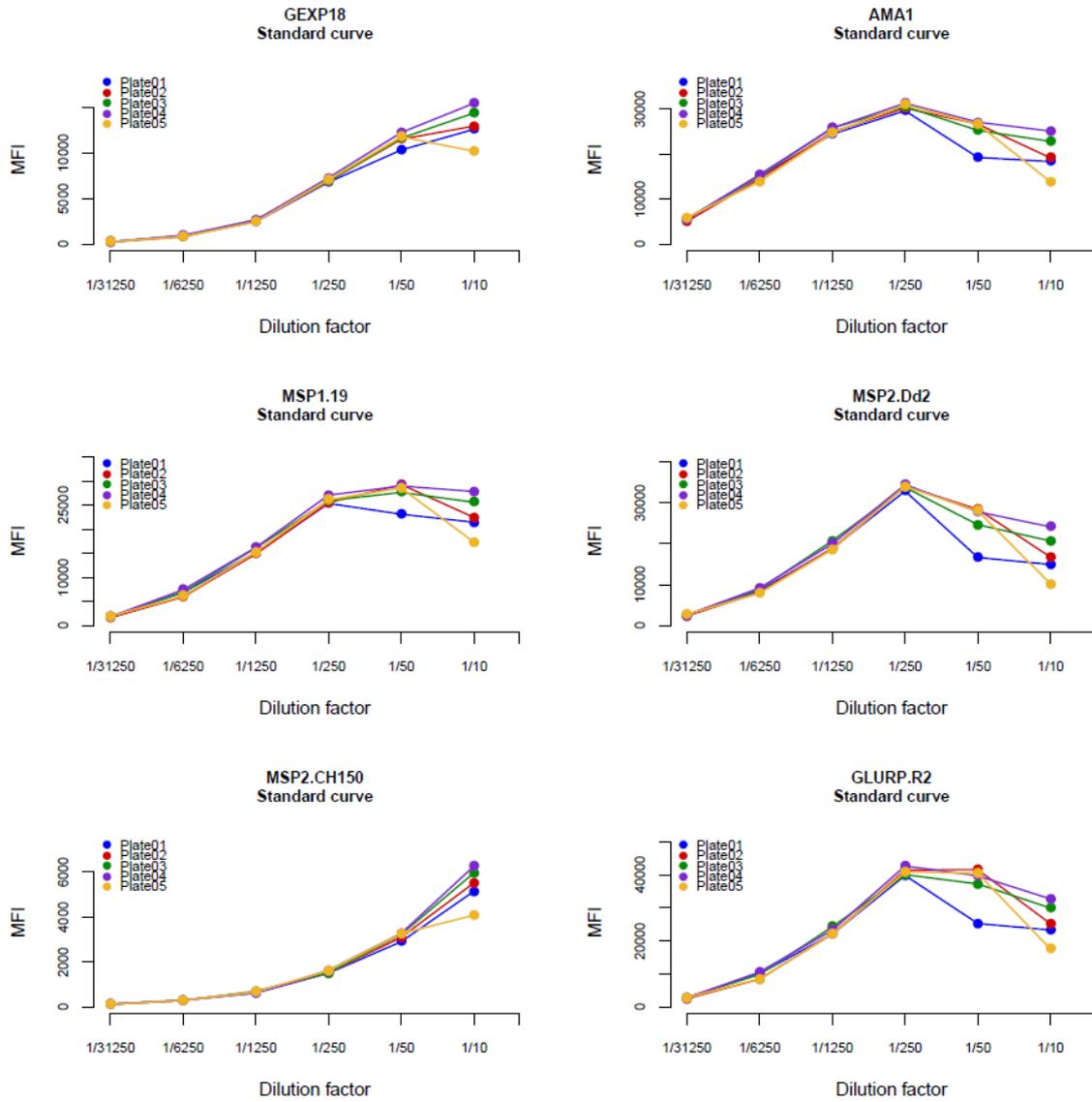
Supplementary Figure 2: Sampling distribution per time of the year for the two sample time points (2013 and 2016). Sample IDs are displayed on the y-axis. Out of all samples selected for the study indicated in the red rectangles, 74% are paired samples between the 2013 and 2016 timepoints.



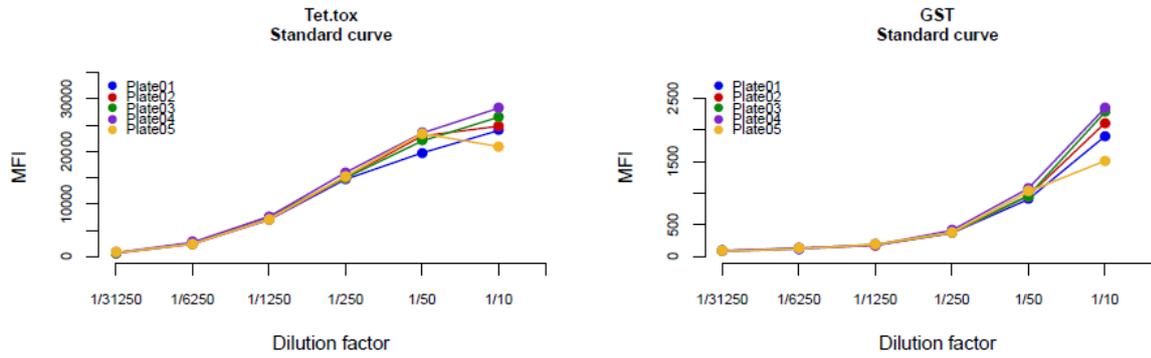
Supplementary Figure 3.1: Control standard curves of PRISM1 from 1/10 down to 1/31250 serum dilution for plates 1 to 5 of Luminex assay for the following antigens: S1_SC: STEVOR1_SC, S1_V2: STEVOR1_V2, S6_SC: STEVOR6_SC, S6_V2: STEVOR6_V2, S5_SC: STEVOR5_SC, R2_SC: RIFIN3_SC.



Supplementary Figure 3.2: Control standard curves of PRISM1 from 1/10 down to 1/31250 serum dilution for plates 1 to 5 of Luminex assay for the following antigens: R3_V2: RIFIN3_V2, HSP40.Ag1, Hyp2, Etramp5.Ag1, Etramp4.Ag2, Rh2.203.



Supplementary Figure 3.3: Control standard curves of PRISM1 from 1/10 down to 1/31250 serum dilution for plates 1 to 5 of Luminex assay for the following antigens: GEXP18, AMA1, MSP1.19, MSP2. Dd2, MSP2, CH150, GLURP.R2.



Supplementary Figure 3.4: Control standard curves of PRISM1 from 1/10 down to 1/31250 serum dilution for plates 1 to 5 of Luminex assay for the following control recombinants: Tet.tox and GST.

Supplementary Table 1: Summary of previously established *Plasmodium falciparum* markers of infection exposure complementing the studied STEVOR and RIFIN recombinant proteins.

Recombinant Antigen	Protein Name	Protein Species	Exposure marker	Expression system
AMA1	Apical Membrane Antigen 1	<i>Plasmodium falciparum</i>	Long-term marker	His-tag <i>E. coli</i>
MSP1.19	Merozoite Surface Protein 1	<i>Plasmodium falciparum</i>	Long-term marker	GST-tag <i>E. coli</i>
GLURP.R2	Glutamate-Rich Protein Region 2	<i>Plasmodium falciparum</i>	Long-term marker	GST-tag <i>E. coli</i>
MSP2.Dd2	Merozoite Surface Protein 2, Dd2 allele	<i>Plasmodium falciparum</i>	Short-term marker	GST-tag <i>E. coli</i>
Etramp5.Ag1	Early Transcribed Membrane Protein 5	<i>Plasmodium falciparum</i>	Short-term marker	GST-tag <i>E. coli</i>
Etramp4.Ag2	Early Transcribed Membrane Protein 4	<i>Plasmodium falciparum</i>	Short-term marker	GST-tag <i>E. coli</i>
Rh2.2030	Reticulocyte Binding Protein Homologue 2	<i>Plasmodium falciparum</i>	Medium-term marker	GST-tag <i>E. coli</i>
HSP40.Ag1	Heat-Shock Protein 40	<i>Plasmodium falciparum</i>	Short-term marker	GST-tag <i>E. coli</i>
Hyp2	Hypothetical Protein 2	<i>Plasmodium falciparum</i>	Short-term marker	GST-tag <i>E. coli</i>
GEXP18	Gametocyte Exported Protein 18	<i>Plasmodium falciparum</i>	Short-term marker	GST-tag <i>E. coli</i>
Tet.tox	Tetanus toxoid	<i>Clostridium tetani</i>	Immunization antigen, serology control	GST-tag <i>E. coli</i>
GST	Glutathione-S-Transferase	<i>Schistosoma japonicum</i>	GST-tagged proteins control	GTS-tag <i>E. coli</i>

Supplementary Table 2: Summary table of mean MFI values of PHE negative controls and calculated seropositivity thresholds of PHE mean MFI plus three times standard deviation, per antigen.

Antigen	Mean PHE (MFI)	3SD cut-off (MFI)	Highest positive MFI
STEVOR1_SC	303.20	1580.84	21942.75
STEVOR1_V2	146.43	463.57	2147.50
STEVOR6_SC	253.65	1182.37	20903.50
STEVOR6_V2	229.97	692.81	3427.75
STEVOR5_SC	221.48	651.31	7734.00
RIFIN3_SC	220.61	1044.30	4422.50
RIFIN3_V2	245.83	767.67	6965.50
HSP40.Ag1	962.02	2861.01	16741.00
Hyp2	894.38	2595.58	12091.00
Etramp5.Ag1	260.74	725.93	42320.50
Etramp4.Ag2	979.62	3044.99	37020.50
Rh2.2030	378.96	1609.28	31749.00
GEXP18	518.18	1712.87	15867.50
AMA1	220.38	662.60	35201.50
MSP1.19	792.07	2362.44	33881.50
MSP2.Dd2	1022.31	11100.98	40267.75
GLURP.R2	153.93	478.74	52300.00
Tet.tox	13019.20	26842.07	29119.50
GST	337.34	985.23	897.25

Supplementary Table 3: Summary of population seroprevalence per survey year for each recombinant antigen.

Age strata	2013 Seroprevalence n (%)			2016 Seroprevalence n (%)		
	6 months – 5 years	5 years – 11 years	>18 years	6 months – 5 years	5 years – 11 years	>18 years
Long-term markers						
PfAMA1	58 (79.45)	144 (94.74)	87 (94.57)	24 (53.33)	120 (86.33)	0 (0.00)
PFMSP1.19	25 (34.25)	65 (42.76)	66 (71.74)	10 (22.22)	48 (34.53)	6 (54.55)
MSP2.Dd2	15 (20.55)	64 (42.11)	51 (55.43)	2 (4.44)	31 (22.30)	7 (63.64)
GLURP.R2	47 (64.38)	124 (81.58)	87 (94.57)	17 (37.78)	89 (64.03)	0 (0.00)
Short-term markers						
HSP40.Ag1	9 (12.33)	7 (4.61)	16 (17.39)	0 (0.00)	2 (1.44)	1 (9.09)
Hyp2	2 (2.74)	4 (2.63)	8 (8.70)	0 (0.00)	4 (2.88)	1 (0.00)
Etramp5.Ag1	34 (46.58)	75 (49.34)	55 (59.78)	16 (35.56)	54 (38.85)	6 (54.55)
Etramp4.Ag2	7 (9.59)	38 (25.00)	36 (39.13)	2 (4.44)	24 (17.27)	3 (27.27)
Rh2.2030	22 (30.14)	78 (51.32)	51 (55.43)	7 (15.56)	77 (55.40)	6 (54.55)
GEXP18	9 (12.33)	24 (15.79)	27 (29.35)	2 (4.44)	8 (5.76)	1 (9.09)
Hypervariable recombinants						
STEVOR1_SC	25 (34.25)	51 (33.55)	20 (21.74)	4 (8.89)	34 (24.46)	3 (27.27)
STEVOR1_V2	1 (1.37)	2 (1.32)	6 (6.52)	0 (0.00)	5 (3.6)	0 (0.00)
STEVOR6_SC	6 (8.22)	20 (13.16)	19 (20.65)	1 (2.22)	14 (10.07)	3 (27.27)
STEVOR6_V2	1 (1.37)	2 (1.32)	9 (9.78)	2 (4.44)	6 (4.32)	1 (9.09)
STEVOR5_SC	4 (5.48)	7 (4.61)	10 (10.87)	1 (2.22)	3 (2.16)	0 (0.00)
RIFIN3_SC	1 (1.37)	2 (1.32)	3 (3.26)	1 (2.22)	4 (2.88)	1 (9.09)
RIFIN3_V2	1 (1.37)	4 (2.63)	4 (4.35)	0 (0.00)	3 (2.16)	0 (0.00)
Internal Controls						
Tet.tox	1 (1.37)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
GST	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	lsh1601985	Title	Miss
First Name(s)	HRISTINA CHAVDAROVA		
Surname/Family Name	VASILEVA		
Thesis Title	BIOCHEMICAL AND BIOINFORMATIC CHARACTERISATION OF UNDERSTUDIED ERYTHROCYTE SURFACE EXPRESSED HYPERVARIABLE PROTEIN FAMILIES IN PLASMODIUM FALCIPARUM.		
Primary Supervisor	Dr. Kevin Tetteh		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	New Frontiers in Malaria
---	--------------------------

Please list the paper's authors in the intended authorship order:	Hristina Vasileva, Ana Chopo-Pizarro, Michael Ooko, Elin Dumont, Lindsey Wu, Isaac Ssewanyana, Anna Last, Kevin K.A. Tetteh
Stage of publication	Not yet submitted

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I have done 90% of the laboratory work, I have done all of the statistical analysis and the manuscript preparation.
--	---

SECTION E

Student Signature	HRISTINA VASILEVA
Date	30.05.2024

Supervisor Signature	ANNA LAST
Date	06.06.2024

Chapter 4: Methods for mapping the variability of STEVOR protein sequences

Linking material to Chapter 5

4.1 Introduction

Following the discussion and conclusions made in Chapter 3, a hypothesis that members of the STEVOR protein family have a potential to be explored as markers of infection exposure to *P. falciparum* was made, although further analysis using more STEVOR variants is needed (1). Given that each parasite genome encodes 30 to 40 unique STEVOR variants, with a single variant being expressed during a polyclonal infection, it is necessary to determine which variants to express and investigate further (2). To do so, it was essential to examine the variability between each of the family members and understand their relationships, which would facilitate the selection of representative variants for the creation of a STEVOR recombinant antigen library.

This chapter describes the analysis of protein sequence data to explore the sequence variability within the STEVOR protein family. The information obtained was used to select the most appropriate methodology to facilitate informed decision-making for the selection of variants for the construction of STEVOR recombinant protein library. Numerous methodologies already exist for exploring the genetic relationships among sequences derived from the same organism, such as sequence motif and pattern analysis, identifying sequence features and functions, and an automated method for identifying patterns such as machine learning and pattern recognition (3). Other methods such as phylogenetic trees and the application of dissimilarity matrices for data clustering using *k*-means (unsupervised algorithm for grouping based on similarity and

choice of number of clusters: k), each with its own set of advantages and limitations are discussed and utilised in this study (4),(5).

Phylogenetic analysis is commonly used to explore the relationship among distinct strains of the same organism in order to track mutations and recombination events (4). It is also a powerful tool in tracking the spread of infections geographically, identifying epidemic hot-spots and movement trends (6). Recent example of this approach has been demonstrated in multi-disease tracking sites such as GISAID (<https://gisaid.org/>) and Nextstrain (<https://nextstrain.org/>) (7),(8). However, these require huge amounts of resources, are driven by large amounts of data and call for the performance of a large number of permutations in the form of bootstrapping in order to select the most probable positioning of the sequences in the tree (4). Moreover, there is a need for the selection of a reference sequence to be used to “root” the tree for graphical representation which could introduce bias to the analysis. With highly variable sequence families such as STEVOR, some of these issues can be mitigated by using a dissimilarity matrix for two-dimensional data representation and grouping via k -means clustering. This method relies on a distance matrix derived from a phylogenetic analysis, offering an alternative approach. However, this approach also presents challenges, particularly in terms of stability and reproducibility (9).

The k -means grouping is a simple, easy to implement and highly efficient when applied on large data sets, an approach aiming to minimise variation within a group (k) while keeping the distance between groups as large as possible (5). However, this approach is highly dependent on the choice of centroid, the central point of each cluster determining the maximum distance allowed within a cluster, as well as the number of clusters selected to force the data in. Selecting correctly both of those parameters is crucial for k -means analysis, where the selection of optimum number of k -means is inheritably arbitrary (5). Moreover, the k -means analysis is sensitive to outliers and assumes a linear data structure, thus does not perform well

on data sets with complex, non-linear structure, such is the relationship between STEVOR V2 variants (5),(9).

This chapter introduces a new computational model designed to explore the similarity and dissimilarity of the STEVOR proteins sequence data. While this method has demonstrated a high-level of reproducibility, it's advantages and limitations are further compared and contrasted with the two conventional techniques introduced above. The three methods are performed to map the variation within STEVOR protein sequences from the 3D7 *P. falciparum* reference strain as proof of concept.

4.2 Methods

4.2.1 Protein sequences alignment and hypervariable domain isolation

A total of 43 STEVOR amino acid sequences, representing all 3D7 *P. falciparum* reference strain STEVOR variants available, were downloaded from PlasmoDB database (<https://plasmodb.org/plasmo/app>) (10). The amino acid sequence “*Pf*3D7_0632500” was excluded from the further analysis due to being annotated as an “*P. falciparum* erythrocyte membrane protein 1 (*PfEMP1*)”, leaving a total number of 42 STEVOR protein sequences, in a *fasta* file format. A multiple alignment of the 42 sequences (Figure 1) was performed using the MAFFT software on a Windows 10 terminal using a G-INS-I strategy with global pair with Smith-Waterman algorithm pairwise alignment and a gap extension penalty for group-to-group variation with a default value of 0.123 (11),(12). This algorithm is designed for sequences with global homology, aligning the entire sequence from two sequences against each other in a pairwise matter and preventing introductions of gaps (13),(14).

No misalignments were observed, and no manual realignment was performed. The large variable domain (V2) was isolated from the alignment by taking the sequences between the two transmembrane domains (2xTM), identified using the trans-membrane prediction software

(TMHMM2.0), found to be between amino acid position 158 and amino acid 215 (15). The protein sequence PlasmoDB names and the adapted V2 sequence names can be found in Table 1.

4.2.2 Mapping STEVOR variation using phylogenetic trees.

A substitution model best fitting the sequences data was tested in IQtree (Version 1.6.12 Windows 10), and the JTT+F+G4 model was selected as best likelihood model according to the Bayesian Information Criterion (BIC) for balancing the goodness of fit in the model with the number of parameters used in the model (16),(17). This model was selected after testing the sequence data through all possible substitution models, where the selected model had the highest likelihood value of compatibility with the provided data. A bootstrap analysis containing 1000 replicates was performed using the selected substitution model, optimised using a hill-climbing nearest neighbour interchange (NNI), for reducing the risk of overestimating branch support, resulting in genetic tree presented in Newick format (18). The tree was visualised using FigTree (v1.4.4) a molecular evolution, phylogenetics and epidemiology software, (Figure 2) (19).

4.2.3 Mapping STEVOR variation using *k*-means and dissimilarity matrix.

The genetic tree was further transformed from Newick format into a “. tre” file and loaded in R 4.3 computational platform to obtain the distance matrix using “*ape*” package. Further the correlation (distance) matrix was transformed into a dissimilarity matrix calculated using “*magrittr*” package and multidimensional scaling (MDS) was applied with 2 dimensions using *k*-means from one to maximum logical cluster number, in this case 11, illustrated in Figure 3. The choice of centroid for each cluster was random.

4.2.4 Mapping STEVOR variation using sequence digitalisation and principal component analysis.

Since protein sequence data is qualitative, the sequences were converted into numerical values in order to calculate the estimated quantitative relationship between them. Therefore, the 42 STEVOR V2 domain aligned amino acid sequences were digitised into binary Boolean vectors, briefly as follows. Each amino acid position of each of the protein sequences was challenged for the presence or absence of each amino acid present in the *P. falciparum* proteome (n=21). Results were recorded in 21 long vectors of data (Boolean vector) with presence recorded as 1 and absence as 0. Furthermore, a score of each 21 vectors was calculated. The 21 Boolean vectors were further piled into a multidimensional matrix and subjected to Singular Value Decomposition (SVD), resulting in distance matrix of Euclidean distances (sPCs) reflecting the distance relationship between each of the protein sequences to one other (20),(21). The sPC matrix was further subjected to a principal component analysis (PCA) and the first two principal components were displayed into two-dimensional plots in Figure 4 (9). All analysis and visualisation were performed using R 4.3 computational platform.

4.3 Results

There was a substantial variability present within the aligned V2 region of all 42 3D7 STEVOR sequences, where Figure 1 provides a visual depiction of the amino acid variability of each sequence.

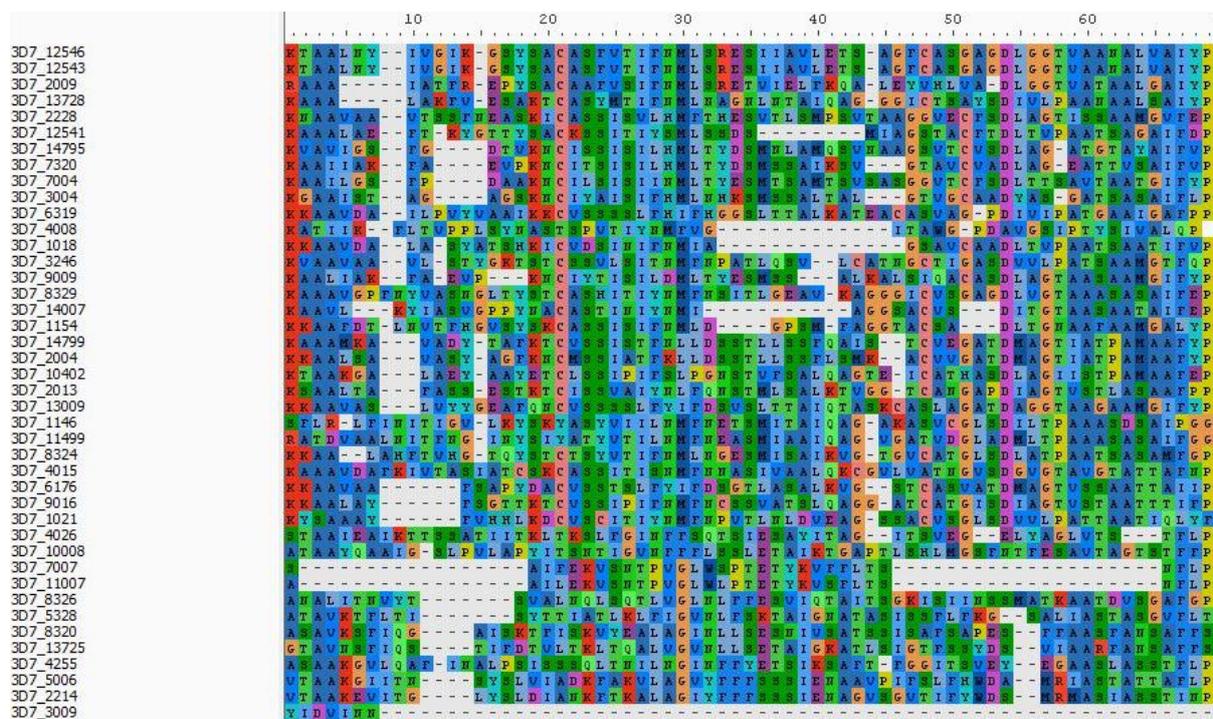


Figure 1: MAFFT multiple alignment of 42-variable domain 3D7 STEVOR sequences performed using G-INS-I with Smith-Waterman pairwise algorithm and gap extension penalty for group-to-group variation value of 0.123 according to BIC, displayed using AliView open-source alignment viewer and editor software. The alignment is color-coded according to amino acids.

Some of the sequences retain the same k -mean group, particularly in k -mean one and k -mean three when the multidimensional scaling was applied with three and with 11 clusters, displayed in blue in Table 1. Increasing the number of k -means splits the original groups into more smaller groups, as the analysis is applied on the same distance matrix and only the number of k -means is changed. Adjusted IDs were generated from the phylogenetic tree's distance matrix.

Table 1: *P. falciparum* 3D7 reference strain STEVOR protein data base IDs, adjusted IDs, and *k*-mean positions when multidimensional scaling is applied with three and 11 *k*-mean clusters.

PlasmoDB ID	Alignment ID	Adjusted ID	<i>k</i>-mean (n=3)	<i>k</i>-mean (n=11)
Pf3D7_1372800	3D7_13728	1372800_19	1	1
Pf3D7_1254100	3D7_12541	1254100_19	1	1
Pf3D7_0700400	3D7_7004	0700400_19	1	1
Pf3D7_0101800	3D7_1018	0101800_19	1	1
Pf3D7_1149900	3D7_11499	1149900_20	1	1
Pf3D7_0102100	3D7_1021	0102100_18	1	1
Pf3D7_0222800	3D7_2228	0222800_19	1	5
Pf3D7_1479500	3D7_14795	1479500_19	1	5
Pf3D7_0732000	3D7_732	0732000_19	1	5
Pf3D7_0900900	3D7_9009	0900900_20	1	5
Pf3D7_0300400	3D7_3004	0300400_19	1	6
Pf3D7_0324600	3D7_3246	0324600_19	1	7
Pf3D7_0114600	3D7_1146	0114600_20	1	7
Pf3D7_0300900	3D7_3009	0300900_18	1	7
Pf3D7_0700700	3D7_7007	0700700_21	2	8
Pf3D7_1100700	3D7_11007	1100700_21	2	8
Pf3D7_0832600	3D7_8326	0832600_21	2	8
Pf3D7_0532800	3D7_5326	0532800_15	2	8
Pf3D7_0832000	3D7_832	0832000_20	2	8
Pf3D7_1372500	3D7_13725	1372500_19	2	8
Pf3D7_0500600	3D7_5006	0500600_20	2	8
Pf3D7_0221400	3D7_2214	0221400_20	2	8
Pf3D7_0402600	3D7_4026	0402600_20	2	11
Pf3D7_1000800	3D7_10008	1000800_18	2	11
Pf3D7_0425500	3D7_4255	0425500_20	2	11
Pf3D7_0400800	3D7_4008	0400800_19	3	2
Pf3D7_1254600	3D7_12546	1254600_19	3	3
Pf3D7_1254300	3D7_12543	1254300_19	3	3
Pf3D7_0200900	3D7_2009	0200900_19	3	3
Pf3D7_0832900	3D7_8329	0832900_19	3	3
Pf3D7_0700700	3D7_7007	0700700_21	3	3
Pf3D7_0115400	3D7_1154	0115400_20	3	3
Pf3D7_1479900	3D7_14799	1479900_20	3	4
Pf3D7_0200400	3D7_2004	0200400_20	3	4
Pf3D7_1040200	3D7_10402	1040200_20	3	4
Pf3D7_0201300	3D7_2013	0201300_20	3	4
Pf3D7_0832400	3D7_8324	0832400_20	3	4
Pf3D7_0617600	3D7_6176	0617600_20	3	4
Pf3D7_0901600	3D7_9016	0901600_20	3	4
Pf3D7_0631900	3D7_6319	0631900_19	3	9
Pf3D7_1300900	3D7_13009	1300900_20	3	9
Pf3D7_0401500	3D7_4015	0401500_20	3	10

From the k -mean analysis the granularity of explaining variation between sequences increases with the number of k -means selected (Figure 2).

When colour coding the phylogenetic tree according to the k -mean grouping it can be seen a logical separation of the three k -means when represented as a tree. When the multidimensional scaling was applied using 11 k -means, the phylogenetic tree has some non-logical colour clustering, specifically for group one and seven represented in magenta and lime colour respectively, in Figure 3.

Colour coding the STEVOR variants in the PCA analysis according to their k -mean grouping clearly shows that the two methods of explaining variation and attempting clustering vary. In Figure 4A can be seen some clustering of k -mean group one according to PC1, but no clustering according to PC2. Similarly, but less pronounced pattern is observed for the sequences from k -mean two and k -mean three. However, applying higher number of k -means improves the variant separation pattern, clearly observed in k -mean group eight in purple in Figure 4B, and k -mean two containing only a single variant in the middle of the PCR plot, clearly distinct to the rest of the variants according to both principal components.

Since MDS does not provide the variance explained by each dimension explicitly, Figure 2 does not display percentage variation on its axis, in contrast to Figure 4, where the axis indicate the percentage of variance explained by each principal component.

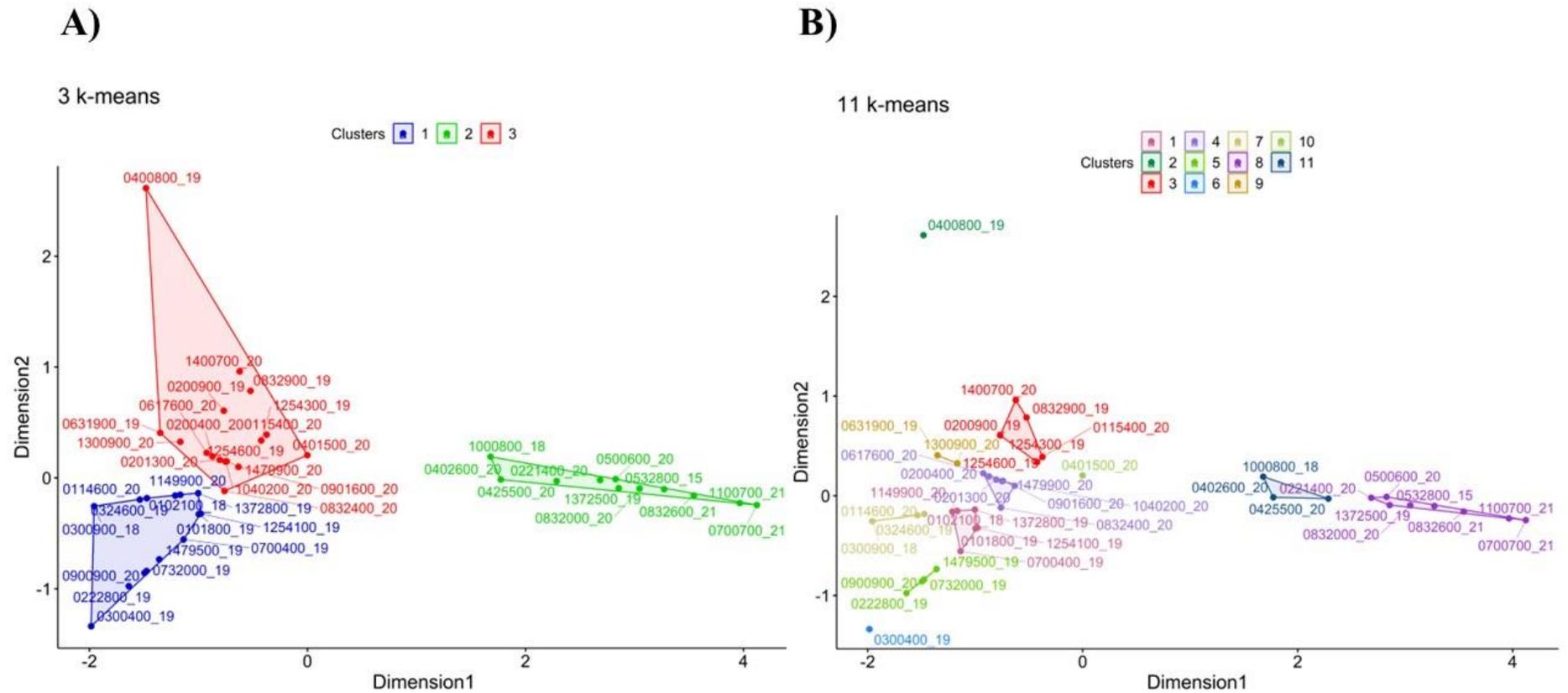
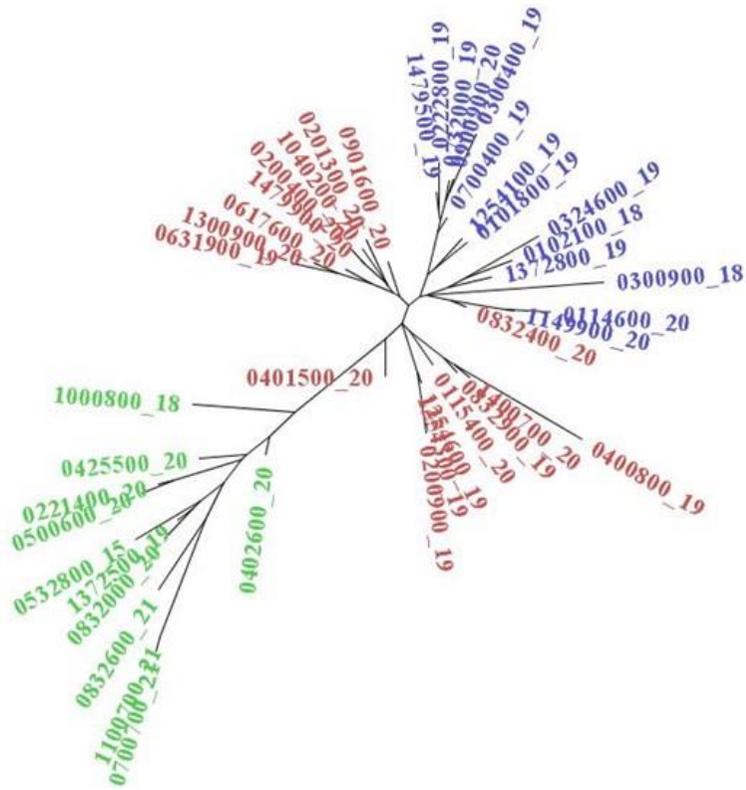


Figure 2: Two-dimensional scaling of dissimilarity matrix from the phylogenetic tree of PlasmoDB STEVOR V2 protein sequences, applied using A) three *k*-means and B) 11 *k*-means.

A)



B)

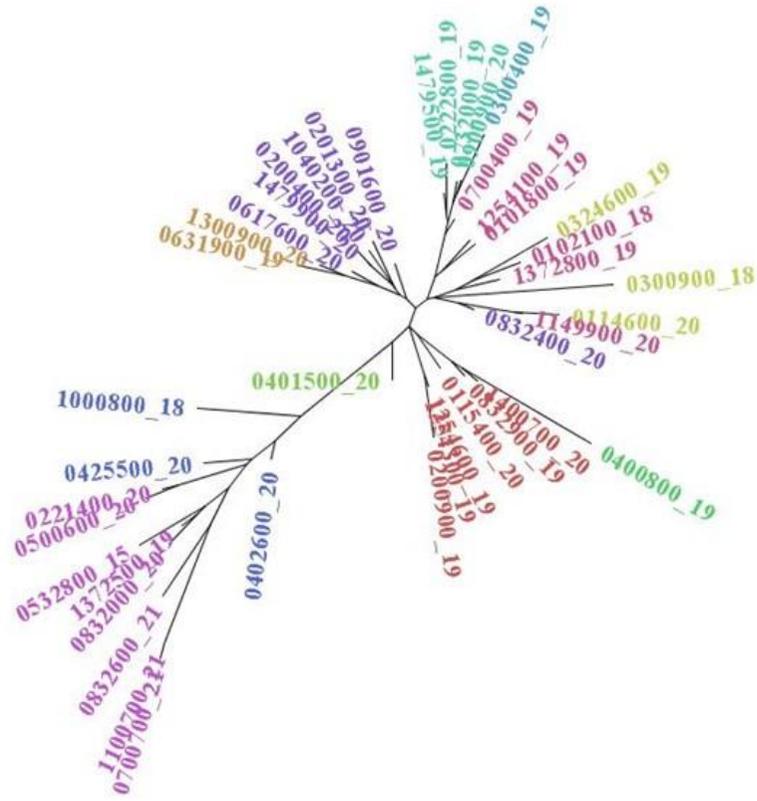


Figure 3: Neighbour joining phylogenetic tree of PlasmoDB database STEVOR V2 protein sequences, colour coded according to the k -mean position from Table 1 according to distance matrix multidimensional scaling applied using A) three k -means and B) 11 k -means.

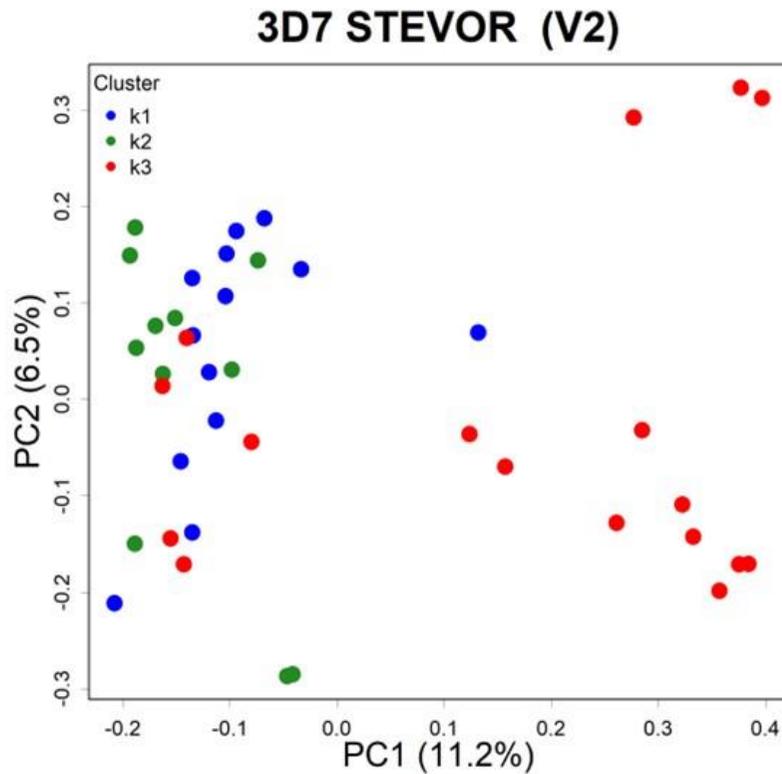
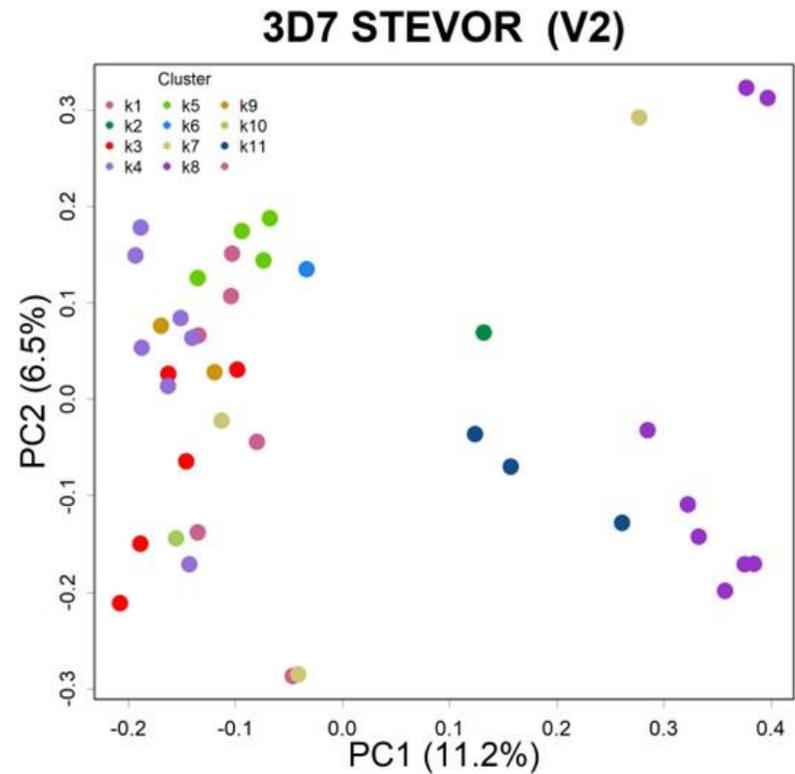
A)**B)**

Figure 4: First and second principal components from the PCA analysis applied on the PlasmDB 3D7 STEVOR V2 sequences after digitalisation and matrix transformation using singular value decomposition. Variants are colour coded according to their k -mean position from Table 1 when multidimensional scaling of dissimilarity matrix was applied using A) three k -means and B) 11 k -means. Percentages indicate the proportion diversity explained by each of the two principal components.

The three methods described and displayed in this chapter possess different attributes as outlined in Table 2. The discussion below examines the advantages and disadvantages associated with these parameters, aiming to draw a conclusion regarding the most suitable method for the subsequent analysis of STEVOR V2 variants in Chapter 5.

Table 2: Parameters summary of sequence clustering methods.

Criteria	Phylogenetic trees	<i>k</i>-means	PCA
Alignment needed	Yes	Yes	Yes
Rooting needed	Yes	No	No
Permutations	Desirable	Desirable	Does not affect results
Selecting groups	No	Yes	No
Shape assumption	Yes	Yes	No
Reproducibility	Low	Low	Absolute

4.4 Discussion

This chapter presents three methodologies: phylogenetic trees, *k*-means analysis, and Boolean vectors PCA analysis, used for exploring the variation of STEVOR protein sequence data originating from 3D7 *Plasmodium falciparum* reference strain. The aim is to identify the most suitable method for mapping the global STEVOR variability, to subsequently select appropriate STEVOR variants for the development of STEVOR recombinant antigen library. The large variable domain of the aligned STEVOR 3D7 protein sequences was specifically isolated to avoid loss of resolution when analysing STEVOR variability, which would likely occur if the more conserved domains of the protein were included. This chapter is focused on the STEVOR protein sequences from the 3D7 *Plasmodium falciparum* reference strain only as a proof-of-concept step for selecting the mapping method, applied to a larger database in Chapter 5. Working with sequences from one strain allows for a manageable data size for multiple analysis and results comparison. The subsequent application of the chosen method, being the sequence digitising and principal component analysis, reported in Chapter 5, extends to the entire available STEVOR protein sequence database from PlasmoDB (10).

Multiple global genetic diversity analysis undertaken on nucleotide and protein sequence data of the *var* genes and their products are predominantly performed using Neighbour-Joining phylogenetic trees with 1000 bootstraps (22), (23). However, for this research, we decided not to use this approach due to the limitations mentioned below.

The decision to use the PCA method for exploring the variation of the STEVOR protein sequences and further selecting representative variants for the library was based on multiple parameters. Phylogenetic trees operate under the assumption that all samples exhibit a tree-shaped relationship, disregarding the possibility of horizontal gene transfer which would create a horizontal relationship between variants. The hypervariability of STEVORs within a strain is

also expected between strains which creates a complex network rather than a strict tree structure. The PCA analysis is a mathematical method that observes the variability between the sequences without harming the data structure as this method rotates the data matrix to find direction of differences and projects them to set of diagonal axes. Additionally, this study seeks to investigate the variability of STEVORs within the 3D7 strain and to identify potential clustering patterns that have not been previously reported. For the *var* gene protein products and protein variants from the RIFIN family, two protein families similar to STEVOR, some clustering has been observed and documented, not only at the sequence level but also based on functional properties and their association with diseases severity (24), (25). In contrast, such grouping has not been yet demonstrated for STEVORs to guide decision for clustering using phylogenetic trees. Furthermore, even with the application of permutations to determine the most probable organisation of the sequences in a tree format, the reproducibility remains low when performed on the same alignment (26). Similarly for the *k*-means method, arbitrary selection of number of *k*-means for clustering allows for the absorption of neighbouring samples around randomly specified starting samples, introducing degree of randomness and unpredictability when changing the number of *k*-means. Selecting different number of *k*-means can lead to drastic changes in the analysis outcomes, suggesting that this method is rather unstable. Moreover, *k*-means analysis assumes grouping of equal sizes, is sensitive to outliers, and does not perform well on non-linear complex data structure (5). This study states that the PCA method demonstrates 100% reproducibility attributed to the conversion of the non-quantitative protein sequences into quantitative binary Boolean vectors which are further subjected to Singular Value Decomposition. This high reproducibility is due to the resulted 2D coordinates for each of the sequences, displayed as the first two principal components in the PCA plot, ensuring that replicating the analysis on identical data yields identical PCA plot

coordinates. This level of reproducibility sets the selected method from the alternative methodologies discussed.

The combined variance explained by the first two principal components (PC1 and PC2) is only 17.7%, which may raise concerns about the adequacy of this PCA in capturing data patterns. This relatively low percentage indicates that a significant portion of the data's variance is dispersed across the remaining principal components, each contributing a smaller percentage than PC1 and PC2. Such distribution of variance is common in high-dimensional data, like protein sequences, where patterns are complex and not easily captured by a simple two-dimensional reduction. Although this limitation restricts the interpretability of the PCA plot, which only visualises the first two dimensions, the method does offer a quantifiable metric of explained variance – a benefit over the two other methods compared in the study, which lack this measure. This scattered variance across multiple components suggests that the data is inherently complex, with no dominant patterns that can be easily reduced. While scaling down to two dimensions may obscure some structural details, this limitation is shared with *k*-means. Additionally, the phylogenetic trees method looks at the differences between sequences in terms of type of amino acids and their positions and does not consider the whole sequences proportion variability between each other. The application of *k*-means also results in the loss of information at specific sites compromising essential data necessary for accurately defining the relationships between the variants. The PCA method analyses the information of the amino acid differences and the samples difference at a same time.

The interpretation of the PCA analysis could be enhanced by incorporating information about amino acids' biochemical properties, such as polarity, charge, size, hydrophobicity etc. This would allow for the analysis to assess whether amino acid variability is functionally significant, where the different biochemical properties could influence the protein structure and function.

Such variability could potentially bring some sequences closer together and separate others in PCA clusters.

To further refine clustering within the PCA, we could implement additional methods to introduce more structured groupings in the data. For instance, clustering algorithms like hierarchical clustering or t-distributed Stochastic Neighbour Embedding (t-SNE) could provide insights into potential clusters not readily apparent in the PCA alone. These methods would offer advantages in capturing subtle, non-linear relationships within high-dimensional data, which could reveal finer distinctions between amino acid variants (27).

Ultimately, a robust approach would involve the sequencing a subset of positive samples from the population to precisely identify the present variants. This strategy would further allow for the creation of recombinant variants tailored to the population's antibody reactivity profile. However, because the aim of this project is to examine antibody responses across populations with varying levels of malaria endemicity, the time and financial constraints make this approach impractical within the current project.

4.5 Conclusion

Based on the analysis conducted on the variable portion of the 3D7 STEVOR variants using the three methodologies shown in Figures 2, 3, and 4—phylogenetic trees, k-means clustering, and PCA analysis—we conclude that the PCA method offers a high level of objectivity and satisfactory reproducibility. Consequently, this method is applied to a larger dataset of STEVOR variants, as presented in Research Paper 2 (Chapter 5) of the thesis.

4.6 References

1. Vasileva H, Chopo-Pizarro A, Ooko M, Dumont E, Wu L, Ssewanyana I, et al. Novel hypervariable erythrocyte surface expressed recombinant proteins show promise as serological markers of exposure to *Plasmodium falciparum* infection. bioRxiv [Internet]. 2023 Nov 2 [cited 2023 Nov 7];2023.10.31.564916. Available from: <https://www.biorxiv.org/content/10.1101/2023.10.31.564916v1>
2. Scherf A, Lopez-Rubio JJ, Riviere L. Antigenic Variation in *Plasmodium falciparum*. Annu Rev Microbiol [Internet]. 2008 Oct 11 [cited 2019 May 20];62(1):445–70. Available from: <http://www.annualreviews.org/doi/10.1146/annurev.micro.61.080706.093134>
3. Hashim FA, Mabrouk MS, Al-Atabany W. Review of Different Sequence Motif Finding Algorithms. Avicenna J Med Biotechnol [Internet]. 2019 [cited 2024 Mar 18];11(2):130. Available from: </pmc/articles/PMC6490410/>
4. Nadler SA. Advantages and Disadvantages of Molecular Phylogenetics: A Case Study of Ascaridoid Nematodes. J Nematol [Internet]. 1995 [cited 2023 Oct 10];27(4):423. Available from: </pmc/articles/PMC2619635/?report=abstract>
5. k-Means Advantages and Disadvantages | Machine Learning | Google for Developers [Internet]. [cited 2023 Oct 10]. Available from: <https://developers.google.com/machine-learning/clustering/algorithm/advantages-disadvantages>
6. Ypma RJF, van Ballegooijen WM, Wallinga J. Relating phylogenetic trees to transmission trees of infectious disease outbreaks. Genetics [Internet]. 2013 [cited 2023 Oct 10];195(3):1055–62. Available from: </pmc/articles/PMC3813836/>

7. GISAID - gisaid.org [Internet]. [cited 2024 Mar 18]. Available from: <https://gisaid.org/>
8. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. NextStrain: Real-time tracking of pathogen evolution. *Bioinformatics*. 2018 Dec 1;34(23):4121–3.
9. Konishi T, Matsukuma S, Fuji H, Nakamura D, Satou N, Okano K. Principal Component Analysis applied directly to Sequence Matrix. *Sci Rep* [Internet]. 2019 Dec 1 [cited 2021 Apr 8];9(1):1–13. Available from: <https://doi.org/10.1038/s41598-019-55253-0>
10. PlasmoDB : The Plasmodium Genomics Resource [Internet]. [cited 2020 Mar 30]. Available from: <https://plasmodb.org/plasmo/>
11. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res* [Internet]. 2019 Jul 7 [cited 2023 Jun 20];47(W1):W5. Available from: </pmc/articles/PMC6602451/>
12. Katoh K, Kuma KI, Toh H, Miyata T. MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* [Internet]. 2005 [cited 2020 Oct 16];33(2):511–8. Available from: </pmc/articles/PMC548345/?report=abstract>
13. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* [Internet]. 2013 Apr [cited 2023 Oct 31];30(4):772. Available from: </pmc/articles/PMC3603318/>
14. Durbin R, Eddy SR, Krogh A, Mitchison G. Multiple sequence alignment methods. *Biol Seq Anal* [Internet]. 1998 Jul 6 [cited 2023 Oct 31];135–60. Available from: <https://www.cambridge.org/core/books/biological-sequence-analysis/multiple-sequence-alignment-methods/217DF949061D16E0D2A4B034233DA7F4>
15. Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane

- protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* [Internet]. 2001 Jan 19 [cited 2023 Jan 5];305(3):567–80. Available from: <https://pubmed.ncbi.nlm.nih.gov/11152613/>
16. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* [Internet]. 2020 May 1 [cited 2023 Jan 5];37(5):1530–4. Available from: <https://academic.oup.com/mbe/article/37/5/1530/5721363>
 17. Bayesian Information Criterion - an overview | ScienceDirect Topics [Internet]. [cited 2023 Oct 12]. Available from: <https://www.sciencedirect.com/topics/social-sciences/bayesian-information-criterion>
 18. Newick Tree Formats [Internet]. [cited 2024 Mar 18]. Available from: <http://marvin.cs.uidaho.edu/Teaching/CS515/newickFormat.html>
 19. FigTree [Internet]. [cited 2020 Oct 14]. Available from: <http://tree.bio.ed.ac.uk/software/figtree/>
 20. Baker K. Singular Value Decomposition Tutorial. 2005;
 21. Alter O, Brown PO, Botstein D. Singular Value Decomposition for Genome-Wide Expression Data Processing and Modeling. *PNAS*. 2000 Aug 29;97(18):10101–6.
 22. Trimmell AR, Kraemer SM, Mukherjee S, Phippard DJ, Janes JH, Flamoe E, et al. Global genetic diversity and evolution of var genes associated with placental and severe childhood malaria. *Mol Biochem Parasitol*. 2006 Aug 1;148(2):169–80.
 23. Mugasa J, Qi W, Rusch S, Rottmann M, Beck HP. Genetic diversity of expressed *Plasmodium falciparum* var genes from Tanzanian children with severe malaria. *Malar*

- J [Internet]. 2012 Jul 16 [cited 2024 Oct 14];11(1):1–13. Available from:
<https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-11-230>
24. Smith JD. The role of PfEMP1 adhesion domain classification in Plasmodium falciparum pathogenesis research. Mol Biochem Parasitol [Internet]. 2014 [cited 2023 Nov 7];195(2):82. Available from: </pmc/articles/PMC4159067/>
 25. Joannin N, Abhiman S, Sonnhammer EL, Wahlgren M. Sub-grouping and sub-functionalization of the RIFIN multi-copy protein family. BMC Genomics [Internet]. 2008 Jan 15 [cited 2020 Mar 9];9(1):19. Available from:
<http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-9-19>
 26. Shen XX, Li Y, Hittinger CT, Chen X xin, Rokas A. An investigation of irreproducibility in maximum likelihood phylogenetic inference. Nat Commun [Internet]. 2020 Dec 1 [cited 2024 Oct 14];11(1). Available from:
</pmc/articles/PMC7705714/>
 27. Anowar F, Sadaoui S, Selim B. Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). Comput Sci Rev. 2021 May 1;40:100378.

Chapter 5: Research Paper 2

***In-silico* mapping of hypervariable STEVOR protein sequences in *Plasmodium falciparum* for the development of a recombinant antigen library.**

Hristina Vasileva^{1,2}, Ernest Diez Benavente³, Jack Bickford-Smith¹, Anna Last^{2*}, Kevin K.A. Tetteh^{1,4*}

¹Department of Infectious Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

²Clinical Research Department, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

³Laboratory of Experimental Cardiology, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

⁴Medical Affairs Department, FIND, Geneva, Switzerland

*These authors share last authorship

Key words: malaria, *Plasmodium falciparum*, STEVOR, recombinant library, *in-silico* model, clustering

Abstract

Plasmodium falciparum pathogenesis involves complex interactions between host and parasite factors, with variant surface antigens such as members of the PfEMP1 protein family playing a pivotal role in modulating disease severity through mechanisms such as immune evasion and cytoadherence. Similarly, the less characterised STEVOR protein family, expressed on the surface of infected erythrocytes, contributes to cytoadherence and rosette formation and exhibits high antigenic variability that may aid parasite immune evasion. This study introduces an *in-silico* model for mapping global STEVOR variability for the development of a comprehensive library of STEVOR recombinant antigens to enable more in-depth investigation of STEVOR antigens involvement in immunity to infection.

Utilising all available STEVOR protein sequence data from the PlasmoDB database, this study maps the variability of STEVOR members within and across 14 *P. falciparum* strains, comprising of both clinical isolates and laboratory adapted strains. Employing bioinformatics and mathematical strategies, we designed an *in-silico* model that consistently replicates findings when applied to identical datasets. The hypervariability domain of the STEVOR protein family showed the greatest sequence variability between family members across all isolates, with a mean diversity of 52.1%, in contrast to the semi-conserved (47.8%) and the conserved (2.9%) domains.

Based on insights from the model, we constructed 11 STEVOR recombinant antigens, representing the hypervariable domain of STEVOR members predominantly from West African isolates. We further expressed them using CyDisCo co-expression plasmid in competent BL21(DE3) *Escherichia coli* expression system, establishing a pioneering library of STEVOR recombinant antigens. Additionally, three 3D7 reference strain STEVOR semi-conserved domain recombinants served as controls, exhibiting greater reactivity compared to

the variable domain recombinants. The antigenicity of the recombinants was confirmed using the multiplex magnetic bead-based assay, Luminex.

This study proposes an *in-silico* model that elucidates the spatial relationships between amino acid sequences, applicable to any organism's sequence data. Furthermore, it presents the first library of recombinant antigens of the STEVOR hypervariable domain, expressed in an *E. coli* system. This library is proposed to be further used in sero-epidemiological studies to explore if infection exposure affects the breadth of responses to members of the STEVOR family and to investigate potential variants linked to different outcomes of malaria disease.

5.1 Introduction

Malaria is a vector borne disease caused by an infection with *Plasmodium spp.*, transmitted via bites of female pregnant *Anopheles* mosquitos (1). Of the 5 species causing disease in humans, *Plasmodium falciparum* is responsible for more than 90% of the world malaria mortality, with the greatest burden felt in Sub-Saharan Africa (1).

Malaria pathogenesis depends on multiple socio-economic, host and parasite factors, resulting in varying severity of the disease from mild and asymptomatic to severe and death (2). Some parasite virulence factors are associated with evasion of the host immune response through antigenic variation and prolonged and efficient infection via cytoadherence and formation of rosettes (2),(3).

Variant surface antigens (VSA) represent several protein families synthesised at distinct phases of the parasite's blood stage life cycle which are subsequently translocated and expressed on the surface of the infected erythrocytes (IEs) (4). Members of these protein families are characterised by a high degree of variation between each other, are highly antigenic, and are found to be associated with a number of important biological processes in malaria pathogenesis, including sequestration and rosetting (3),(5). The most well-characterised protein family is the *P. falciparum* Erythrocyte Membrane Protein 1 family (*PfEMP1*). *PfEMP1* protein variants have been shown to be important for parasite survival, transmission, and virulence, due to their role in IEs cytoadherence in the microvasculature and formation of rosettes (6). Ongoing research focuses on deciphering the associations between *PfEMP1* variants, severe malaria, and their common pathogenic mechanisms with a goal of targeting these for therapeutic interventions (7).

PfEMP1 variants encoded by group A *var* genes, with an N-terminal tandem domain cassette 8 (DC8) and domain cassette 13 (DC13), have been identified to be associated with adherence

of IEs to the microvasculature via binding to the endothelial protein C receptor (EPCR), a mediator of cytoprotective effects of the active protein C (8),(9). These variants are also found to be highly upregulated in parasites associated with cerebral malaria, and are also associated with alternative sequestration sites in severe malaria, including the lungs, dermis, and the heart (10),(11). Furthermore, children with severe malaria are found to serorecognise DC8 and DC13 *PfEMP1* variants compared to those with non-complicated disease, thus detected exposure to these variants can be regarded as marker of severe malaria which offers insights into potential targets for interventions or preventative measures against children's severe malaria (11),(10). Therefore, *PfEMP1* sequence classification and diversity characterisation enables the predictions of host receptor interactions and provides foundation for the development of anti-adhesion strategies (7).

The *var* genes protein products: *PfEMP1* protein family is encoded of approximately 60 *var* genes per parasite with limited genotypic overlap, enabling the parasite to escape the host immune system and to exhibit different tissue tropisms, particularly in cerebral and placental malaria (11),(12). This remarkable diversity of variants is concentrated in few protein domains, the Duffy Binding-Like (DBL) and Cysteine-Rich Interdomain region (CIDR) domains, also associated with binding specific endothelial receptors (13). A subgroup of *PfEMP1* is found to be often expressed in severe malaria cases and particular *var* gene subset are linked to adverse clinical outcomes in malaria across all ages and geographical population, with specific *var* phenotypes associated with cerebral malaria in children (14).

Another VSA protein family, the Sub-Telomeric Variable Open Reading frame family (STEVOR) remains relatively uncharacterised (5). STEVOR is a multi-copy gene family encoded on the sub-telomeric region of all *P. falciparum* chromosomes, except for chromosome 5, with multiple variants per chromosome. A single parasite genome encodes approximately 40 *stevor* genes each coding for a different variant. However, only a small subset of these genes

is transcribed, with each parasite expressing a single 30-40 kDa variant during a polyclonal infection (5),(15).

STEVORs are found to be expressed in gametocytes, the trophozoite and schizont stages of the *P. falciparum* life cycle, and some are found in the apical tips of merozoites (16),(17).

The primary linear protein structure of STEVOR proteins is composed of a small signal peptide (SP), followed by a short variable domain (V1), a five amino acid long conserved motif encoding a protein trafficking peptide (PEXEL), a semi conserved domain (SC), a large variable domain (V2) flanked by two transmembrane domains (TM), and a short, terminal conserved domain (C), graphically represented in Figure 1 (18).

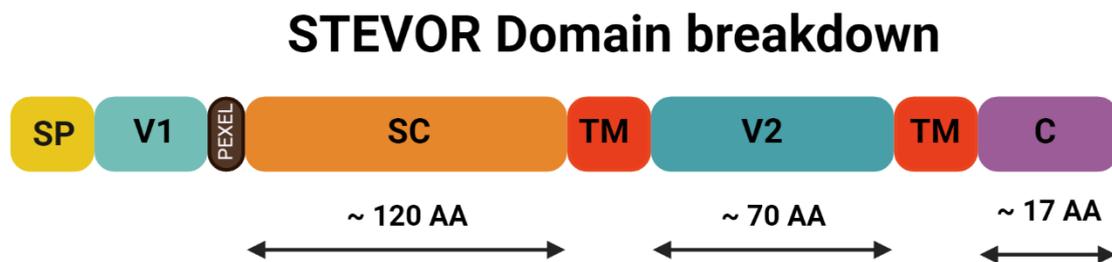


Figure 1: Schematic representation of a generic STEVOR protein. The architecture is in the following sequence: signal peptide (SP); short variable domain (V1); PEXEL motif; approximately 120 amino acid long semi-conserved domain (SC); two trans-membrane domains (TM) flanking a large hypervariable domain (V2) composed of approximately 70 amino acids, length slightly varying between variants; ending with a 17 amino acid long conserved domain (C).

Functionally STEVORs are found to be trafficked to the membrane of IE, firstly transported and cleaved in the endoplasmic reticulum and then translocated via Maurer's cleft, localising in proximity to knobs (19), (20). Previous research using antibody inhibition assays and anti-STEVOR serum on enriched rosetting-positive *P. falciparum* cultures have observed a

reduction in rosette formation. Additionally, with the use of dual-micropipette aspiration force assays it has been demonstrated that STEVOR-positive IE exhibit over four times stronger binding than controls, where further disrupted rosettes re-establish stable binding within 30 minutes, indicating that STEVOR proteins play a key role in rosette formation and stabilisation (19), (21). STEVOR protein members differ mostly by their large variable domain (V2) which is the part of the protein that protrudes to the extracellular space and has been shown to be antigenic (18),(22). This antigenic variability is an adaptive parasitic mechanism to help evade the host immune response during an infection (5),(23). Studies using peptide arrays have shown that there is an age and disease severity dependent seroreactivity and serorecognition to the STEVOR V2 and SC domains (18). Additional studies using recombinant STEVOR V2 and SC proteins have suggested the generation of an age and exposure dependent antibody responses against them, indicating their potential as markers of infection (24),(25). Antibodies targeting VSAs are considered potential candidates for both antimalarial therapeutics and vaccine development. However, there is a limited comprehension of the global antibody repertoire against STEVOR protein family, with poorly understood antibody reactivity variation in children, and lack of identified key protective targets (26). To be able to further explore the antibody responses to these antigens and their role in immunity, or for use in serosurveillance, further research is required. Specifically, a comprehensive mapping of the STEVOR protein family's diversity at the sequence level is required to investigate how antigenic properties vary and relate to infection exposure.

The aim of this study is to generate a library of STEVOR recombinant antigens which represents the repertoire of *P. falciparum* STEVOR protein family for future serological studies in exploring the role of STEVORs in immunity to *P. falciparum*.

5.2 Methods

5.2.1 Protein sequences alignment and inspection

A total of 546 STEVOR protein sequences from 14 sequenced isolates were analysed: one reference strain, six laboratory strains and seven clinical isolates, were obtained from the PlasmoDB database, corresponding to all available protein sequence ‘hits’ under the search terms of “STEVOR and *Plasmodium falciparum*” (27). Sequences from ‘ML01’ and ‘TG01’ laboratory strains were not downloaded due to the large number of corresponding ‘hits’ of 236 and 272, respectively, annotated as STEVOR/RIFIN representing both families. Moreover, ‘ML01’ and ‘TG01’ strains are known to come from complex infections with more than one parasite strain, possibly explaining the large number of sequences under STEVOR/RIFIN annotation. Another 19 sequences were excluded from this database due to wrong annotation, or considerably too long or too short sequences in comparison to the rest, leaving a database of 527 protein sequences. The new protein database was subjected to multiple alignment with MAFFT on Windows 10 terminal, using G-INS-I strategy with a global pair pairwise alignment, computed with the Smith-Waterman algorithm (28),(29). This algorithm was chosen as it is specifically tailored for sequences that have global homology (the entire length of each sequence is related to the entire length of every other sequence), as is the case for STEVORs, and the global pair is aligning the entirety of each two sequences against each other in a pairwise matter, rather than aligning only the portions of the sequences where they best match (30). The alignment included a gap extension penalty for group-to-group with a default value of 0.123, to deter the introduction of gaps unless they are truly needed, and to refine the alignment when dealing with a mixture of closely and distantly related sequences (31). Subsequently, 34 shorter protein sequences with large gaps and/or annotated as ‘pseudogenes’ were excluded, resulting in the final database of 493 STEVOR aligned protein sequences, summarised in Table 1.

5.2.2 Domain sequences isolation

Isolation of the aligned large hypervariable domain (V2), semi-conserved domain (SC) and conserved domain (C), of all 493 STEVOR protein sequences was manually performed briefly as follows. The two transmembrane domains (2xTM) flanking the V2 domain were identified using TMHMM2.0 trans-membrane domain prediction software, isolating the approximately 70 amino acids long V2 protein sequences being between amino acid 158 and amino acid 215, also supported by literature (32),(33). All TMHMM2.0 prediction plots for the selected 11 variants for the library can be found in Supplementary Figure 1. The C domain situated right after the second TM domain was isolated, identified to be 17 amino acids long. The SC domain was isolated by taking the amino acid sequences between the detected five amino acid long PEXEL motif and the first TM domain, identified to be approximately 123 amino acids long (34). Amino acid sequences per isolated domain were realigned using the same method outlined above and no further sequences were excluded from either domain group.

5.2.3 Sequence variability mapping model and variant library selection.

Amino acid sequences per domain were digitised into binary Boolean vectors. Each amino acid position for each sequence was challenged if it is one of the 21 amino acids found in *Plasmodium falciparum* proteome, with results recorded as positive = 1 and not positive = 0, resulting in 21 binary vectors, one for each amino acid. (35). All Boolean vectors were then piled into a matrix, which was further subjected to Singular Value Decomposition (SVD), resulting in a distance matrix of Euclidean distances (sPC) between sequences. sPC were used in a Principal Component Analysis (PCA) and visualised as 2D plots. All analysis and data visualisation were performed on R 4.3 computational platform. Alignment diversity for each of the isolated domains was calculated first as Shannon Entropy index, measuring the diversity at each position and second as overall mean diversity of the pairwise distances, analysed using “Biostrings” package, Bioconductor on R 4.3 computational platform (36). The V2 PCA was

further subdivided into nine equal quadrants and 11 sequences were selected to represent the STEVOR protein family, briefly as follows. The selection of sequences was based on (i) their two-dimensional position on the PCA plot and (ii) the geographical location of the isolates. In the case of variants from multiple isolates clustering together, the variant coming from a Western Sub-Saharan African isolate was selected. Three SC sequences from the *P. falciparum* reference strain 3D7 found to have potential as markers of infection exposure were also selected to serve as controls for STEVOR antigenicity (24).

A phylogenetic tree was generated using all V2 sequences with the JTT+F+G4 model and 1000 bootstrap replicates, optimised with hill-climbing nearest neighbour interchange (NNI) using IQtree 1.6.12 software, and the selected 11 library sequences were mapped on the tree (37). This phylogenetic analysis served to validate the PCA results and illustrate the relationships among the sequences. The visualisation and annotation of the phylogenetic tree was performed using iTOL: Interactive Tree of Life (Supplementary Figure 6) (38).

5.2.4 Recombinant antigens library expression.

E. coli BL21(DE3) competent cells (Trans, China) were initially transformed with pMJS226 CyDisCo plasmid (University of Oulu, Finland), containing sulfhydryl oxidase (Evr1p), disulfide bond isomerase (PDI), and chloramphenicol (CMP) resistance cassette (39),(40),(41). Transformed cells were grown in LB media with 100 µg/ml CMP and were further transformed with pGEX-5X-1(RBS) plasmids (GenScript, UK), each containing one of the 11 STEVOR V2 variants, or one of the three *Pf*3D7 STEVOR SC variants, to be expressed as N-terminal GST-tag recombinant proteins. Transformed bacteria was grown in ZY-autoinduction media with 100 µg/ml Ampicillin (Amp) and 100 µg/ml CMP, and were further lysed using LM20 microfluidizer (Analytik LTD, UK). Recombinant proteins were affinity purified in a batch mode using Glutathione Sepharose 4B beads and quantified using Bradford protein quantification assay (Bio-Rad, UK). All recombinants were run on 4-15% precast Mini-

PROTEAN TGX gels (Bio-Rad, UK) with 2% SDS and 5% β -mercaptoethanol loading buffer after undergoing denaturation at 95°C for 5 minutes. The gels were further stained with Coomassie Brilliant Blue R-250 (Merck, UK) and imaged in Bio-Rad ChemiDoc MP viewer (Supplementary Figure 2).

To confirm recombinant antigenicity, the proteins were chemically coupled to MagPlex microsphere beads (Luminex, UK), following established protocols (42). Briefly, coupled beads were titrated at six-point eight-fold dilutions, starting from 1000 $\mu\text{g/ml}$ down to 0.031 $\mu\text{g/ml}$, against a five-point two-fold dilutions of positive control serum pool of malaria hyperimmune Ugandan adults (PRISM), starting from 1/100 down to 1/1600 serum concentration (24). Data in form of median fluorescence intensity (MFI) was obtained for each titration point and a four-parameter logistic regression was applied to each titration curve to obtain the EC50 point on sigmoidal curve. The median EC50 point across all dilutions was selected as the optimum protein coupling concentration for each recombinant, titration plots per antigen can be found in the Supplementary Figure 3.

5.3 Results

5.3.1 STEVOR protein panel selection.

The total number of 493 STEVOR sequences used in this study represent 90.30% of all protein sequences ‘hits’ under the defined search terms from PlasmoDB database, summarised in Table 1 (43). The remaining 9.7% (n=53) have been removed from the database due to detected mis annotations.

Table 1: Summary of *Plasmodium falciparum* STEVOR protein sequences.

Isolate	Location	Type	PlasmoDB hits (n)	Analysis sequences (n)
3D7	Netherlands	Reference strain	43	41
Dd2	Cambodia	Laboratory strain	40	37
GA01	Gabon	Laboratory strain	43	38
GB4	Cambodia	Laboratory strain	40	36
HB3	Cambodia	Laboratory strain	32	28
IT	Cambodia	Laboratory strain	41	37
SD01	Sudan	Laboratory strain	25	22
7G8	Brazil	Clinical isolate	25	21
CD01	Congo	Clinical isolate	44	40
GN01	Guinea	Clinical isolate	46	41
KE01	Kenya	Clinical isolate	37	34
KH01	Cambodia	Clinical isolate	43	39
KH02	Cambodia	Clinical isolate	41	38
SN01	Senegal	Clinical isolate	46	41

STEVOR V2 region showed highest domain variability with overall mean diversity of pairwise distances in the alignment of 52.1% as compared to 47.8% and 2.9% for STEVOR SC and STEVOR C domains, respectively (Figure 2). Overall mean diversity of SC and V2 did not appear significantly different, however the amino-acid alignment length was significantly different of around 150 amino-acid compared to around 90 amino-acids, respectively, as demonstrated in the Shannon Entropy plot in Figure 2. For the semi-conserved region of all aligned STEVOR sequences, two specific positions exhibit notably low entropy, indicating high conservation. These regions span amino acids 60 – 71 and 118 – 129, each encompassing 11 amino acids, and display entropy values of approximately 0.3 and 0.1, respectively, as outlined in Supplementary Figure 4, focused solely on the SC region. The entropy of these two regions is significantly lower than that of other sections within the SC region, which reach values as high as 2.0 on the Shannon Entropy index, as shown in Figure 2. Further examination of SC alignment reveals that these conserved portions appear as gaps in most sequences, where the minority of variants in the alignment contain sequences aligning in these gap regions with high conservation. These variants come from different strains which shows a higher inter strain diversity compared to intra strain. This pattern suggests that semi-conservation in these regions is more likely to be variant-dependent rather than a result of inter-strain variation.

Similarly, the rest of the SC domain shows variability across groups of sequences with higher conservation between variants from different laboratory strains or isolates, rather than between variants from one strain. The major group contains 348 variants, while the smaller groups are made of 16, 15, 13, 28, 12, 29, and 32 variants, each exhibit localised conservation. This diversity distribution aligns with the PCA analysis in Figure 3B, where a primary large cluster and several smaller clusters correspond to these alignment groupings.

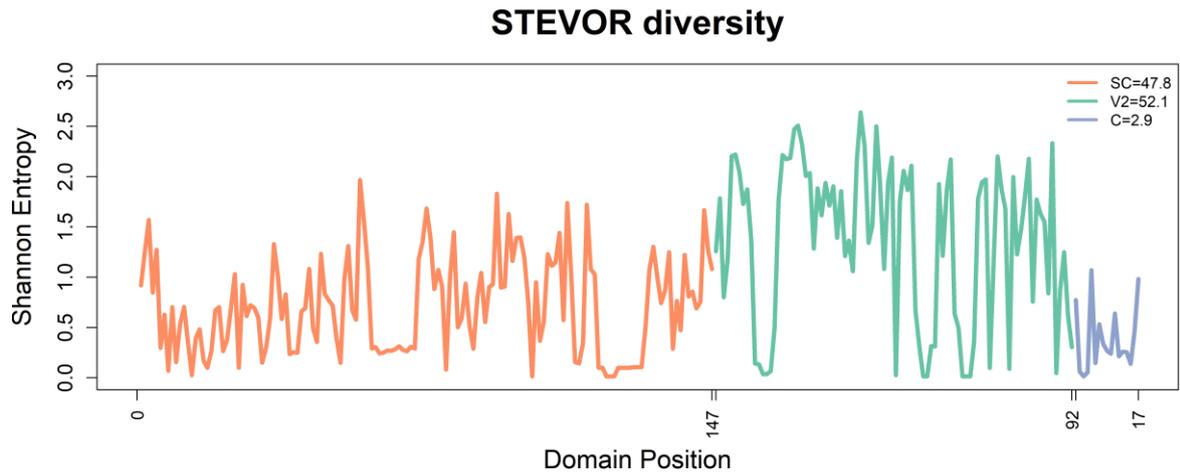


Figure 2: Shannon Entropy Index plot representing the level of diversity in each amino acid position of the STEVOR domain alignments. The higher the value, the higher the diversity. Double lines on the x-axis represent the end of one and the beginning of the next (position 0) protein domain. Legend values represent the overall mean diversity of the pairwise distances in each of the alignments.

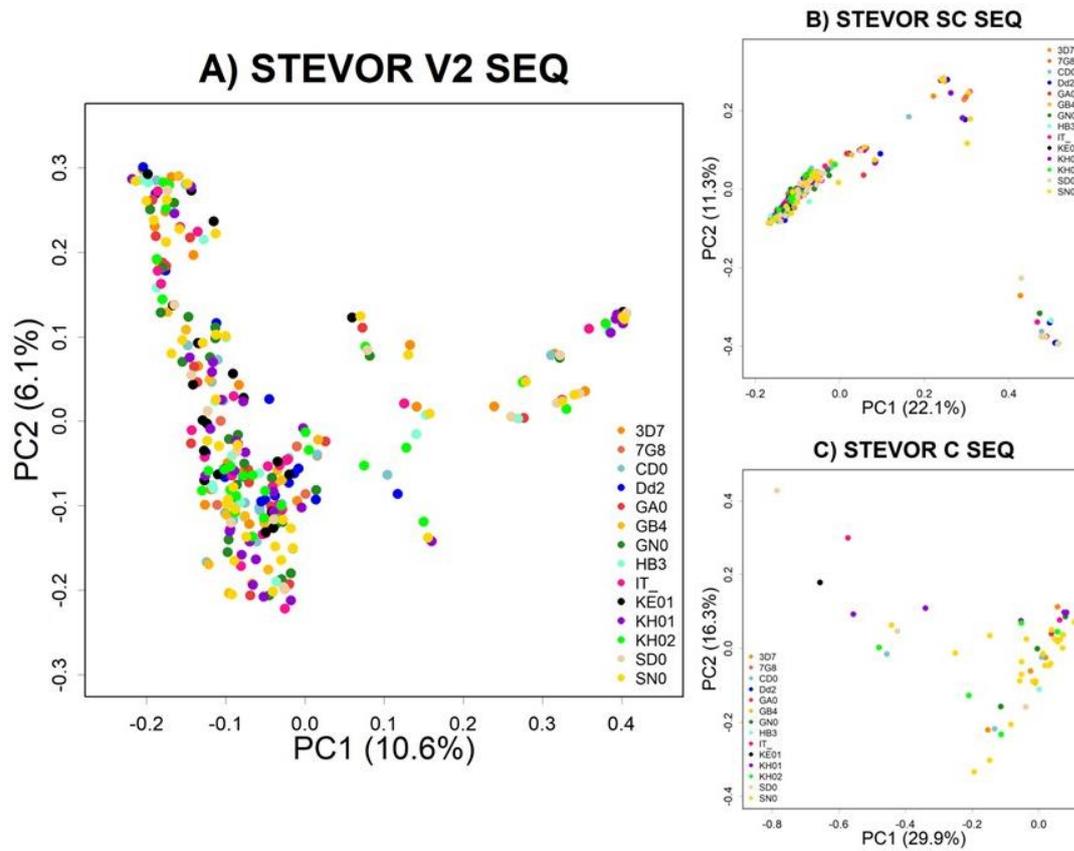


Figure 3: Principal Component Analysis showing the first two principal components (PC1 and PC2) for A) Large variable domain (V2), B) Semi-conserved domain (SC), and C) Conserved domain (C) of the STEVOR protein sequences. Sequences are colour coded according to the isolate they belong to. Percentage values represent the proportion of variation explained by each of the two principal components.

To select sequences which encompass the diversity across the STEVOR V2 sequences we selected 11 sequences (Figure 4): seven variants from West Sub-Saharan Africa (four from Guinea and three from Senegal), one variant from Central Africa: Gabon, one variant from Kenya, East Africa, and two Cambodian variants. The STEVOR library panel was selected to develop recombinant serological tools to help interrogate immunity to this protein family and expand our understanding of the STEVOR function, in addition providing a template for the generation of protein products for other hypervariable protein families of pathological importance. Since *P. falciparum* malaria is mostly found in Sub-Saharan Africa, most of the library representatives were selected from isolates from the same region. The other African variants were selected from countries with high *P. falciparum* prevalence. The remaining two variants from Cambodia were included to explore potential geographical differences in exposure to STEVOR variants. The selection of these 11 sequences was also restricted by plasmid availability, as only 80 V2 recombinant plasmids from the analysed strains were accessible, reducing the selection from 493 to 80. Available plasmids across global STEVOR variability in relation to library variants is presented in Supplementary Figure 5. Moreover, due to resource limitations, we could express and include only up to 11 recombinants in the library.

STEVOR V2 Library

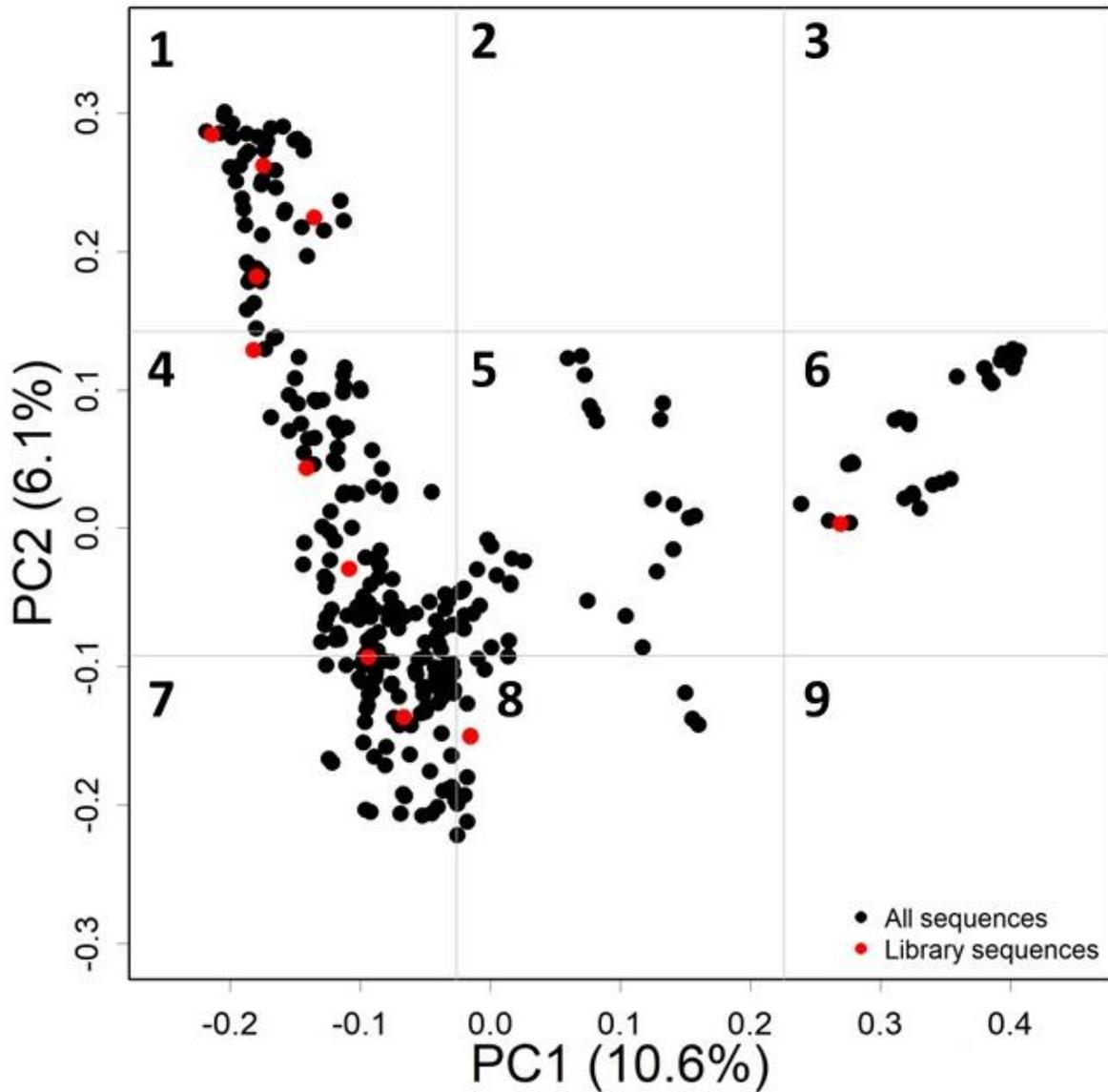


Figure 4: Visualisation of the large variable domain (V2) sequences selected to be expressed as STEVOR recombinant protein library, shown in red, among all 493 V2 amino acid sequences used in the PCA analysis (shown in black). The numbers indicate the quadrant number (labelled 1-9) in which each of the library sequences is positioned, summarised in Table 2.

5.3.2 Expression of selected recombinant proteins in *Escherichia coli*.

The 11 V2 variants were expressed as GST-tag recombinant proteins in BL21(DE3) *E. coli* competent cells, co-transformed with the CyDisCo plasmid (pMJS226). The sequences ranged between 272 and 283 amino-acids in length and expressed as recombinant protein products between 32 and 33 kDa, including the 211 amino acids long, 26 kDa GST-tag, summarised in Table 2. Although of similar sizes, the V2 recombinants varied substantially in terms of expression concentrations between 1.45 mg/ml (SN01_115) and 10.51 mg/ml (IT_13776 and KE01_158). However, the differences in expression concentrations were not reflected in the optimum protein-bead coupling concentrations for the MagPlex assay (median EC50 point), demonstrating a recombinant specific seroreactivity. Coomassie stained SDS-PAGE annotated gel images are included as Supplementary Figure 2.

Table 2: Summary of STEVOR V2 library recombinant antigens.

Protein Name (PlasmoDB)	Abbreviated protein name	Quadrate Number (PCA)	Length (amino acids)*	Molecular size (kDa)*	Recombinant concentration (mg/ml)	EC50 point (mg/ml)
PfGN01_040031100	GN01_4311	Q1	283	33	1.87	0.46
PfIT_130077600	IT_13776	Q1	273	32	10.51	0.48
PfSN01_140006300	SN01_1463	Q1	277	32	6.19	0.33
PfGN01_100006100	GN01_161	Q4	283	33	4.12	0.15
PfKE01_100005800	KE01_158	Q4	272	32	10.51	1.17
PfSN01_030005600	SN01_356	Q4	277	32	3.28	1.56
PfHB3_040028900	HB3_4289	Q6	279	33	8.63	0.20
PfGA01_070034600	GA01_7346	Q7	281	32	5.62	1.97
PfGN01_020006800	GN01_268	Q7	278	32	3.77	0.21
PfSN01_000011500	SN01_115	Q7	273	32	1.45	0.73
PfGN01_130006600	GN01_1366	Q8	279	32	4.33	0.58

***The length and weight of the recombinants represent the V2 domain including the 211 amino acids long, 26 kDa GST molecular tag.**

The SC recombinant proteins ranged between 334 and 340 amino acids in length with size ranging between 40 and 41 kDa (Table 3). The significantly much lower EC50 concentrations calculated for the SC recombinants highlight that they exhibit much higher antigenicity than the V2 recombinants.

Table 3: Summary of STEVOR SC recombinant antigens.

Protein Name (PlasmoDB)	Abbreviated protein name	Length (amino acids)*	Molecular size (kDa)*	Recombinant concentration (mg/ml)	EC50 point (mg/ml)
Pf3D7_1300900	3D7_139	336	40	1.91	0.01
Pf3D7_0832000	3D7_832	340	41	5.34	0.24
Pf3D7_0832600	3D7_8326	334	41	3.28	0.04

***The length and weight of the recombinants represent the V2 domain including the 211 amino acids long, 26 kDa GST molecular tag.**

5.4 Discussion

We developed an *in-silico* model, simplifying the complex amino-acid polymorphism inherent in hypervariable gene families. This novel approach allowed us to develop the first recombinant protein STEVOR variant library in a *E. coli* expression system, providing novel serological tools with which to explore the human immune response to the understudied but associated with pathology STEVOR protein family.

Other studies investigating the variable domain of STEVORs and dissecting the protein structure have been mostly focused on laboratory strains such as 3D7, Dd2 and HB3 (18),(44). However, the variability between STEVORs is firstly between variants within isolates, termed antigenic variability, and secondly between the variants across isolates, due to high recombination rate in the *P. falciparum* genome, which we have demonstrated in the multiple alignment provided as a supplementary documentation (44),(45). Additionally, the SC domain alignment analysis and the V2 domain PCA plot revealed that STEVOR variability is greater among variants within the same strain than between different strains, indicating higher antigenic diversity within strains rather than across strains. Since the majority of the protein sequence of STEVOR variants is highly conserved, exploring the variability relationship between variants using the full protein sequence would reduce the resolution level in examining the differences. Hence, after the initial alignment, the V2 domain was isolated for the variability analysis. The SC and C domains were also isolated to test the reliability of the model, expecting closer relationship due to higher conservation, resulting in lower variability. Furthermore, it is worth considering that the expression of proteins in *E. coli* that encompass domains such as signal peptides, PEXEL motifs, and trans-membrane domains, would most likely result in packaging of the proteins in inclusion bodies, thereby rendering them insoluble (46).

Since protein sequences are qualitative data, the numerical conversion of the amino acid sequences into binary Boolean vectors with a quantifiable score allowed for the illustration of the relationships between variants in a quantitative way (35). The Singular Value Decomposition method was used to construct the Euclidean distance matrix from the Boolean vectors data. This method was chosen as it effectively converts interrelated variables into a set of orthogonal variables, while identifying and arranging the dimensions along which data points display the greatest variability, therefore finding the optimum approximation of the original data points using reduced number of dimensions. (47). Furthermore, protein sequence data is a multivariable data due to the multiple types of amino acids present, therefore using a multivariate analysis of the distance matrix such as PCA was appropriate to be able to observe the matrix data without harming its structure (48). The direct PCA analysis rotates the data and projects it to set of diagonal axes, finding direction of differences considering both the amino acid composition and the genetic distance between the samples (48). Moreover, this analysis demonstrated high reproducibility when performed on the same dataset, unlike other sequence data clustering methods such as phylogenetic trees and *k*-means clustering, exhibiting variance in their outcomes with each permutation or with the change of numbers of *k*-means used, respectively (49).

The importance of conducting variant down selection for the STEVOR library from the clustering model was derived from practical constrains, specifically the inability to express every variant from each strain due to resource and time limitations, as well as the concern of high likelihood of cross reactivity between closely related variants when further tested on samples. A method of creating specific monoclonal antibodies against variants which are close to each other in the PCA analysis could be performed to further calculate the difference in reactivity to measure the distance needed between variants for the lowest cross-reactivity. However, this method is highly resource intensive approach beyond the scope of this study.

The PCA analysis indicates that the majority of the variation in the V2 domain is accounted for by the first principal component (PC1). However, most of the sequences cluster primarily in quadrants one (Q1, four (G4) and seven (Q7), with minimal variation along PC1. These sequences show distinct differences when viewed through the second principal component (PC2). From each of these three quadrants we have selected two to four variants for the library, majority coming from West Sub-Saharan African stains. The last variant coming from a laboratory strain from Cambodia, 'HB3_4289', was selected from Q6. This variant was selected to introduce geographical diversity into the recombinant library, as well as due to the high prevalence of Cambodian laboratory strains variants in the sequence database.

The further three variants selected to be expressed, represented the SC domain of STEVOR variants from the *P. falciparum* reference strain 3D7. The reason behind expressing SC domains alongside the V2 library is to serve as library controls, expected to be recognised by a diverse panel of malaria immune sera due to the conservation of the domain (18). The specific selection of the three variants: 'Pf3D7_1300900', 'Pf3D7_0832000' and 'Pf3D7_0832600' was based on previously published study showing a potential of those SC variants recombinants as markers of *P. falciparum* exposure (18),(24).

The decision was made to create a library of recombinant STEVOR proteins instead of applying sequencing methods on the isolates or employing peptide arrays. This choice was based on several factors. Firstly, the *stevor* genes are located on the sub telomeric region of the *P. falciparum* chromosomes, adjacent to the telomer, where abundant repetitions of heterogeneous sequences pose challenges for mapping and sequencing (50). Furthermore, using recombinant proteins for the detection of antibody responses to the variants would offer a more precise evaluation of serorecognition compared to the peptide arrays, since recombinant proteins possess a folded structure while peptide arrays are made of only primary linear structure. Using a CyDisCo plasmid technology to transform *E. coli* competent cells for the

expression of the recombinants was selected to improve the chances of proper protein folding and post-translational modifications (39). Proteins expressed in the cytoplasm of *E. coli* often require the formation of disulfide bonds between an even number of cysteine molecules (51). Manual examination of the V2 variant amino acid sequences confirmed that they contain 2 cysteine molecules. The CyDisCo plasmid is therefore likely to benefit the correct folding of the proteins in the cytoplasm of *E. coli* due to the co-expression of sulfhydryl oxidase (Evr1p), allowing the production of disulfide-bonded proteins. Additionally, the expression of disulfide bond isomerase (PDI) ensures the editing of the disulfide bonds during protein folding via cleavage and formation cycles, a type of post-translational modification (39),(40).

The study utilises the Luminex serological assay to verify the antigenic properties of the expressed recombinant proteins. All recombinant proteins possess differing reactivity characteristics to the PRISM serum pool positive control from Uganda (24). Specifically, ‘SN01_115’ and ‘SN01_356’ exhibit the lowest reactivity while ‘GA01_7346’ displays the highest for the V2 recombinants. The GA01 isolate comes from Gabon, geographically closer to Uganda, whereas SN01 strain comes from Senegal. This geographical difference may account for the observed disparity in seroreactivity between variants from the two strains. Furthermore, the protein concentration of the recombinants can also be linked to this diversity in reactivity signal. As expected, the SC recombinants exhibited considerably higher serorecognition when compared to all of the V2 recombinants, due to the substantial conservation of the domain across all STEVORs, and an expected cross reactivity between SC domains from different STEVOR variants.

5.5 Limitations

Selection of variants for the library from the database of 493 STEVOR V2 protein sequences was based on their position in the PCA analysis and focused primarily on variants from West Sub-Saharan Africa, as this library is intended for testing against serological samples coming from populations of the same region. However, the library of these 11 STEVOR variants, may not fully capture the diversity of the protein family based on amino acid sequences, since 10 out of the 11 recombinants do not vary significantly according to sPC1, although most of all variants were clustering according to the sPC1. The phylogenetic tree shown in the supplementary material highlights the imbalance in clade representation, with some clades overrepresented and others entirely absent. This discrepancy arises from the limited availability of plasmids of 80 outlined in Supplementary Figure 5, specifically those mapping to quadrant five in the PCA analysis. The available plasmids from Q5, which also cluster into a distinct clade in Supplementary Figure 6, exhibit a unique single transmembrane domain structure, setting them apart from the other variants selected for the library, hence were not selected to be expressed.

Additionally, the *in-silico* model does not account for amino acid properties such as polarity, charge size and shape, which could potentially enhance the accuracy of diversity mapping. Incorporating these properties might produce a different distribution of variants across PCA coordinates.

The selected PRISM positive control to assess the reactivity of these variants is uncertain as we don't know whether the serum pool from the Ugandan hyperimmune individuals would specifically recognise the selected variants. On the contrary, we also anticipated some degree of cross-reactivity among the V2 variants. Unfortunately, similar to the case for other variable antigen families, such as RIFIN and *Pf*EMP1, there is no definitive mathematical or biological method to gauge the necessary level of divergence between the variants to minimise the

potential cross-reactivity, in order to identify key targets of protective immunity when further using the library against serum samples. Additionally, further research of inter-strain conserved variants should be performed and those identified should be added to the library going forward.

5.6 Future work

To incorporate biological reasoning into mapping global STEVOR variability using the computational model, it is essential to include information on the physicochemical properties of amino acids. Changes in amino acids can alter protein folding, charge, interactions with other molecules, immune recognition, and overall function and structure. Including these properties could potentially further distinguish the variants from one another, increasing the explained variability. Currently, only 16% of variability is captured by the first two principal components in the PCA, with variability dispersed across all sPCs.

5.7 Conclusion

The approach described in this study provides a robust *in-silico* model to help interrogate complex sequence dynamics by highlighting the spatial relationship between sequences. This model is applicable to any amino acid sequence alignment data with high reproducibility when performing permutations, irrespective of the species. Additionally, this is the first study to generate a recombinant antigen library, expressed in *E. coli* system, representing the variable domain of the surface expressed protein family STEVOR. The protein library tool of V2 variants could be a valuable resource for subsequent investigations. It can be employed to explore the degree of seroreactivity and recognition using serum or plasma samples from various populations, potentially identifying variants associated with enhanced parasite fitness, or clinical disease outcome, warranting further exploration through functional assays *in vitro*. This chapter, presented in a scientific manuscript format, outlines a methodology for selecting STEVOR variants to represent the global variation of the protein family using the available

protein sequence database. However, addressing the limitations and implementing the suggested future work would strengthen the analysis and increase confidence in selecting accurate/additional variants for the library.

5.8 Declarations

5.8.1 Ethical approval and consent to participate.

The serum samples used in this study come from the Program for Resistance, Immunology, Surveillance, and Modelling of Malaria in Uganda longitudinal cohort (PRISM), approved by the Ethics Committee of LSHTM (reference number: 15823) and the Research and Ethics Committee of the Makerere University School of Medicine in Kampala, Uganda (reference number: 2011-167).

5.8.2 Availability of data

The data sets supporting the conclusions of this article are included within the article and its additional files. The computational method is included in as an additional file to this manuscript and can be found in the Appendices section of the thesis.

5.8.3 Competing interests.

All authors declare that the research was conducted in the absence of any financial, personal, or professional relationships that could be construed to have influenced the work.

5.8.4 Funding

This project is nested within the clinical trial: “Adjunctive Ivermectin Mass Drug Administration for Malaria Control (MATAMAL)” ClinicalTrials.gov Identifier: NCT04844905, funded by the Joint Global Health Trial scheme (JGHTs) with PI Dr Anna Last, (Grant Code MR/S005013/1). This scheme is jointly funded by the Foreign, Commonwealth and Development Office (FCDO), the Medical Research Council (MRC), the National Institute

for Health Research (NIHR) and the Wellcome Trust foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

5.8.5 Authors contributions

HV was responsible for the design, acquisition and analysis, interpretation of results and drafting of the work. KT was involved in the design of the work and interpretation of results. EB was involved in the design of the work. JBS was involved in laboratory work. AL and KT were responsible for securing funds for the study. All authors have approved the submitted version and have agreed to be personally accountable for their own contributions of this work.

5.8.6 Acknowledgements

Special acknowledgements to Dr James Ashall for contributing to the graphical illustrations of this study.

5.9 List of abbreviations

- STEVOR: Sub telomeric Open Reading Frame protein family
- PfEMP1: *Plasmodium falciparum* Erythrocyte Membrane Protein 1
- WHO: World Health Organisation
- LLIN: Long-Lasting Insecticidal Net
- MSP1: Merozoite Surface Protein 1
- AMA1: Apical Membrane Antigen 1
- VSA: Variable Surface Antigens
- IE: Infected Erythrocytes
- RIFIN: Repetitive Interspersed protein family
- DC: Domain Cassette
- EPCR: Endothelial Protein C Receptor
- DBL: Duffy-Binding Like domain
- CIDR: Cysteine-Rich Interdomain Region
- SP: Signal Peptide
- V1: Small Variable domain
- PEXEL: Translocating Element
- SC: Semi-Conserved domain
- V2: Large Variable domain
- TM: Trans-Membrane domain
- C: Conserved domain
- SVD: Singular Value Decomposition
- sPC: Principal Component
- PCA: Principal Component Analysis

- CMP: Chloramphenicol
- LB: Lysogeny Broth
- GST: Glutathione S-transferase
- MFI: Median Fluorescent Intensity
- Q: Quadrant

5.10 References

1. World Health Organization. World Malaria Report 2023 [Internet]. 2023. Available from: https://cdn.who.int/media/docs/default-source/malaria/world-malaria-reports/world-malaria-report-2023-spreadview.pdf?sfvrsn=bb24c9f0_4
2. Miller LH, Baruch DI, Marsh K, Doumbo OK. The pathogenic basis of malaria. *Nature* [Internet]. 2002 Feb 7 [cited 2021 Nov 17];415(6872):673–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/11832955/>
3. Niang M, Bei AK, Madnani KG, Pelly S, Dankwa S, Kanjee U, et al. The variant STEVOR protein of *Plasmodium falciparum* is a red cell binding protein important for merozoite invasion and rosetting. *Cell Host Microbe* [Internet]. 2014 Jul 7 [cited 2023 Aug 29];16(1):81. Available from: </pmc/articles/PMC4382205/>
4. Chan JA, Fowkes FJI, Beeson JG. Surface antigens of *Plasmodium falciparum*-infected erythrocytes as immune targets and malaria vaccine candidates. [Internet]. Vol. 71, Cellular and molecular life sciences : CMLS. Springer; 2014 [cited 2020 Oct 16]. p. 3633–57. Available from: <https://link.springer.com/article/10.1007/s00018-014-1614-3>
5. Scherf A, Lopez-Rubio JJ, Riviere L. Antigenic Variation in *Plasmodium falciparum*. *Annu Rev Microbiol* [Internet]. 2008 Oct 11 [cited 2019 May 20];62(1):445–70. Available from: <http://www.annualreviews.org/doi/10.1146/annurev.micro.61.080706.093134>
6. Hayward RE, Tiwari B, Piper KP, Baruch DI, Day KP. Virulence and transmission success of the malarial parasite *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* [Internet]. 1999 Apr 4 [cited 2023 Dec 8];96(8):4563. Available from: </pmc/articles/PMC16372/>

7. Chan JA, Boyle MJ, Moore KA, Reiling L, Lin Z, Hasang W, et al. Antibody Targets on the Surface of *Plasmodium falciparum*-Infected Erythrocytes That Are Associated With Immunity to Severe Malaria in Young Children. *J Infect Dis* [Internet]. 2019 Mar 3 [cited 2023 Nov 7];219(5):819. Available from: [/pmc/articles/PMC6376912/](#)
8. Lavstsen T, Turner L, Saguti F, Magistrado P, Rask TS, Jespersen JS, et al. *Plasmodium falciparum* erythrocyte membrane protein 1 domain cassettes 8 and 13 are associated with severe malaria in children. *Proc Natl Acad Sci U S A* [Internet]. 2012 Jun 26 [cited 2024 Jan 5];109(26):E1791. Available from: [/pmc/articles/PMC3387094/](#)
9. Turner L, Lavstsen T, Berger SS, Wang CW, Petersen JEV, Avril M, et al. Severe malaria is associated with parasite binding to endothelial protein C receptor. *Nature* [Internet]. 2013 Jun 6 [cited 2024 Jan 5];498(7455):502. Available from: [/pmc/articles/PMC3870021/](#)
10. Claessens A, Adams Y, Ghumra A, Lindergard G, Buchan CC, Andisi C, et al. A subset of group A-like var genes encodes the malaria parasite ligands for binding to human brain endothelial cells. *Proc Natl Acad Sci U S A* [Internet]. 2012 Jun 26 [cited 2024 Jan 5];109(26):E1772. Available from: [/pmc/articles/PMC3387129/](#)
11. Avril M, Tripathi AK, Brazier AJ, Andisi C, Janes JH, Soma VL, et al. A restricted subset of var genes mediates adherence of *Plasmodium falciparum*-infected erythrocytes to brain endothelial cells. *Proc Natl Acad Sci U S A* [Internet]. 2012 Jun 26 [cited 2024 Jan 5];109(26):E1782. Available from: [/pmc/articles/PMC3387091/](#)
12. Kraemer SM, Kyes SA, Aggarwal G, Springer AL, Nelson SO, Christodoulou Z, et al. Patterns of gene recombination shape var gene repertoires in *Plasmodium falciparum*: Comparisons of geographically diverse isolates. *BMC Genomics* [Internet]. 2007 Feb 7 [cited 2023 Nov 7];8(1):1–18. Available from:

<https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-8-45>

13. Smith JD. The role of PfEMP1 adhesion domain classification in Plasmodium falciparum pathogenesis research. *Mol Biochem Parasitol* [Internet]. 2014 [cited 2023 Nov 7];195(2):82. Available from: [/pmc/articles/PMC4159067/](https://pubmed.ncbi.nlm.nih.gov/24159067/)
14. Duffy F, Bernabeu M, Babar PH, Kessler A, Wang CW, Vaz M, et al. Meta-analysis of Plasmodium falciparum var Signatures Contributing to Severe Malaria in African Children and Indian Adults. *MBio* [Internet]. 2019 Mar 1 [cited 2023 Nov 7];10(2). Available from: [/pmc/articles/PMC6495371/](https://pubmed.ncbi.nlm.nih.gov/336495371/)
15. Kaviratne M, Khan SM, Jarra W, Preiser PR. Small variant STEVOR antigen is uniquely located within Maurer's clefts in Plasmodium falciparum-infected red blood cells. *Eukaryot Cell* [Internet]. 2002 Dec [cited 2020 Oct 7];1(6):926–35. Available from: <https://pubmed.ncbi.nlm.nih.gov/12477793/>
16. Wichers JS, Scholz JAM, Strauss J, Witt S, Lill A, Ehnold LI, et al. Dissecting the gene expression, localization, membrane topology, and function of the plasmodium falciparum STEVOR protein family. *MBio*. 2019 Aug 27;10(4).
17. Blythe JE, Xue YY, Kuss C, Bozdech Z, Holder AA, Marsh K, et al. Plasmodium falciparum STEVOR proteins are highly expressed in patient isolates and located in the surface membranes of infected red blood cells and the apical tips of merozoites. *Infect Immun*. 2008 Jul;76(7):3329–36.
18. Zhou AE, Berry AA, Bailey JA, Pike A, Dara A, Agrawal S, et al. Antibodies to Peptides in Semiconserved Domains of RIFINs and STEVORs Correlate with Malaria Exposure. *mSphere*. 2019 Mar 20;4(2).
19. Singh H, Madnani K, Lim YB, Cao J, Preiser PR, Lim CT. Expression dynamics and

- physiologically relevant functional study of STEVOR in asexual stages of *Plasmodium falciparum* infection. *Cell Microbiol* [Internet]. 2017 Jun 1 [cited 2023 Nov 8];19(6). Available from: <https://pubmed.ncbi.nlm.nih.gov/28030753/>
20. Russo I, Babbitt S, Muralidharan V, Butler T, Oksman A, Goldberg DE. Plasmeprin V licenses *Plasmodium* proteins for export into the host erythrocyte. *Nature* [Internet]. 2010 Feb 2 [cited 2024 Jan 5];463(7281):632. Available from: </pmc/articles/PMC2826791/>
 21. Sherman IW, Eda S, Winograd E. Cytoadherence and sequestration in *Plasmodium falciparum*: defining the ties that bind. *Microbes Infect*. 2003 Aug 1;5(10):897–909.
 22. Cheng Q, Cloonan N, Fischer K, Thompson J, Waine G, Lanzer M, et al. *stevor* and *rif* are *Plasmodium falciparum* multicopy gene families which potentially encode variant antigens. *Mol Biochem Parasitol* [Internet]. 1998 Nov 30 [cited 2020 Oct 7];97(1–2):161–76. Available from: <https://pubmed.ncbi.nlm.nih.gov/9879895/>
 23. Lavazec C, Sanyal S, Templeton TJ. Hypervariability within the *Rifin*, *Stevor* and *Pfmc-2TM* superfamilies in *plasmodium falciparum*. *Nucleic Acids Res* [Internet]. 2006 Dec [cited 2020 Oct 5];34(22):6696–707. Available from: </pmc/articles/PMC1751529/?report=abstract>
 24. Vasileva H, Chopo-Pizarro A, Ooko M, Dumont E, Wu L, Ssewanyana I, et al. Novel hypervariable erythrocyte surface expressed recombinant proteins show promise as serological markers of exposure to *Plasmodium falciparum* infection. *bioRxiv* [Internet]. 2023 Nov 2 [cited 2023 Nov 7];2023.10.31.564916. Available from: <https://www.biorxiv.org/content/10.1101/2023.10.31.564916v1>
 25. Schreiber N, Khattab A, Petter M, Marks F, Adjei S, Kobbe R, et al. Expression of *Plasmodium falciparum* 3D7 STEVOR proteins for evaluation of antibody responses

- following malaria infections in naïve infants. *Parasitology* [Internet]. 2008 Feb [cited 2020 Oct 16];135(2):155–67. Available from:
<https://pubmed.ncbi.nlm.nih.gov/17931459/>
26. Kanoi BN, Nagaoka H, White MT, Morita M, Palacpac NMQ, Ntege EH, et al. Global Repertoire of Human Antibodies Against *Plasmodium falciparum* RIFINs, SURFINs, and STEVORs in a Malaria Exposed Population. *Front Immunol* [Internet]. 2020 May 12 [cited 2023 Aug 24];11. Available from:
<https://pubmed.ncbi.nlm.nih.gov/32477363/>
 27. PlasmoDB : The Plasmodium Genomics Resource [Internet]. [cited 2020 Mar 30]. Available from: <https://plasmodb.org/plasmo/>
 28. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res* [Internet]. 2019 Jul 7 [cited 2023 Jun 20];47(W1):W5. Available from: [/pmc/articles/PMC6602451/](https://pubmed.ncbi.nlm.nih.gov/31211111/)
 29. Katoh K, Kuma KI, Toh H, Miyata T. MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* [Internet]. 2005 [cited 2020 Oct 16];33(2):511–8. Available from: [/pmc/articles/PMC548345/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/15655203/)
 30. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* [Internet]. 2013 Apr [cited 2023 Oct 31];30(4):772. Available from: [/pmc/articles/PMC3603318/](https://pubmed.ncbi.nlm.nih.gov/23745288/)
 31. Durbin R, Eddy SR, Krogh A, Mitchison G. Multiple sequence alignment methods. *Biol Seq Anal* [Internet]. 1998 Jul 6 [cited 2023 Oct 31];135–60. Available from:
<https://www.cambridge.org/core/books/biological-sequence-analysis/multiple-sequence-alignment-methods/217DF949061D16E0D2A4B034233DA7F4>

32. Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* [Internet]. 2001 Jan 19 [cited 2023 Jan 5];305(3):567–80. Available from: <https://pubmed.ncbi.nlm.nih.gov/11152613/>
33. Hypervariability within the Rifin, Stevor and Pfmc-2TM superfamilies in *Plasmodium falciparum* [Internet]. [cited 2020 Mar 9]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1751529/>
34. Boddey JA, Moritz RL, Simpson RJ, Cowman AF. Role of the *Plasmodium* export element in trafficking parasite proteins to the infected erythrocyte. *Traffic* [Internet]. 2009 [cited 2020 Oct 7];10(3):285–99. Available from: [/pmc/articles/PMC2682620/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/pmc/articles/PMC2682620/?report=abstract)
35. Konishi T, Matsukuma S, Fuji H, Nakamura D, Satou N, Okano K. Principal Component Analysis applied directly to Sequence Matrix. *Sci Rep* [Internet]. 2019 Dec 1 [cited 2021 Apr 8];9(1):1–13. Available from: <https://doi.org/10.1038/s41598-019-55253-0>
36. Pagès H, Aboyou P GR. Biostrings: Efficient manipulation of biological strings version 2.58.0 from Bioconductor [Internet]. 2003 [cited 2023 Nov 1]. Available from: <https://rdrr.io/bioc/Biostrings/>
37. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* [Internet]. 2020 May 1 [cited 2023 Jan 5];37(5):1530–4. Available from: <https://academic.oup.com/mbe/article/37/5/1530/5721363>
38. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: Recent updates to the

- phylogenetic tree display and annotation tool. *Nucleic Acids Res.* 2024 Jul 5;52(W1):W78–82.
39. Gaciarz A, Khatri NK, Velez-Suberbie ML, Saaranen MJ, Uchida Y, Keshavarz-Moore E, et al. Efficient soluble expression of disulfide bonded proteins in the cytoplasm of *Escherichia coli* in fed-batch fermentations on chemically defined minimal media. *Microb Cell Fact* [Internet]. 2017 Jun 15 [cited 2023 Aug 28];16(1):1–12. Available from:
<https://microbialcellfactories.biomedcentral.com/articles/10.1186/s12934-017-0721-x>
40. Matos CFRO, Robinson C, Alanen HI, Prus P, Uchida Y, Ruddock LW, et al. Efficient export of prefolded, disulfide-bonded recombinant proteins to the periplasm by the Tat pathway in *Escherichia coli* CyDisCo strains. *Biotechnol Prog* [Internet]. 2014 [cited 2024 Jun 11];30(2):281–90. Available from:
<https://pubmed.ncbi.nlm.nih.gov/24376243/>
41. Tungekar AA, Recacha R, Ruddock LW. Production of neutralizing antibody fragment variants in the cytoplasm of *E. coli* for rapid screening: SARS-CoV-2 a case study. *Sci Rep* [Internet]. 2023 Dec 1 [cited 2024 Jun 11];13(1). Available from:
</pmc/articles/PMC10019796/>
42. Wu L, Hall T, Ssewanyana I, Oulton T, Patterson C, Vasileva H, et al. Optimisation and standardisation of a multiplex immunoassay of diverse *Plasmodium falciparum* antigens to assess changes in malaria transmission using sero-epidemiology. *Wellcome Open Res* [Internet]. 2020 Apr 23 [cited 2020 Oct 15];4:26. Available from:
<https://doi.org/10.12688/wellcomeopenres.14950.1>
43. Otto TD, Böhme U, Sanders M, Reid A, Bruske EI, Duffy CW, et al. Long read assemblies of geographically dispersed *Plasmodium falciparum* isolates reveal highly

- structured subtelomeres. Wellcome Open Res [Internet]. 2018 [cited 2023 Dec 5];3.
Available from: [/pmc/articles/PMC5964635/](#)
44. Catherine Lavazec, Sohini Sanyal and TJT. Hypervariability within the Rifin, Stevor and Pfmc-2TM superfamilies in Plasmodium falciparum. Nucleic Acids Res [Internet]. 2006;34(22):6696–707. Available from:
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1751529/>
45. Jiang H, Li N, Gopalan V, Zilversmit MM, Varma S, Nagarajan V, et al. High recombination rates and hotspots in a Plasmodium falciparum genetic cross. Genome Biol [Internet]. 2011 Apr 4 [cited 2023 Sep 13];12(4):R33. Available from:
[/pmc/articles/PMC3218859/](#)
46. Bhatwa A, Wang W, Hassan YI, Abraham N, Li XZ, Zhou T. Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in Escherichia coli and Strategies to Address Them for Industrial Applications. Front Bioeng Biotechnol [Internet]. 2021 Feb 10 [cited 2023 Jul 18];9. Available from:
[/pmc/articles/PMC7902521/](#)
47. Baker K. Singular Value Decomposition Tutorial. 2005;
48. Jolliffe I, Jolliffe, Ian. Principal Component Analysis. In: Wiley StatsRef: Statistics Reference Online [Internet]. Chichester, UK: John Wiley & Sons, Ltd; 2014 [cited 2017 Oct 18]. Available from: <http://doi.wiley.com/10.1002/9781118445112.stat06472>
49. k-Means Advantages and Disadvantages | Machine Learning | Google for Developers [Internet]. [cited 2023 Oct 10]. Available from:
<https://developers.google.com/machine-learning/clustering/algorithm/advantages-disadvantages>

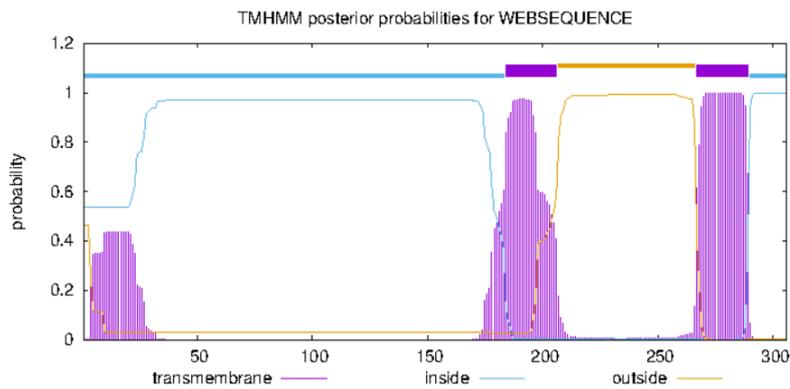
50. Kwapisz M, Morillon A. Subtelomeric Transcription and its Regulation. *J Mol Biol* [Internet]. 2020 Jul 7 [cited 2023 Aug 29];432(15):4199. Available from: [/pmc/articles/PMC7374410/](https://pubmed.ncbi.nlm.nih.gov/374410/)
51. Berkmen M. Production of disulfide-bonded proteins in *Escherichia coli*. *Protein Expr Purif*. 2012 Mar 1;82(1):240–51.

5.11 Supplementary Materials

PfGN01_004003110-t41_1

TMHMM result

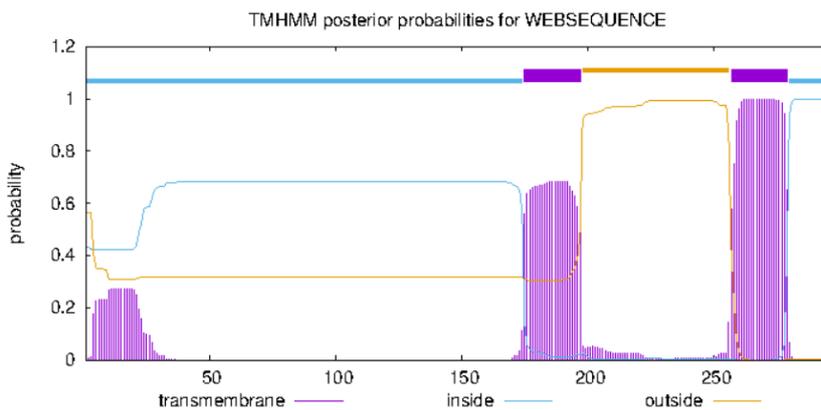
```
# WEBSEQUENCE Length: 306
# WEBSEQUENCE Number of predicted TMHs: 2
# WEBSEQUENCE Exp number of AAs in TMHs: 53.46687
# WEBSEQUENCE Exp number, first 60 AAs: 8.92912
# WEBSEQUENCE Total prob of N-in: 0.53838
WEBSEQUENCE TMHMM2.0 inside 1 183
WEBSEQUENCE TMHMM2.0 TMhelix 184 206
WEBSEQUENCE TMHMM2.0 outside 207 266
WEBSEQUENCE TMHMM2.0 TMhelix 267 289
WEBSEQUENCE TMHMM2.0 inside 290 306
```



PfIT_130077600-t41_1

TMHMM result

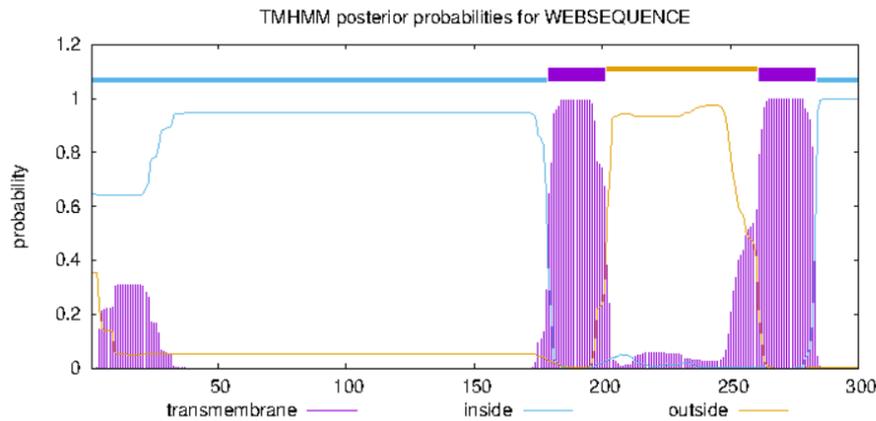
```
# WEBSEQUENCE Length: 296
# WEBSEQUENCE Number of predicted TMHs: 2
# WEBSEQUENCE Exp number of AAs in TMHs: 43.63084
# WEBSEQUENCE Exp number, first 60 AAs: 5.37615
# WEBSEQUENCE Total prob of N-in: 0.43571
WEBSEQUENCE TMHMM2.0 inside 1 174
WEBSEQUENCE TMHMM2.0 TMhelix 175 197
WEBSEQUENCE TMHMM2.0 outside 198 256
WEBSEQUENCE TMHMM2.0 TMhelix 257 279
WEBSEQUENCE TMHMM2.0 inside 280 296
```



PfSN01_140006300-t41_1

TMHMM result

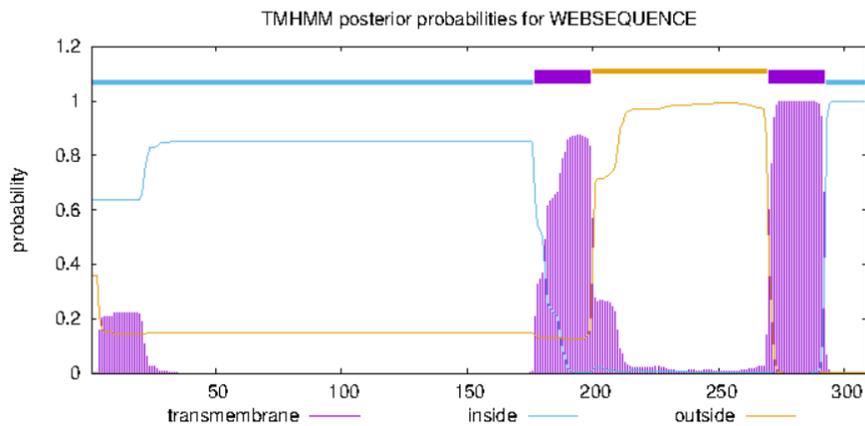
```
# WEBSEQUENCE Length: 300
# WEBSEQUENCE Number of predicted TMHs: 2
# WEBSEQUENCE Exp number of AAs in TMHs: 57.10805
# WEBSEQUENCE Exp number, first 60 AAs: 6.41147
# WEBSEQUENCE Total prob of N-in: 0.64427
WEBSEQUENCE TMHMM2.0 inside 1 178
WEBSEQUENCE TMHMM2.0 TMhelix 179 201
WEBSEQUENCE TMHMM2.0 outside 202 260
WEBSEQUENCE TMHMM2.0 TMhelix 261 283
WEBSEQUENCE TMHMM2.0 inside 284 300
```



PfGN01_100006100-t41_1

TMHMM result

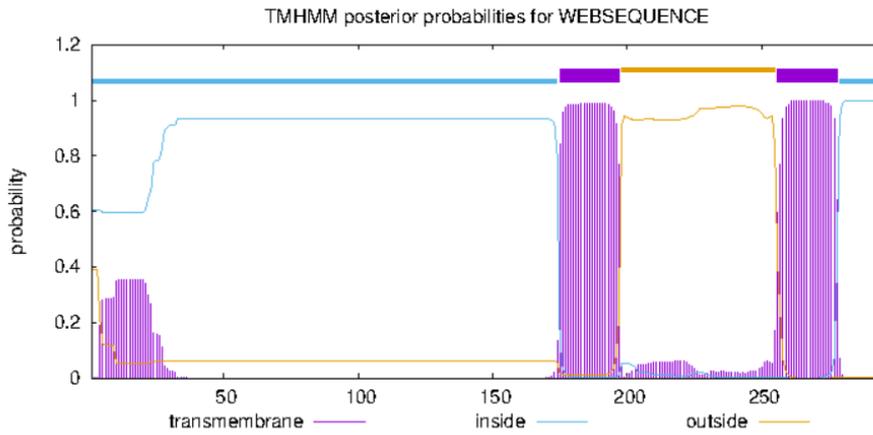
```
# WEBSEQUENCE Length: 309
# WEBSEQUENCE Number of predicted TMHs: 2
# WEBSEQUENCE Exp number of AAs in TMHs: 46.13525
# WEBSEQUENCE Exp number, first 60 AAs: 4.1073
# WEBSEQUENCE Total prob of N-in: 0.64013
WEBSEQUENCE TMHMM2.0 inside 1 176
WEBSEQUENCE TMHMM2.0 TMhelix 177 199
WEBSEQUENCE TMHMM2.0 outside 200 269
WEBSEQUENCE TMHMM2.0 TMhelix 270 292
WEBSEQUENCE TMHMM2.0 inside 293 309
```



PfKE01_100005800-t41_1

TMHMM result

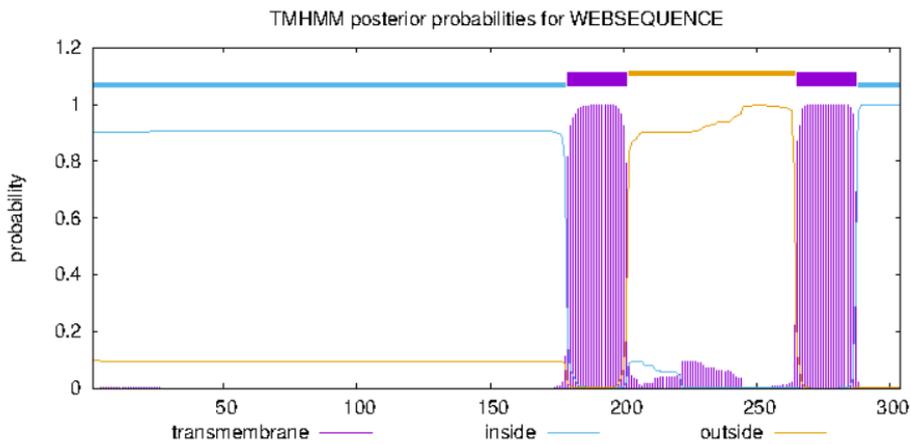
```
# WEBSEQUENCE Length: 295
# WEBSEQUENCE Number of predicted TMHs: 2
# WEBSEQUENCE Exp number of AAs in TMHs: 53.60472
# WEBSEQUENCE Exp number, first 60 AAs: 7.14763
# WEBSEQUENCE Total prob of N-in: 0.60721
WEBSEQUENCE TMHMM2.0 inside 1 174
WEBSEQUENCE TMHMM2.0 TMhelix 175 197
WEBSEQUENCE TMHMM2.0 outside 198 255
WEBSEQUENCE TMHMM2.0 TMhelix 256 278
WEBSEQUENCE TMHMM2.0 inside 279 295
```



PfSN01_030005600-t41_1

TMHMM result

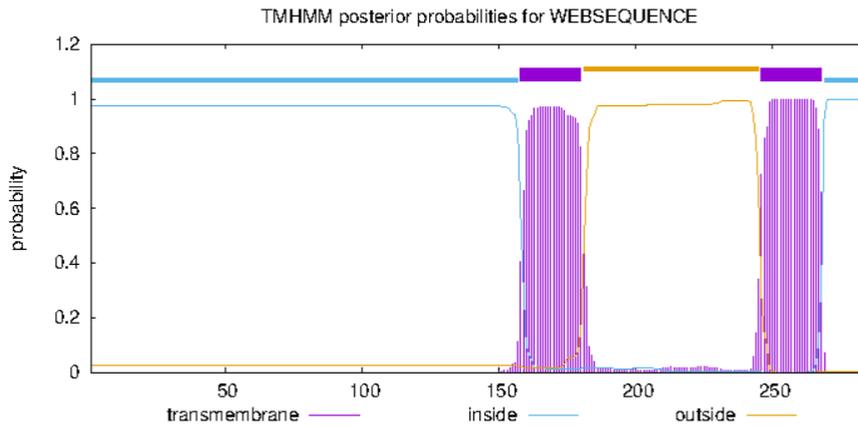
```
# WEBSEQUENCE Length: 304
# WEBSEQUENCE Number of predicted TMHs: 2
# WEBSEQUENCE Exp number of AAs in TMHs: 47.49211
# WEBSEQUENCE Exp number, first 60 AAs: 0.03045
# WEBSEQUENCE Total prob of N-in: 0.90325
WEBSEQUENCE TMHMM2.0 inside 1 178
WEBSEQUENCE TMHMM2.0 TMhelix 179 201
WEBSEQUENCE TMHMM2.0 outside 202 264
WEBSEQUENCE TMHMM2.0 TMhelix 265 287
WEBSEQUENCE TMHMM2.0 inside 288 304
```



PfHB3_040028900

TMHMM result

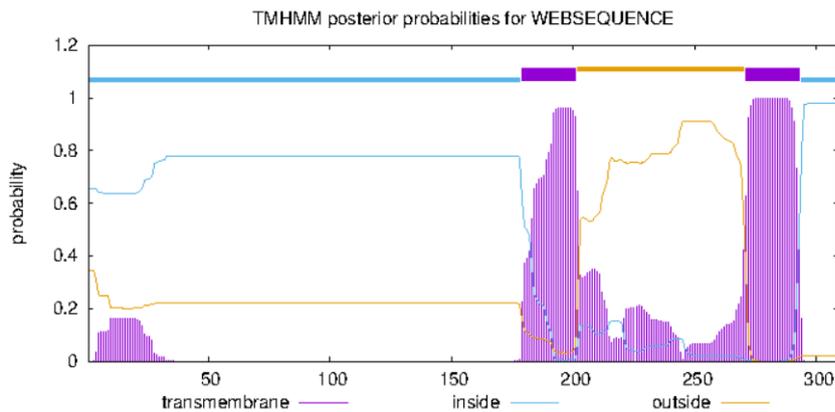
```
# WEBSEQUENCE Length: 285
# WEBSEQUENCE Number of predicted TMHs: 2
# WEBSEQUENCE Exp number of AAs in TMHs: 45.42818
# WEBSEQUENCE Exp number, first 60 AAs: 0
# WEBSEQUENCE Total prob of N-in: 0.97680
WEBSEQUENCE TMHMM2.0 inside 1 157
WEBSEQUENCE TMHMM2.0 TMhelix 158 180
WEBSEQUENCE TMHMM2.0 outside 181 245
WEBSEQUENCE TMHMM2.0 TMhelix 246 268
WEBSEQUENCE TMHMM2.0 inside 269 285
```



PfGA01_070034600-t41_1

TMHMM result

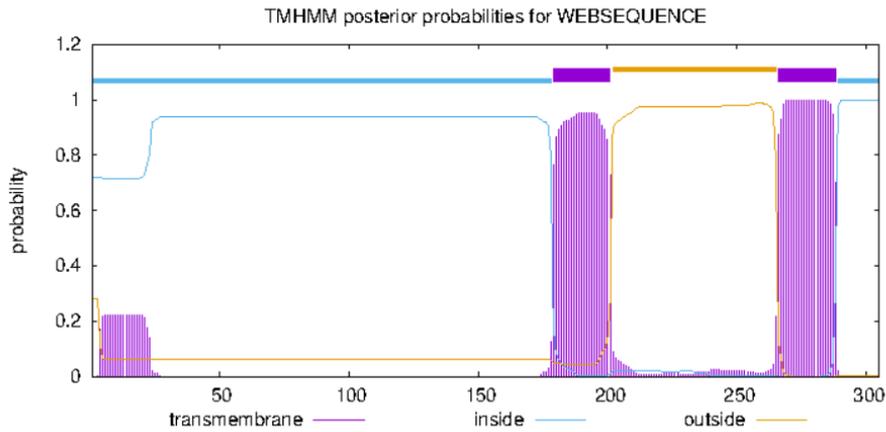
```
# WEBSEQUENCE Length: 309
# WEBSEQUENCE Number of predicted TMHs: 2
# WEBSEQUENCE Exp number of AAs in TMHs: 54.03043
# WEBSEQUENCE Exp number, first 60 AAs: 3.30002
# WEBSEQUENCE Total prob of N-in: 0.65561
WEBSEQUENCE TMHMM2.0 inside 1 178
WEBSEQUENCE TMHMM2.0 TMhelix 179 201
WEBSEQUENCE TMHMM2.0 outside 202 270
WEBSEQUENCE TMHMM2.0 TMhelix 271 293
WEBSEQUENCE TMHMM2.0 inside 294 309
```



PfGN01_020006800-t41_1

TMHMM result

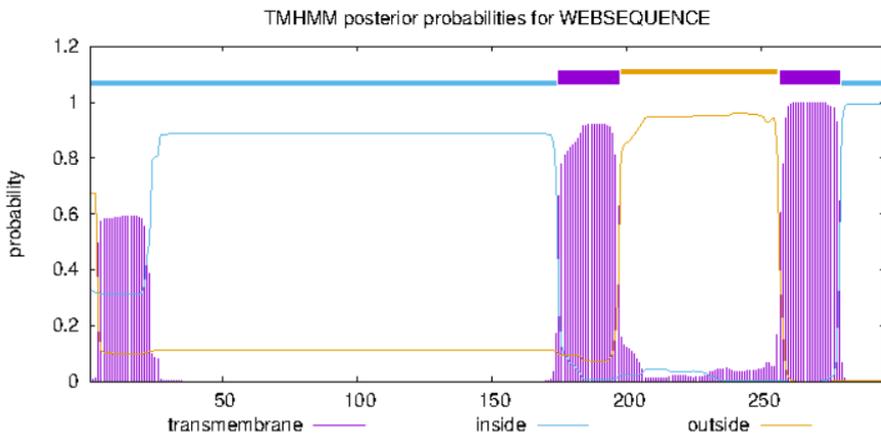
```
# WEBSEQUENCE Length: 305
# WEBSEQUENCE Number of predicted TMHs: 2
# WEBSEQUENCE Exp number of AAs in TMHs: 49.32162
# WEBSEQUENCE Exp number, first 60 AAs: 4.31078
# WEBSEQUENCE Total prob of N-in: 0.71886
WEBSEQUENCE TMHMM2.0 inside 1 178
WEBSEQUENCE TMHMM2.0 TMhelix 179 201
WEBSEQUENCE TMHMM2.0 outside 202 265
WEBSEQUENCE TMHMM2.0 TMhelix 266 288
WEBSEQUENCE TMHMM2.0 inside 289 305
```



PfSN01_000011500-t41_1

TMHMM result

```
# WEBSEQUENCE Length: 296
# WEBSEQUENCE Number of predicted TMHs: 2
# WEBSEQUENCE Exp number of AAs in TMHs: 56.03364
# WEBSEQUENCE Exp number, first 60 AAs: 11.57495
# WEBSEQUENCE Total prob of N-in: 0.32850
# WEBSEQUENCE POSSIBLE N-term signal sequence
WEBSEQUENCE TMHMM2.0 inside 1 174
WEBSEQUENCE TMHMM2.0 TMhelix 175 197
WEBSEQUENCE TMHMM2.0 outside 198 256
WEBSEQUENCE TMHMM2.0 TMhelix 257 279
WEBSEQUENCE TMHMM2.0 inside 280 296
```

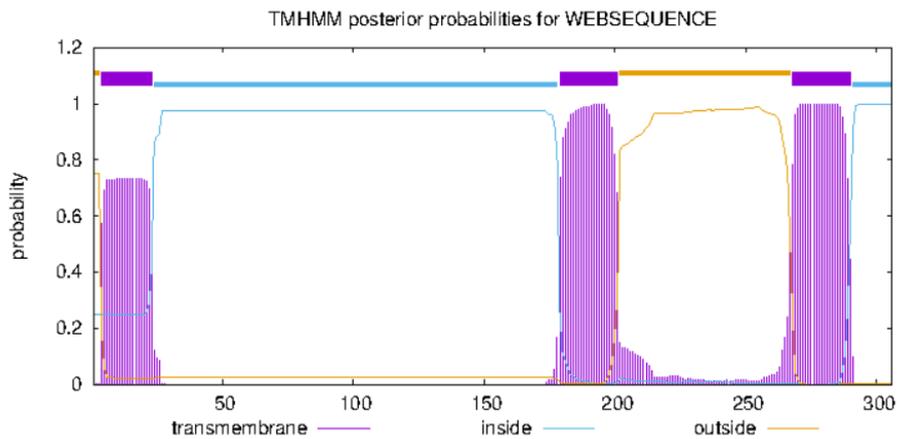


TMHMM result

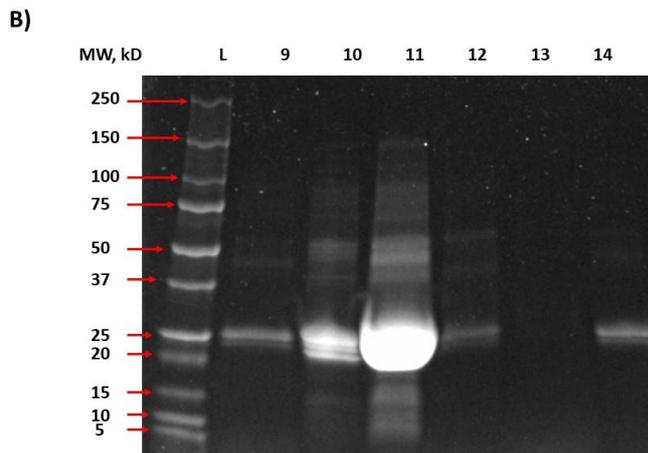
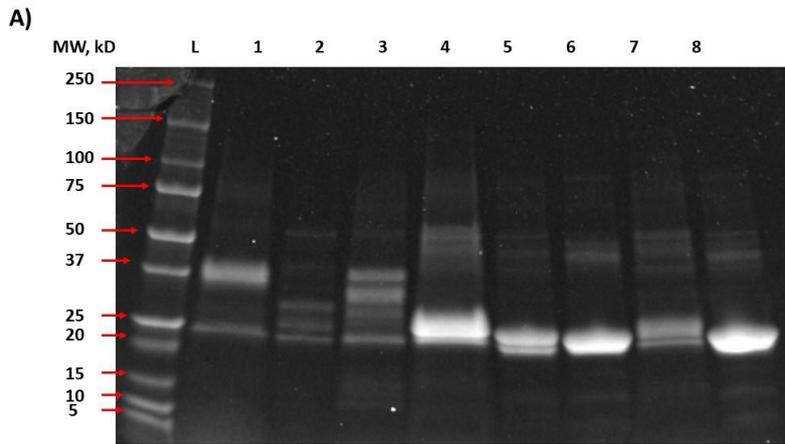
```

# WEBSEQUENCE Length: 306
# WEBSEQUENCE Number of predicted TMHs: 3
# WEBSEQUENCE Exp number of AAs in TMHs: 61.58025
# WEBSEQUENCE Exp number, first 60 AAs: 14.51679
# WEBSEQUENCE Total prob of N-in: 0.25040
# WEBSEQUENCE POSSIBLE N-term signal sequence
WEBSEQUENCE TMHMM2.0 outside 1 3
WEBSEQUENCE TMHMM2.0 TMhelix 4 23
WEBSEQUENCE TMHMM2.0 inside 24 178
WEBSEQUENCE TMHMM2.0 TMhelix 179 201
WEBSEQUENCE TMHMM2.0 outside 202 267
WEBSEQUENCE TMHMM2.0 TMhelix 268 290
WEBSEQUENCE TMHMM2.0 inside 291 306

```

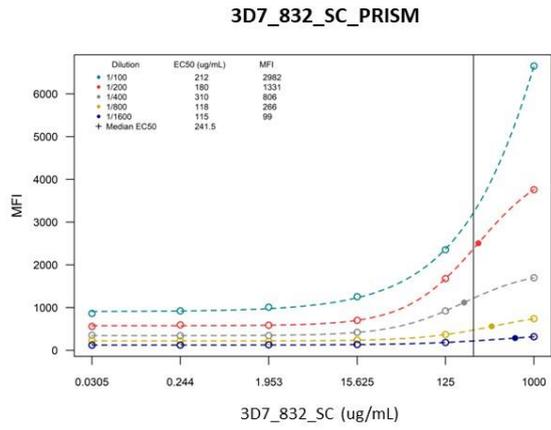
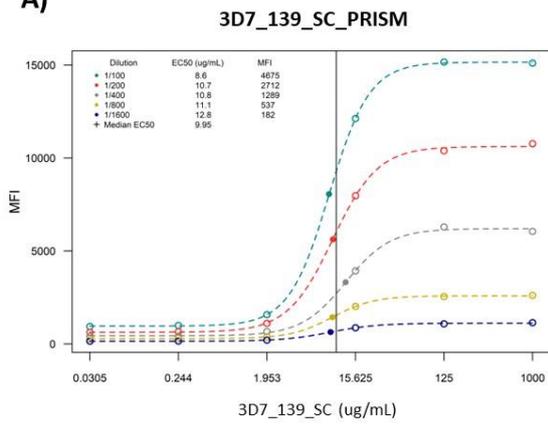


Supplementary Figure 1: Output of the TMHMM 2.0 transmembrane domain predicting software performed on the full-length amino acid sequences of each of the selected variants for the STEVOR V2 recombinant library. Transmembrane domains are predicted and displayed in purple, intercellular portions of the protein in blue and extracellular in orange. Signal peptides, marking the beginning of each protein sequence are also displayed in purple. All 11 graphics are displayed, and the identification name of each sequence is displayed above its representative graph.

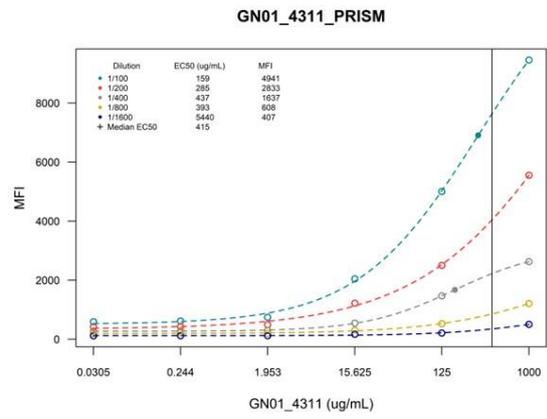
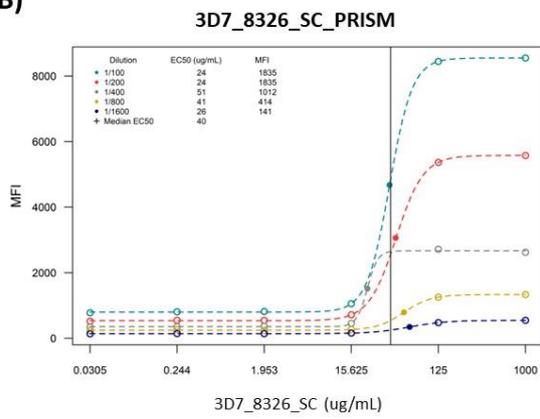


Supplementary Figure 2: Coomassie Brilliant Blue R-250 stain of 4-15% BioRad precast Mini-PROTEAN TGX SDS-PAGE gels of **A)** Molecular weight ladder (L), Pf3D7_1300900_SC (1), Pf3D7_0832000_SC (2), Pf3D7_0832600_SC (3), PfGN01_040031100_V2 (4), PfIT_130077600_V2 (5), PfSN01_140006300_V2 (6), PfGN01_100006100_V2 (7), PfKE01_10000580_V2 (8), and **B)** Molecular weight ladder (L), PfSN01_030005600_V2 (9), PfHB3_040028900_V2 (10), PfGA01_070034600_V2 (11), PfGN01_020006800_V2 (12), PfSN01_000011500_V2 (13), PfGN01_130006600_V2 (14).

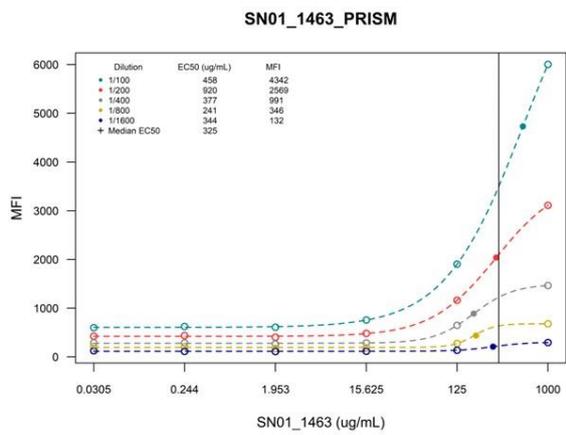
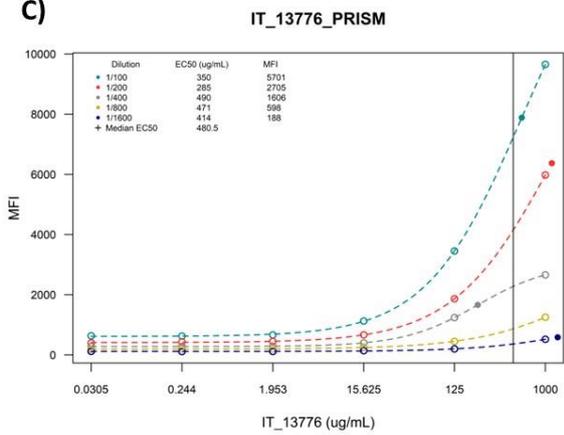
A)



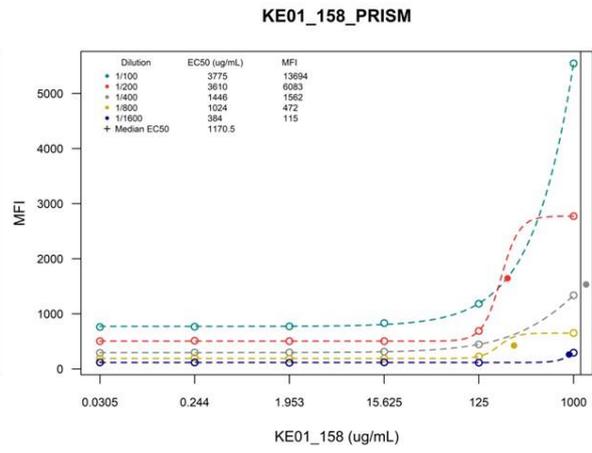
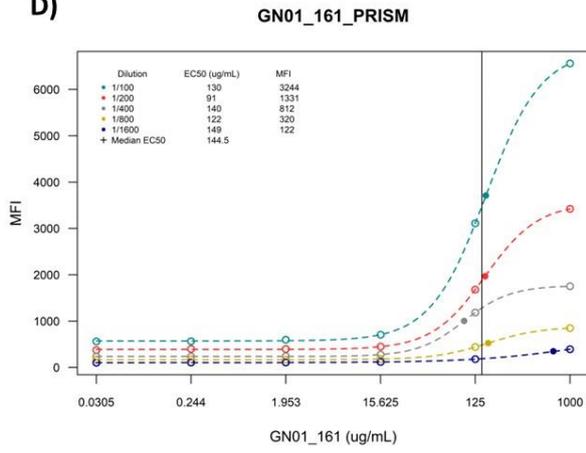
B)



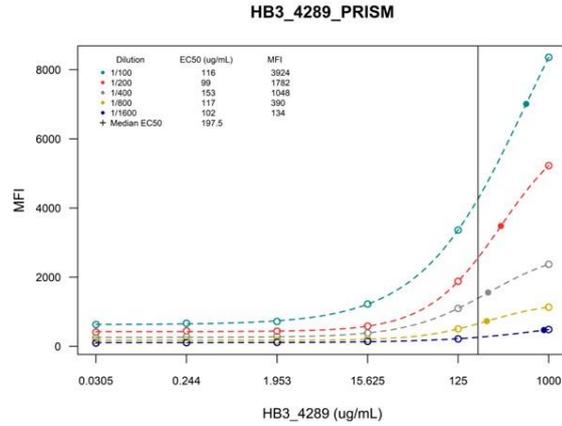
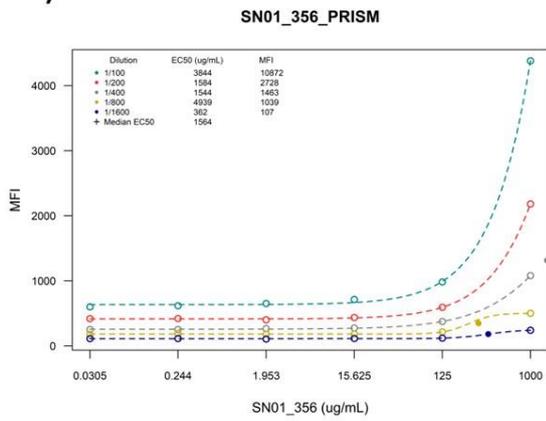
C)



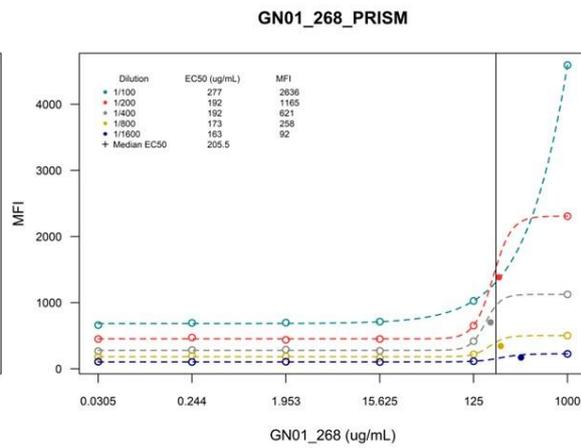
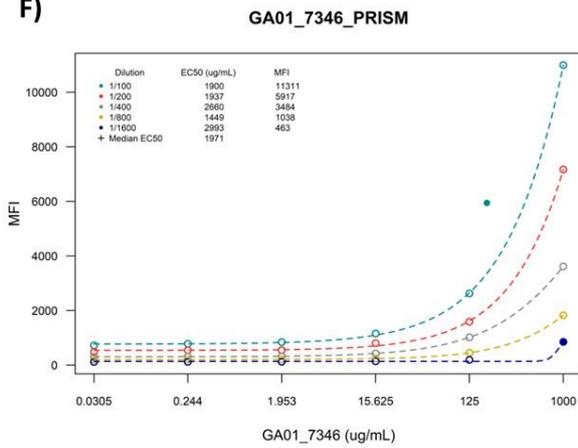
D)

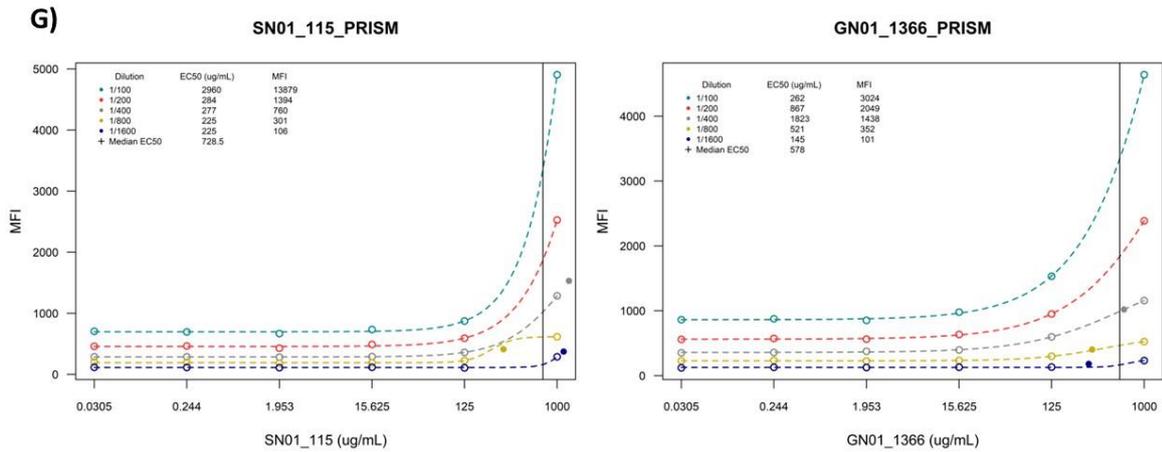


E)

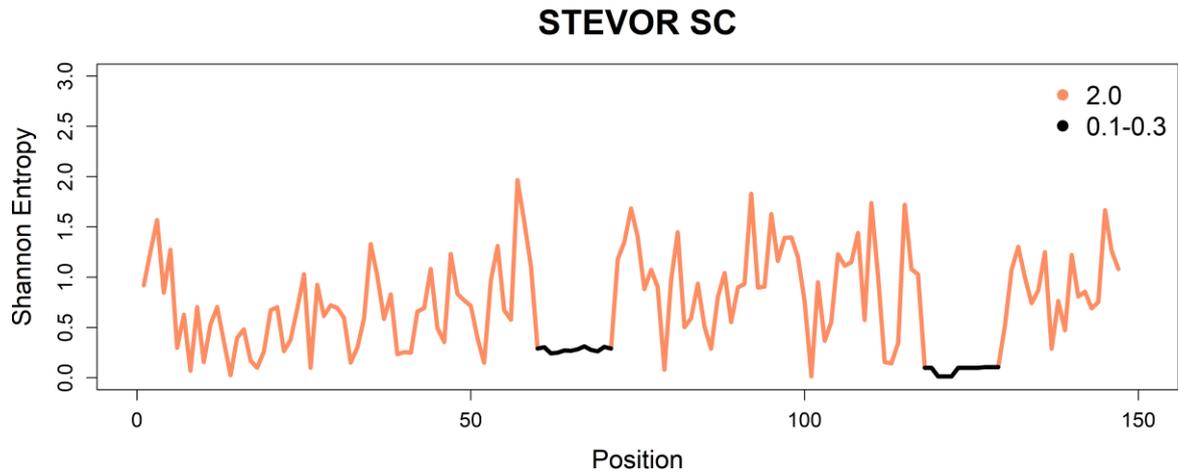


F)



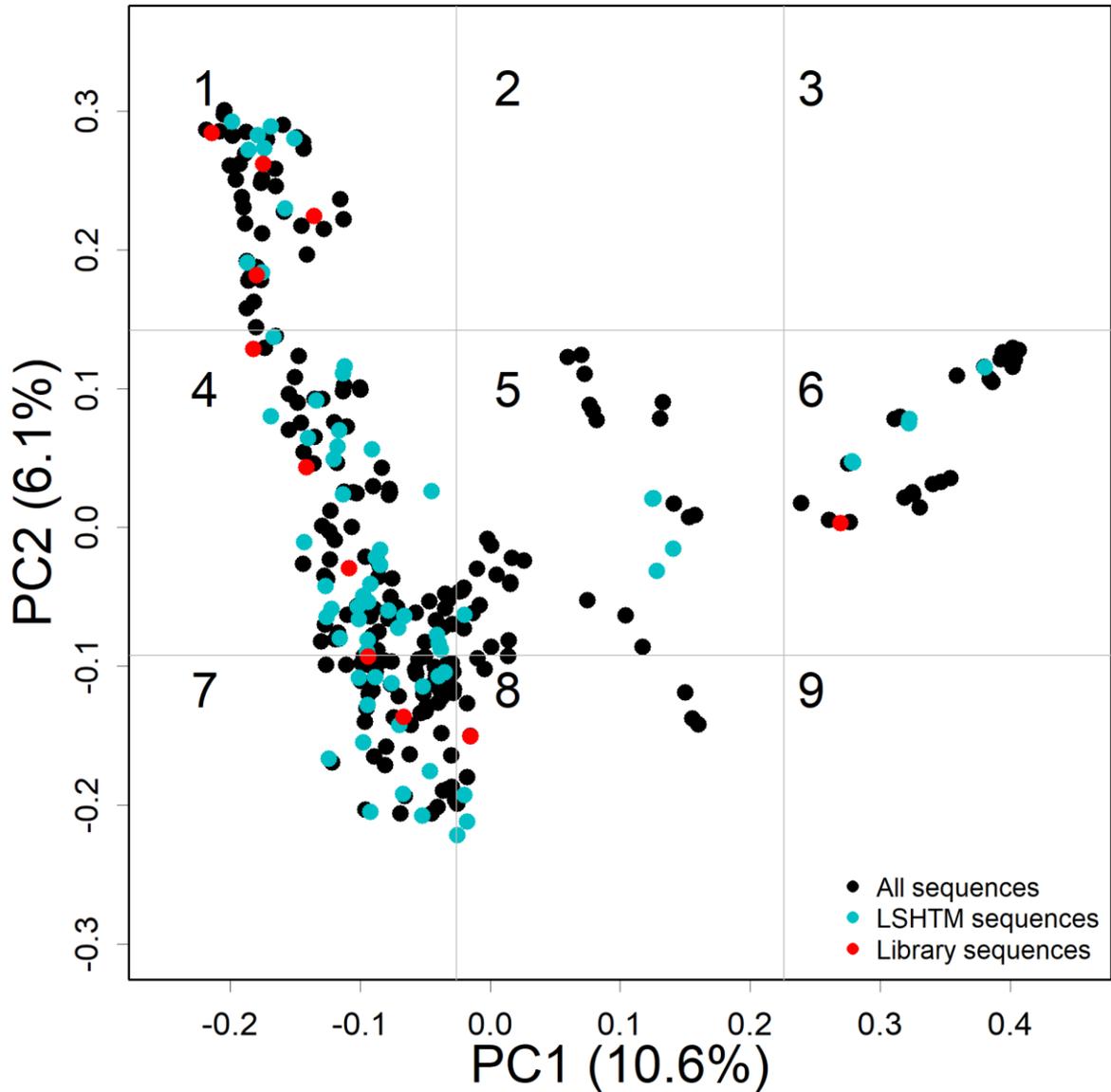


Supplementary Figure 3: Graphical summary of titration of all 14 recombinants at six-point eight-fold dilution from 1000 $\mu\text{g/ml}$ down to 0.0305 $\mu\text{g/ml}$ against five-point two-fold dilution series of the PRISM positive control serum pool from 1/100 down to 1/1600. The first three graphs, both from A) and first from B) represent the semi-conserved domain recombinants and the rest, down to G) represent the variable domain recombinants.

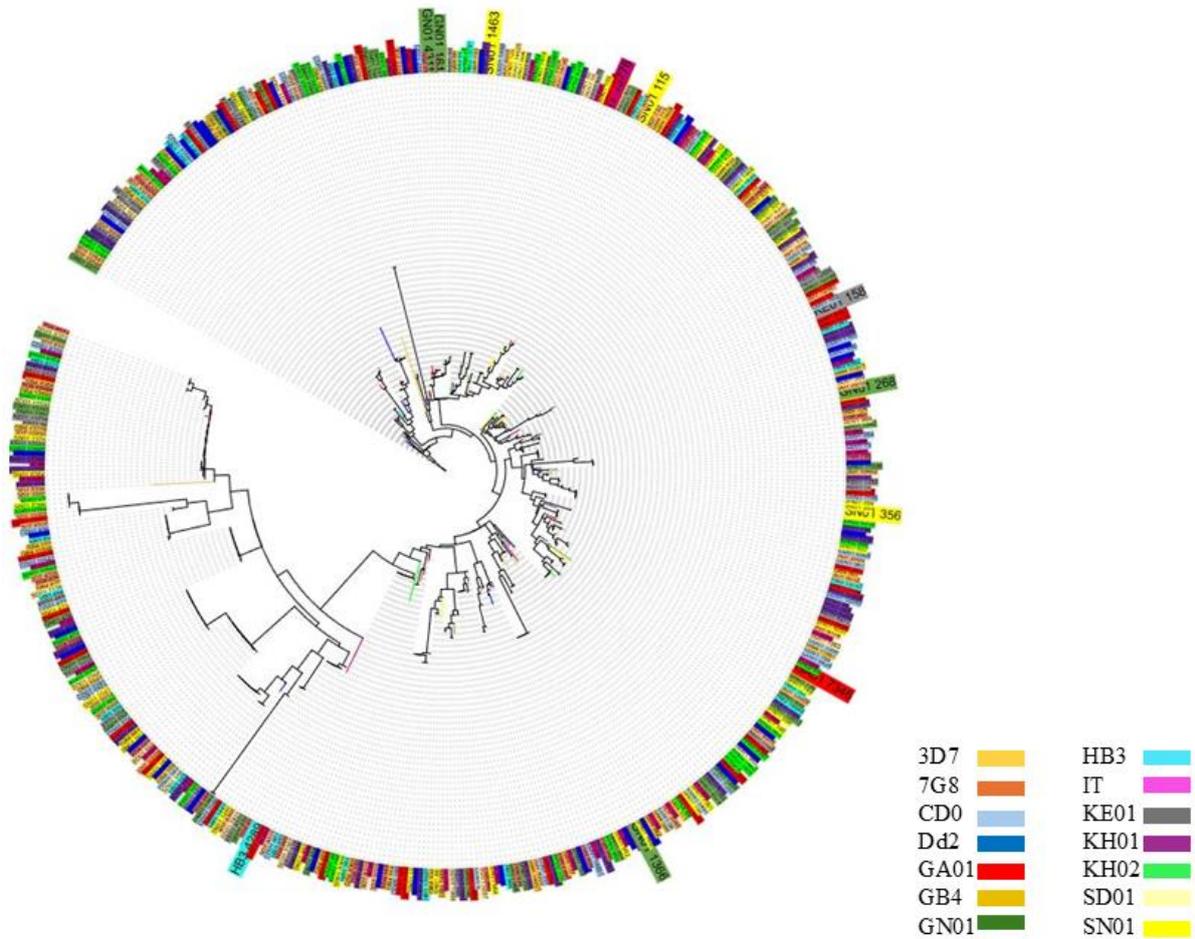


Supplementary Figure 4: Shannon Entropy Index plot representing the level of diversity in each amino acid position of the STEVOR semi-conserved domain alignment. The higher the value, the higher the diversity. The black portions represent the highly conserved regions of the domain. The x-axis represents the amino acid positions of the domain. Legend values represent the approximate entropy of the different parts of the SC domain.

STEVOR V2 Library



Supplementary Figure 5: Visualisation of the large variable domain (V2) sequences selected to be expressed as STEVOR recombinant protein library, shown in red, among the 80 available plasmids in LSHTM from all 493 V2 amino acid sequences used in the PCA analysis (shown in black). The numbers indicate the quadrant number (labelled 1-9) in which each of the library sequences is positioned, summarised in Table 2.



Supplementary Figure 6: A graphical representation of a phylogenetic tree of all 493 STEVOR V2 domain sequences generated on IQtree using JTT+F+G4 model and 1000 bootstrap replicates, optimised using hill-climbing nearest neighbour interchange (NNI) and visualised using iTOL: Tree of Life tool. Sequences names are colour coded according to the strain they belong to, indicated in the legend. Sequences names with larger font represent the 11 selected variants for the recombinant library.

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	lsh1601985	Title	Miss
First Name(s)	HRISTINA CHAVDAROVA		
Surname/Family Name	VASILEVA		
Thesis Title	BIOCHEMICAL AND BIOINFORMATIC CHARACTERISATION OF UNDERSTUDIED ERYTHROCYTE SURFACE EXPRESSED HYPERVARIABLE PROTEIN FAMILIES IN PLASMODIUM FALCIPARUM.		
Primary Supervisor	Dr Kevin Tetteh		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Frontiers of Parasitology
---	---------------------------

Please list the paper's authors in the intended authorship order:	Hristina Vasileva, Ernest Diez Benavente, Jack Bickford-Smith, Anna Last. Kevin K.A. Tetteh
Stage of publication	Submitted

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I have done all laboratory work, data analysis and manuscript preparation.
--	--

SECTION E

Student Signature	HRISTINA VASILEVA
Date	30.05.2024

Supervisor Signature	ANNA LAST
Date	06.06.2024

Chapter 6: Description of Study Sites for sample selection for Research Paper 3: “Breadth of endemic serum antibody responses to *Plasmodium falciparum* STEVOR recombinant protein antigen library”.

Linking material to Chapter 7

6.1 Introduction

To address the hypothesis presented in Research Paper 3 (Chapter 7) that higher malaria endemicity levels in a population characterised by increased exposure to *P. falciparum* will increase the breadth of responses to the STEVOR V2 library, described in Research Paper 2 (Chapter 5), in an age and exposure dependent manner, samples from three malaria intervention studies in Sub-Saharan Africa were selected (1). The aim was to choose samples representing individuals from a high malaria endemicity setting, such as the FIGHTMAL study in Northern Uganda (2010), as well as samples representing individuals from low malaria endemicity settings, such as the MASSIV and MATAMAL clinical trials conducted in The Gambia (2019) and the Bijagos islands of Guinea-Bissau (2021-2022), respectively (2),(3),(4). This chapter provides detailed description to the three studies from which residual serum and dried blood spots samples were sub-selected to be screened against the STEVOR V2 library alongside a panel of established *P. falciparum* markers of infection on the multiplex bead-based serological platform Luminex, presented in Research Paper 3 (5).

6.2 FIGHTMAL

The FIGHTMAL serum samples used in the study were collected as part of a longitudinal cohort study entitled “Correlating protection against malaria with serum profiles against *Plasmodium falciparum* antigen repertoires (FIGHTMAL) - epidemiological studies in Uganda”, conducted in 2010 in the Abedi parish, in Apac district, in Northern Uganda.

6.2.1 Study site and malaria endemicity.

The Abedi parish in Apac district in Northern Uganda is a rural area situated between Lake Kyoga and the Victoria Nile (latitude: 1°59'0''N and longitude: 32°32'0''E), covering an area of 3,255.9 square kilometres, indicated in Figure 1, with a population of over 282 thousand people (6). Apac district has a bimodal rainfall pattern with two short rainy seasons from April to May and from September to October, followed by a dry season from November to March. The district is regarded as holoendemic malaria area, characterised by intense malaria transmission with overall *Plasmodium falciparum* parasite prevalence by microscopy of 37.5% and by rRNA qPCR of 57.7% in 2009, both decreasing with increasing age. The highest detected prevalence was found in children under five years of age of 55.8% by microscopy and 71.9% by qPCR (7).



Figure 1: Figure adapted from the official site of Apac District, Government of Uganda. The area in red indicates the borders of the Apac district within Uganda in yellow (6).

6.2.2 Study design and population.

The longitudinal cohort consisted of a total number of 509 participants from 300 recruited households, stratified in three age groups: under 5 years of age, individuals with neither clinical nor parasitic immunity (n=249), due to their immature immune systems and little exposure to malaria infection; individuals between 6 and 10 years of age, individuals with clinical but no parasitological immunity (n=126), characterised by high exposure to the infection due to their behaviour nature, and thus suggested clinical immunity; and individuals over the age of 20, assumed to possess both clinical and parasitological immunity (n=134). Clinical immunity to malaria is defined as protection against symptomatic disease achieved through repeated exposure, resulting in asymptomatic infection. This immunity included anti-parasitic responses

(anti-parasitic immunity) that control parasitaemia, maintaining parasites as a low density. Acquired immunity encompasses both the prevention of symptoms and the regulation of parasite density, thereby reducing the risk of severe malaria or death (8). summarised in the schematic in Figure 2 (9),(10),(11). The exclusion criteria were Z-score lower than 3 (weight-for-height/height-for-age), severe anaemia, or the presence of any chronic disease. All participants were clinically assessed for detection of malaria (fever) and received an antimalarial treatment with artemether/lumefantrine (ACT) at a standard dose (six doses of 20mg/120mg) at enrolment, following baseline sample collection. Three timepoint sample collections were conducted, during which a finger prick blood sample was collected for microscopy thin and thick smear, qPCR, RDT and haemoglobin levels to check for study inclusion. Additionally, a dried blood spot (DBS) sample on filter paper and matching serum samples were collected. Sample collection was performed at baseline prior to giving treatment, post-parasite clearance (after 6 weeks), and post-peak transmission season (16 weeks after baseline). Each participant was also given a long-lasting insecticidal net (LLIN) and was followed for up to six months using passive case detection, receiving treatment for malaria if they experienced a fever above 37.5°C with any parasite density (2). Written consent form was obtained from each participant or guardian, and an assent form was obtained from adolescents aged 12 and 17 years.

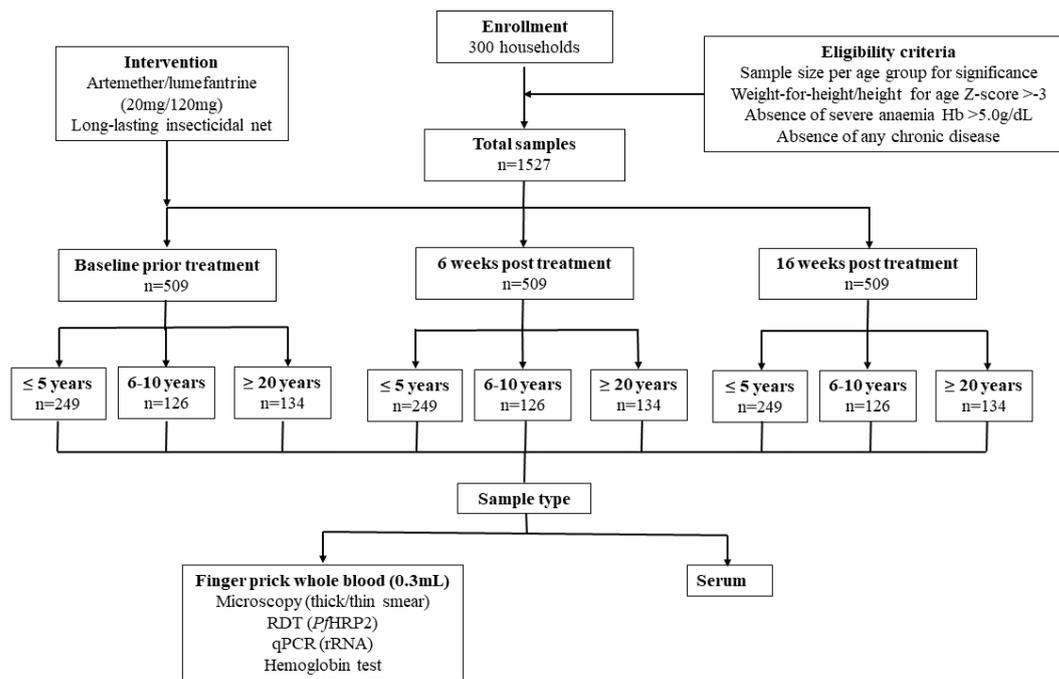


Figure 2: FIGHTMAL longitudinal cohort study profile with participants per time point of sampling, intervention type, study eligibility criteria and type of collected samples.

6.3 MASSIV

The MASSIV dried blood spot samples were collected during the clinical trial “Mass drug administration of Ivermectin and Dihydroartemisinin-piperaquine as an additional intervention for malaria elimination (MASSIV)” cluster-randomised trial (ClinicalTrials.gov ID: NCT03576313) conducted in 2018 and 2019 in the Upper-River Region in The Gambia.

6.3.1 Study site and malaria endemicity.

The MASSIV trial was conducted in the Upper River Region (URR) of the eastern part of The Gambia (latitude: 13°23'40''N and longitude: 14°10'31''W). This area spans two thousand square kilometres and has a population of 240 thousand people, illustrated in Figure 3. The region experiences a single rainy season a year, between June and October, followed by a long dry season from October to June. Malaria transmission is seasonal with highest transmission

rates recorded immediately after the rainy season (October-November). The baseline malaria prevalence of clusters selected for the trial was approximately 15% in November, determined via *varATS* qPCR (12).

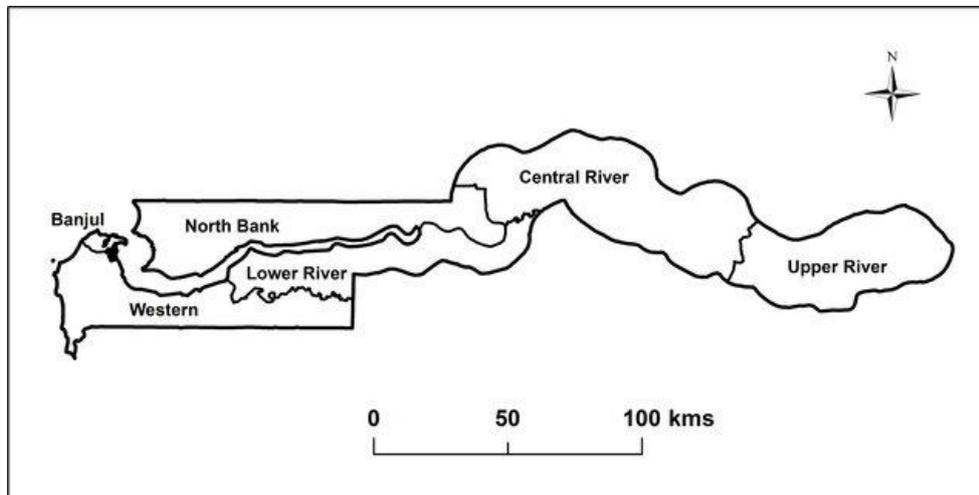


Figure 3: Map of the demographic division of The Gambia, MASSIV was conducted in the Upper River region. Figure is adapted from Robello M. *et. al* (2015), *PLOS NTD* (13).

6.3.2 Study design and population.

MASSIV is a two-arm cluster-randomised mass drug administration (MDA) trial. The trial was conducted in 32 clusters (villages), randomized in 1:1 ratio (16 clusters in each arm). The intervention arm received dihydroartemisinin-piperaquine (DP) according to body weight per manufacturer’s guidelines and ivermectin (IVM) at 300-400 $\mu\text{g}/\text{kg}/\text{day}$ both once daily for three days across three subsequent months during the transmission season starting from July in both 2018 and 2019. There was no placebo control arm; the control arm did not receive any MDA but did receive seasonal malaria chemoprophylaxis (SMC). Both arms received standard malaria control interventions, including insecticide treated nets (ITNs), indoor residual spraying (IRS) in 2019, as well as SMC and intermittent preventive treatment during pregnancy (IPTp). SMC was not administered to eligible children under 5 years of age in the intervention arm to avoid double antimalarial treatment. The participant exclusion criteria included children

below the age of 6 months (DP) and below a weight of ≤ 15 kg (IVM); individuals with chronic illnesses, those who were pregnant (IVM: any trimester; DP: first trimester), breastfeeding mothers (IVM), people who had travelled to *Loa loa* endemic countries (IVM) and hypersensitivity to both treatments.

DBS samples from 200 individuals per cluster were collected in June 2019, nine months after the first intervention, and in November 2019, one month after the second MDA intervention. This resulted in a total of 3,103 samples in June (1,470 in the intervention and 1,540 in the control arm) and 2,975 samples in November (1,479 in the intervention and 1,496 in the control arm), as depicted in the schematic in Figure 4. All samples were analysed for *P. falciparum* infection using *varATS* qPCR and were analysed for the presence of antigen-specific antibodies against an established panel of *P. falciparum* markers of infection using Luminex MagPlex© technology (3), (Kozits C. *et. al*, unpublished data). Written informed consent was obtained from all adults; for children consent was provided by their guardians, and assent was obtained for adolescents aged 12-17 years.

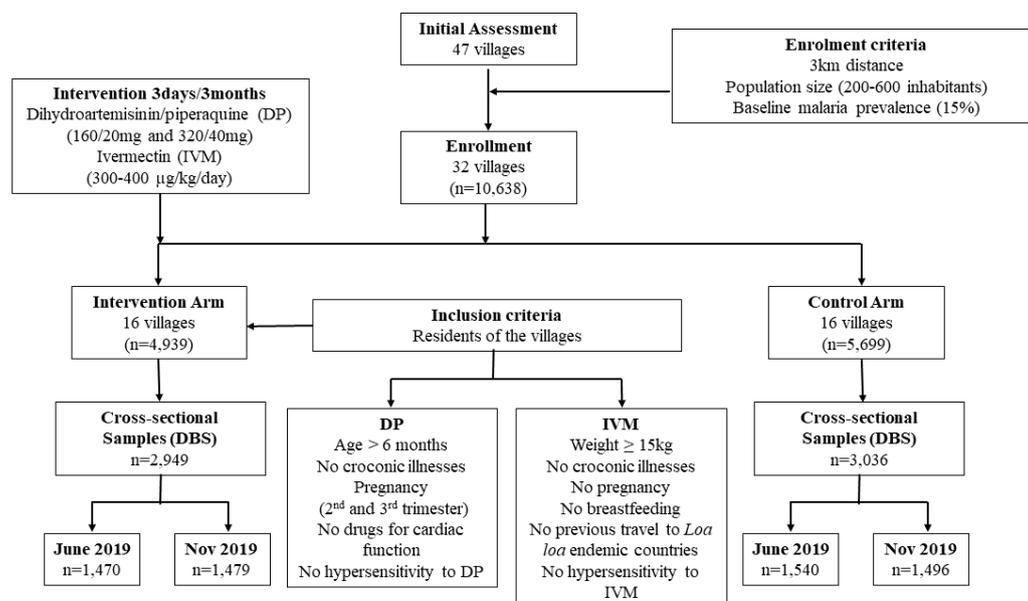


Figure 4: MASSIV clinical trial study profile with participants per trial arm, intervention type, study inclusion criteria, and collected samples.

6.4 MATAMAL

The MATAMAL dried blood spots samples come from the quadruple-blinded cluster randomised placebo-controlled trial “Adjunctive Ivermectin Mass Drug Administration of Malaria Control on the Bijagos Archipelago of Guinea-Bissau” (ClinicalTrials.gov ID: NCT04844905) conducted in 2021 and 2022 in the Bijagos Archipelago of Guinea-Bissau.

6.4.1 Study site and malaria endemicity.

The Bijagos archipelago consists of 19 permanently inhabited islands approximately 50 km off the coast of Guinea-Bissau, with a population of approximately 25 thousand people. Similar to The Gambia, the archipelago is characterised with a long dry season and a short rainy season from June to October (14). Malaria transmission is highly seasonal, with the highest 18S qPCR *P. falciparum* prevalence recorded of 17.5% in November and lowest of 8.5% recorded in the month of January in 2018. An exception is the island of Soga, which presented a qPCR prevalence as high as 40%; however, this island was not included in the MATAMAL trial (Hutchins H *et al*, unpublished data). MATAMAL was conducted on 18 out of the 19 inhabited islands of the Bijagos Archipelago, as indicated on the map in Figure 5. Highest *P. falciparum* prevalence was recorded in children between 6 and 15 years of age.



Figure 5: Map of the Bijagos Archipelago of Guinea-Bissau with MATAMAL clusters marked as stars and landmarks of islands separated into different clusters marked as lines. Figure is adapted from Hutchins H. *et al.* (2023), *BMJ Open* (4).

6.4.2 Study design and population.

Individual islands were selected as individual clusters, but three large islands (Bubaque, Canhabaque and Uno) were sub-divided into three individual clusters each, separated by at least 2km as a buffer zone. This resulted in total of 24 clusters, randomized in 1:1 ratio (12 clusters in each arm). Island Soga was not included in the trial but was treated as if it was in the control arm, serving as a buffer zone due to its very high baseline malaria prevalence, which made it an outlier and could have led to imbalance between the arms. All islands received ITNs through the national malaria control program, with a coverage of 92%. IPTp was also administered, although IRS is not deployed in Guinea Bisau. In the MATAMAL trial, the intervention arm received dihydroartemisinin-piperaquine (DP) according to body weight per manufacturer’s guidelines and ivermectin (IVM) at 300 µg/kg/day, both administered once daily for three subsequent days across three subsequent months during the transmission season

starting from July for both 2021 and 2022. Participants in the control arm received DP according to body weight per manufacturer's guidance and placebo IVM, which was undistinguishable from the intervention IVM. The exclusion criteria of the intervention included severe illness; children under 6 months (for DP); weight under 15kg (IVM/placebo); pregnancy (any trimester for IVM/placebo and first trimester for DP); breastfeeding (for IVM/placebo); and travel to *Loa loa* endemic countries (IVM/placebo); usage of drugs for cardiac function (DP), and hypersensitivity to either or both treatments.

DBS samples were collected from 200 participants from each cluster, one month after the final MDA intervention in each of both years of the trial, giving a total of 4,441 samples in 2021 (2,277 in the control and 2,164 in the intervention arm) and 4,383 samples in 2022 (2,300 in the control and 2,083 in the intervention arm), as summarised in the schematic in Figure 6. A cohort of 50 children aged 5 to 14 years were followed monthly throughout the malaria transmission season in 18 clusters each year. At each visit, a DBS sample was collected (4). All samples were analysed for *P. falciparum* infection using *varATS* qPCR and for the presence of antigen-specific antibodies against an established panel of *P. falciparum* markers of infection using Luminex MagPlex© technology. Written informed consent was obtained for all participants or their guardians, and assent form was obtained from adolescents aged 12 to 17 years.

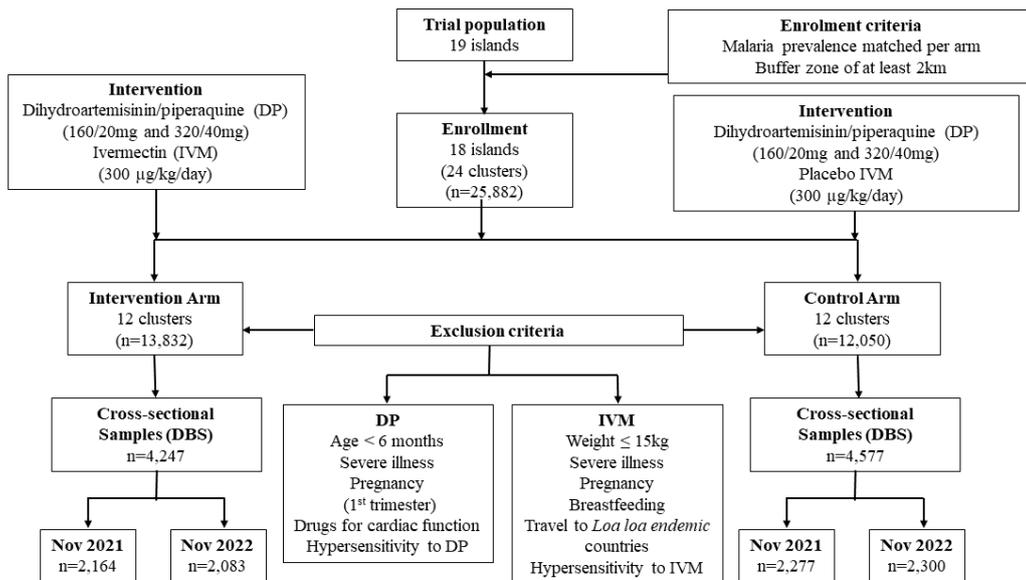


Figure 6: MATAMAL clinical trial study profile with participants per trial arm, intervention type, study exclusion criteria and collected samples.

6.5 Discussion

All available serum samples from the FIGHTMAL study were selected and a sub-selection of a total of 1,278 DBS from MASSIV and MATAMAL was performed to be tested against the newly developed library of STEVOR hypervariable recombinant antigens to explore the breadth of responses in populations from settings characterised by different *P. falciparum* endemicity levels. The sub-selection of the DBS samples was based on all malaria positive samples by *varATS P. falciparum* qPCR one month after the second year of intervention for MASSIV in November 2019 and two months after intervention for both years for MATAMAL, November 2021 and November 2022, and the same sample number of qPCR malaria negative DBS matched by timepoint, age-group and arm, per trial.

The aim was to use the FIGHTMAL samples to represent an area with high endemicity and MASSIV/MATAMAL samples to represent areas with low endemicity. Although the reported baseline prevalences for both The Gambia and the Bijagos Archipelago suggest moderate endemicity, endpoint samples from MASSIV and samples from both cross-sectional sample points for MATAMAL show qPCR prevalence in the control arms of 5.1%, 1.75%, and 6.64%, respectively, indicating areas with low *P. falciparum* endemicity (15),(Hutchins H *et. al*, unpublished data). Furthermore, samples from all three studies come from individuals of all ages, including children under the age of 5 years, making them good sample sets for the research question, as breadth of responses against *P. falciparum* antigens is associated with age in addition to exposure (9),(16).

6.6 References

1. Vasileva H, Benavente ED, Last A. In-silico classification and antigen library expression of *Plasmodium falciparum* STEVOR hypervariable infected erythrocyte surface-expressed multivariant protein family. 2023 Nov 14 [cited 2024 Mar 19]; Available from: <https://www.researchsquare.com>
2. Proietti C, Verra F, Bretscher MT, Stone W, Kanoi BN, Balikagala B, et al. Influence of infection on malaria-specific antibody dynamics in a cohort exposed to intense malaria transmission in northern Uganda. *Parasite Immunol* [Internet]. 2013 May 1 [cited 2024 Mar 8];35(5–6):164–73. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1111/pim.12031>
3. Debira E, et.al. Mass drug administration with high-dose ivermectin and dihydroartemisinin-piperaquine for malaria elimination in an area of low transmission with high coverage of malaria control interventions (MASSIV): a study protocol for a cluster-randomized clinical t. *JMIR Res Protoc*. 2020;
4. Hutchins H, Bradley J, Pretorius E, Teixeira Da Silva E, Vasileva H, Jones RT, et al. Protocol: Protocol for a cluster randomised placebo-controlled trial of adjunctive ivermectin mass drug administration for malaria control on the Bijagós Archipelago of Guinea-Bissau: the MATAMAL trial. *BMJ Open* [Internet]. 2023 Jul 7 [cited 2023 Sep 27];13(7):72347. Available from: [/pmc/articles/PMC10335573/](https://pmc/articles/PMC10335573/)
5. Wu L, Hall T, Ssewanyana I, Oulton T, Patterson C, Vasileva H, et al. Optimisation and standardisation of a multiplex immunoassay of diverse *Plasmodium falciparum* antigens to assess changes in malaria transmission using sero-epidemiology. *Wellcome Open Res* [Internet]. 2020 Apr 23 [cited 2020 Oct 15];4:26. Available from: <https://doi.org/10.12688/wellcomeopenres.14950.1>

6. APAC DISTRICT | LOCAL GOVERNMENT [Internet]. [cited 2024 Mar 8]. Available from: <https://www.apac.go.ug/>
7. Proietti C, Pettinato DD, Kanoi BN, Ntege E, Crisanti A, Riley EM, et al. Continuing Intense Malaria Transmission in Northern Uganda. *Am J Trop Med Hyg* [Internet]. 2011 May 5 [cited 2024 Mar 8];84(5):830. Available from: </pmc/articles/PMC3083756/>
8. Doolan DL, Dobaño C, Baird JK. Acquired immunity to Malaria [Internet]. Vol. 22, *Clinical Microbiology Reviews*. American Society for Microbiology (ASM); 2009 [cited 2020 Oct 20]. p. 13–36. Available from: </pmc/articles/PMC2620631/?report=abstract>
9. Aponte JJ, Menendez C, Schellenberg D, Kahigwa E, Mshinda H, Vountasou P, et al. Age interactions in the development of naturally acquired immunity to *Plasmodium falciparum* and its clinical presentation. *PLoS Med* [Internet]. 2007 [cited 2024 Apr 24];4(7):1259–67. Available from: <https://pubmed.ncbi.nlm.nih.gov/17676985/>
10. Ranjha R, Singh K, Baharia RK, Mohan M, Anvikar AR, Bharti PK. Age-specific malaria vulnerability and transmission reservoir among children. *Glob Pediatr*. 2023 Dec 1;6:100085.
11. Guinovart C, Dobaño C, Bassat Q, Nhabomba A, Quintó L, Manaca MN, et al. The role of age and exposure to *Plasmodium falciparum* in the rate of acquisition of naturally acquired immunity: a randomized controlled trial. *PLoS One* [Internet]. 2012 Mar 7 [cited 2024 Apr 24];7(3). Available from: <https://pubmed.ncbi.nlm.nih.gov/22412865/>
12. Mwesigwa J, Achan J, Di Tanna GL, Affara M, Jawara M, Worwui A, et al. Residual malaria transmission dynamics varies across The Gambia despite high coverage of

- control interventions. *PLoS One* [Internet]. 2017 Nov 1 [cited 2024 Mar 8];12(11):e0187059. Available from:
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0187059>
13. Rebollo MP, Sambou SM, Thomas B, Biritwum NK, Jaye MC, Kelly-Hope L, et al. Elimination of Lymphatic Filariasis in The Gambia. *PLoS Negl Trop Dis* [Internet]. 2015 Mar 18 [cited 2024 Mar 18];9(3):e0003642. Available from:
<https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0003642>
 14. Durrans S, Last A, Boiro H, Goncalves A, Mabey D, Greenland K. “Moving like birds”: A qualitative study of population mobility and health implications in the Bijagós Islands, Guinea Bissau. *Soc Sci Med* [Internet]. 2019 Jun 1 [cited 2019 May 20];230:204–13. Available from:
<https://www.sciencedirect.com/science/article/pii/S0277953619301613>
 15. Dabira ED, Soumare HM, Conteh B, Ceesay F, Ndiath MO, Bradley J, et al. Mass drug administration of ivermectin and dihydroartemisinin-piperaquine against malaria in settings with high coverage of standard control interventions: a cluster-randomised controlled trial in The Gambia. *Lancet Infect Dis* [Internet]. 2022 Apr 1 [cited 2024 Apr 24];22(4):519–28. Available from: <https://pubmed.ncbi.nlm.nih.gov/34919831/>
 16. Wu L, Mwesigwa J, Affara M, Bah M, Correa S, Hall T, et al. Antibody responses to a suite of novel serological markers for malaria surveillance demonstrate strong correlation with clinical and parasitological infection across seasons and transmission settings in The Gambia. *BMC Med* [Internet]. 2020 Sep 25 [cited 2023 Mar 13];18(1):304. Available from: </pmc/articles/PMC7517687/>

Chapter 7: Research Paper 3

Breadth of endemic serum antibody responses to *Plasmodium falciparum* STEVOR recombinant protein antigens library

Hristina Vasileva^{1,2}, Christian Kositz², Helena Brazal-Monzo¹, Harry Hutchins², Eunice Teixeira da Silva^{3,4}, Edgard Dabira⁵, Mamadou Ousmane Ndiath⁵, Chris Drakeley¹, Umberto D'Alessandro⁵, Anna Last², Kevin K.A. Tetteh^{1,6}

¹Department of Infectious Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

²Clinical Research Department, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

³Projecto de Saúde Bandim, Bissau, Guinea-Bissau

⁴Ministério de Saúde Pública, Bissau, Guinea-Bissau

⁵Diseases Control and Elimination, Medical Research Council Unit the Gambia at the London School of Hygiene and Tropical Medicine, Fajara, The Gambia

⁶Medical Affairs Department, FIND, Geneva, Switzerland

Abstract

The STEVOR protein family, part of the *Plasmodium falciparum* variable surface antigens contributes to immune evasion and infected erythrocyte sequestration. Evidence suggests that immunity against malaria involves a broad and geographically specific antibody responses to different *PfEMP1* variants, which play a crucial role in both uncomplicated and severe malaria outcomes. Despite limited research on the STEVOR protein family, there is evidence showing that antibody responses to variants correlate with age and confer protective immunity.

A library of 11 purified STEVOR hypervariable domain recombinants from non-reference strains was used on the Luminex platform, to assess antibody responses across varying ages and malaria endemicity. The aim was to explore the breadth of antibody responses to the STEVOR library in settings with contrasting levels of malaria endemicity and to identify STEVOR variants associated with early immune response and potential protection against infection, while also comparing them to antibody responses to established serological markers of seroincidence.

This study does not provide evidence that high malaria endemicity, characterised by an increased exposure to *P. falciparum*, enhances the breadth of antibody responses to the STEVOR hypervariable domain panel. Furthermore, the antibody reactivity levels to the STEVOR recombinants did not correlate with level of endemicity, nor with clinical or infectious status in individuals with uncomplicated or asymptomatic malaria. Additionally, three recombinants were consistently serodominant across different levels of endemicity, and three were minimally recognised in low endemicity settings, results attributed to differences in host immunity and random effect. A new hypothesis that breadth of responses to STEVOR variants is not driven by the setting's *P. falciparum* endemicity level is made, and the high breadth of responses even in short exposure in children is attributed to a mechanism of parasite survival and level of cross-reactivity of antibodies to the STEVOR variants.

Further research involving individuals with varying malaria clinical outcomes is needed to clarify the significance of STEVOR hypervariability in disease pathology.

7.1 Introduction

World Health Organisation (WHO) approved *Plasmodium falciparum* vaccines contain the circumsporozoite *P. falciparum* protein (CSP) as the vaccine target to stimulate antibody response in the recipients and potential protection of infection (1),(2). However, due to the complexity of the parasite life cycle and its antigenic variation, the approach of “one antigen one vaccine” may not be the most appropriate in achieving long-lasting protection (3),(4).

For instance, the RTS,S vaccine faces efficacy challenges, as it is based on the 3D7 reference strain and on a single CSP protein type, which was found to be not well represented in some targeted populations such as Ghana, due to high polyclonality of the infection as well as genetic mutations of the targeted antigen. As a result, some parasites evade immune responses through antigenic variation, progressing to the blood stage of the infection and bypassing vaccine-induced immunity (5). Given the parasite's complex lifecycle and genetic variability, multi-epitope or multi-antigen approaches could potentially address these challenges, offering broader and more robust protection against *Plasmodium falciparum*.

Other vaccine candidates, such as the *P. falciparum* Apical Membrane Antigen 1 (*PfAMA1*) and the *P. falciparum* Merozoite Surface Expressed Protein 1 (*PfMSP1.19*) which possess attributes such as exposure to the host circulation, antigenic properties and association with diseases pathology, have been extensively investigated and have progressed to malaria vaccine phase trials (6),(7). Both vaccine candidates have been explored in phase 1 clinical trials, demonstrating similar safety profiles with transient local and systemic adverse events (6), (7). For AMA1, dose-dependent antibody responses were functional, exhibiting partial growth inhibition *in vitro*. Moreover, antibody levels declined significantly within one year post-vaccination, and challenges such as antigenic polymorphism in AMA1 could hinder broader efficacy (6). For MSP1, all vaccinees seroconverted, producing long-lasting IgG and IgM titres, with immune responses exceeding those observed in naturally exposed individuals. Functional

assays showed that MSP1-specific antibodies activated neutrophils and induced complement fixation but failed to directly inhibit merozoite invasion or growth *in vitro*. Despite robust humoral and memory T-cell responses, the lack of direct inhibitory activity raised questions about the vaccine's clinical efficacy (7). Together, these studies highlight the limitations of *P. falciparum* vaccine targets, such as short-lived antibody responses, antigenic variability, and incomplete functional activity.

Furthermore, those and other proteins such as the Early Transcribed Membrane Protein 5 (Etramp5.Ag1) and the Heat-Shock Protein 40 (HSP40.Ag1) have been established as markers of *P. falciparum* exposure and have been extensively used in serosurveillance studies (8),(9). Nevertheless, certain proteins with analogous vaccine-desired characteristics, such as the members of the Sub-Telomeric Variable Open Reading frame protein family (STEVOR), have not been as thoroughly examined in scientific literature regarding their antigenicity and patterns of antibody responses in understanding their immunological significance.

STEVOR protein family is part of the Variable Surface Antigens (VSA) superfamily of proteins identified in *P. falciparum*, alongside the *var* genes products *P. falciparum* Erythrocyte Membrane Protein 1 (*Pf*EMP1) and the Repetitive Interspersed protein family (RIFIN), characterised by hypervariability between members or each family (10). VSA proteins are expressed by the parasite blood stage of infection and are transported to the membrane of the infected erythrocytes (IE), associated with immune evasion and sequestration, two important aspects of the parasite virulence (11),(12). *Pf*EMP1 variants differ primarily by their Duffy-Binding Like (DBL) domains, where hundreds of DBL variants have already been identified (13). Seropositivity to variants from the *Pf*EMP1 protein family has been demonstrated to increase with age and infection. In adults, there is a higher breadth of responses, the number of variants an individual is seropositive to, which stabilises over time. Additionally, seropositivity is higher in children with confirmed infection compared to those without. Furthermore, specific

variants have been associated with seroprevalence early in life, while to other emerges later and some are rarely recognised in a population study in Papa New Guinea, suggesting geographically specific or serodominant variants (14). Evidence also suggests that immunity to uncomplicated malaria is associated with broad repertoire of antibodies to *PfEMP1* variants, where protection to severe diseases is associated with immunity to specific variants (15). Other VSA research suggests a wide breadth of responses to *PfEMP1* variants early after first infection which showed no correlation with clinical outcome, unlike breadth of responses to other merozoite proteins found to be predictive of clinical outcome of the infection (16),(17).

Although less research has been conducted on the STEVOR protein family, a study using a large number of VSA families variants from the 3D7 *P. falciparum* reference strain, including STEVORs, among healthy individuals aged 6 to 20 years in Uganda, revealed that the overall breadth of responses correlated with age but not with clinical malaria, with one STEVOR variant identified to stimulate antibody responses related to protective immunity (18). The variability of STEVOR proteins is predominantly concentrated in their hypervariable domain (V2), while the remainder of the protein exhibits high conservation across variants. The V2 domain is also suggested to protrude into the extracellular space upon translocation and expression on the IE and alongside the semi-conserved domain, it is shown that it possess antigenic properties (19),(20).

This study proposes the use of a previously established library of purified STEVOR non-reference strain V2 domain recombinant antigens to examine the antibody responses in individuals across diverse age groups and malaria endemicity settings (21). The two samples' settings selected for the study were samples from a longitudinal cohort from Uganda in 2010 (FIGHTMAL), characterised by high *P. falciparum* transmission intensity and subsequent high exposure to infection in all ages. Around fifteen percent of the study individuals from FIGHTMAL presented with clinical malaria (fever above 37°C). This study compares the

serological responses to the STEVOR recombinant antigens in individuals from FIGHTMAL to those from settings characterised by low malaria transmission and infection exposure and individuals presenting with asymptomatic malaria only. The aim was to investigate the relationship between antibody responses to STEVORs and malaria endemic settings characterised by contrasting levels of *P. falciparum* endemicity, transmission intensity, and infection exposure.

This study investigated the breadth of antibody responses to the STEVOR library, using the multiplex bead-based platform Luminex, across serological samples from individuals in different age groups, with and without current *P. falciparum* infection, and in settings with contrasting malaria endemic status. This approach seeks to identify variants linked to early immunity, which may be important in conferring protection against the infection. The study compared the antibody responses to the STEVOR library to validated biomarkers of serological incidence (*PfAMA1*, *PfMSP1.19*, *Rh5.1*, *Etramp5.Ag1*, *HSP40.Ag1* and *CSP*) (22).

The study hypothesis was that individuals from higher *P. falciparum* endemicity settings, characterised by higher infection exposure, are expected to have higher breadth of antibody responses to the STEVOR proteins in age and exposure dependent manner.

7.2 Methods

7.2.1 Samples source

The samples used in the study were dried blood spots (DBS) and serum samples coming from three larger studies (described in detail in Chapter 6): 1) Serum samples from the longitudinal cohort “Correlating protection against malaria with serum profiles against *Plasmodium falciparum* antigen repertoires: FIGHTMAL” (LSHTM Ethics #5539; Med Biotech Laboratories in Kampala and the Uganda National Council of Science and Technology Ethics #HS699), conducted in the Apac district, Northern Uganda in 2010 (23); 2) DBS samples from the cluster-randomised two-arm mass drug administration (MDA) trial “Mass drug administration of Ivermectin and Dihydroartemisinin-piperaquine as an additional intervention for malaria elimination: MASSIV” (LSHTM Ethics #15823; The Gambia Government/MRC joint Ethics Committee #1593), conducted in the Upper River region, The Gambia, in 2018 and 2019 (24); and 3) DBS samples from the quadruple-blinded cluster-randomised MDA placebo-controlled trial “Adjunctive Ivermectin Mass Drug Administration for Malaria Control on the Bijagos Archipelago of Guinea-Bissau: MATAMAL” (LSHTM Ethics: #19156; Comite Nacional de Eticas de Saude Guinea-Bissau #084/CNES/INASA/2020), conducted in the Bijagos Archipelago of Guinea-Bissau in 2021 and 2022 (25). The samples selected for this study received ethical approval from the London School of Hygiene and Tropical Medicine ethical review committee (#21505). Additional demographics data, clinical malaria recorded data, infectious status (microscopy and/or PCR), and age and gender data collected as part of the original studies was also used in this study and covered by the outlined above ethics.

7.2.2 Study populations

A total of 1,278 DBS samples and 1,270 serum samples were selected and used in the study as follows (described in detail in Chapter 6): 1) FIGHTMAL (n=1,270), a longitudinal cohort conducted in the Apac district in Northern Uganda in 2010, with participants sampled at

baseline (n=498) prior to treatment, 6 weeks after intervention with artemether/lumefantrine combination therapy (ACT) (n=491), and 16 weeks after ACT intervention (n=281) (23). 2) MASSIVE (n=561), a cluster-randomised two-arm massive drug administration trial with dihydroartemisinin-piperaquine (DP) and ivermectin (IVM) in 2018 and 2019, samples subset one month after the second year of intervention in November 2019 (24). 3) MATAMAL (n=717), a quadruple-blinded cluster-randomised placebo-controlled trial with DP and IVM in 2021 and 2022, samples subset two months after intervention in November 2021 (n=204) and samples subset two months after intervention in November 2022 (n=513) (25). All available samples from FIGHTMAL were used in the study and for the MASSIV and MATAMAL trials, samples sub-selection was based on all malaria positive samples identified through *varATS P. falciparum* qPCR at the sampling timepoints per trial, indicated above. Additionally, an equal number of malaria-negative samples also based on *varATS* qPCR results were selected for comparison. These malaria-negative samples were matched to the positive samples by time point, age group and study arm for each trial.

Samples from November 2019 (MASSIV), November 2021 (MATAMAL) and November 2022 (MATAMAL) were merged to a single data set, referred to MASSIV/MATAMAL in this paper, resulting in a sample set (n=1,278) representing low malaria endemicity, providing a more robust sample size for the analysis (26),(Hutchins H *et. al*, unpublished). Merging these two data sets across three timepoints was justified as the populations were from similar ethnic groups, similar geographical and malaria transmission settings, and similar if not the same malaria endemicity from trials involving similar intervention strategies (24),(25). No samples were collected during 2020 due to the COVID-19 pandemic.

Antibody responses data for both FIGHTMAL and MASSIV/MATAMAL were stratified by age groups: i) below 5 years of age, representing individuals lacking clinical immunity; ii) individuals aged between 5 and 15 years, characterised by the absence of clinical immunity but

possessing some degree of parasitological immunity, also the age group associated with the highest exposure to malarial infection; iii) individuals above 15 years of age, individuals with mature immunity and presumed clinical and parasitological immunity attributed to naturally acquired immunity (NAI) to the parasite (27),(28). The Mann-Whitney non-parametric test was performed to assess significant differences in MFI continuous data between MASSIV and MATAMAL, and again on seropositivity data between the two datasets, to evaluate the justification for merging them.

7.2.3 Recombinant proteins panel

A total of 21 recombinant antigens were chemically coupled onto MagPlex magnetic beads, according to defined optimum concentration per antigen (EC50) (21),(29). The recombinants were as follows: six established markers of malaria seroincidence (*PfAMA1*, *PfMSP1.19*, *Etramp5.Ag1*, *HSP40.Ag1*, *Rh5.1*, and *CSP*), an internal serological control (*Tetanus.toxoid*), three STEVOR semi-conserved domain (SC) recombinant antigens and 11 STEVOR large hypervariable domain (V2) recombinant antigens as part of previously described STEVOR recombinant antigen library, including: four variants from Guinean clinical isolate (*PfGN01*), three variants from Senegalese clinical isolate (*PfSN01*), one variant from Kenyan clinical isolate (*PfKE01*), two variants from laboratory strains from Cambodia (*PfHB3* and *PfIt*) and one variant from a laboratory strain from Gabon (*PfGA01*) (21).

7.2.4 Multiplex bead-based serological assay

For MASSIV/MATAMAL samples, one 6mm punch per DBS was previously eluted in 400 µl of antibody elution buffer (0.05% Tween 20, 0.5% BSA, 0.02% NaN₃, and 0.002% *E. coli* extract), resulting in 1/400 sample dilution. For FIGHTMAL samples, 1µl of serum was eluted in 400 µl of the same antibody elution buffer, resulting in 1/400 sample dilution. All eluates were stored at -20°C. Samples were thawed overnight and 50µl per sample was used to perform the Luminex serology assay in a 96-well plate, following an optimised standard operating

procedure, described elsewhere (29). In brief, samples were incubated with the antigen coupled beads, washed with 1xPBS/Tween 20 and further incubated with Goat anti-human IgG R-phycoerythrin (R-PE) labelled antibody (Jackson ImmunoResearch ©) for signal detection using the MagPix© bioanalyzer to obtain antibody levels data in form of median fluorescence intensity (MFI). MFI is a proxy measure of antibody titres in the eluted sample. A five-fold standard curve from 1/50 to 1/31250 concentration of pooled positive serum from Ugandan adults (PRISM) was used as positive control and for assessing plate-to-plate variations of the assay (30). Two wells of elution buffer were used per plate to assess the background noise of the assay and two wells per plate of malaria naïve serum samples provided by UKHSA (formally Public Health England: PHE) in 1/400 dilution were used as negative controls. MFI data was quality assessed using the standard curves and background adjusted (subtracting the background readings from the sample readings per antigen), resulting in clean MFI signal data ready for analysis.

7.2.5 Statistical Analysis

All data were cleaned, coded, and analysed using R 4.2.3 computational platform (R Core Team, 2023). The seroprevalence of each recombinant antigen was calculated as the proportion of participants with MFI levels above the established seropositivity threshold per antigen (29). This was defined as an MFI above the mean MFI plus three standard deviations (SD) of the malaria naïve negative samples. A logistic regression model was used to analyse the effect of age group, sampling timepoint (FIGHTMAL) or intervention arm (MASSIV/MATAMAL), and their interaction on seropositivity for each antigen. The model predictors were age, sampling time/intervention arm and interaction term. The seropositivity percentage was converted to proportions, and logistic regression was performed using `logit()` function with binomial family in R computational platform. Separate models were fit for each antigen to evaluate antigen-specific effects. Statistical significance of predictors was determined using Z-tests for model

coefficients and directionality of differences, and p-values < 0.05 were considered significant. Breadth of antibody responses was calculated per individual in each study and was categorised as being seropositive for 1-3, 4-6, 7-9, or 10-11 of the STEVOR V2 recombinants. Grouping the breadth of responses was justified due to large differences in seropositivity sample size of breadth of responses to individual number of antigens. MFI continuous data per recombinant per study was presented as raincloud plots.

7.3 Results

7.3.1 Samples statistics

The total number of FIGHTMAL serum samples used in the study was 1,270, out of which 52% (n=664) were female and 48% (n=606) were male. There were no samples of individuals above 15 years of age for the 16-week post-ACT treatment group, highlighted in bold text in Table 1.

Table 1: FIGHTMAL samples characteristics.

	Baseline	6 weeks	16 weeks
Sample size (n)	498	491	281
Age strata, n (%)			
< 5 years	197 (39.6)	201 (40.9)	201 (71.5)
5 years – 15 years	168 (33.7)	166 (33.8)	80 (29.5)
> 15 years	133 (26.7)	124 (25.3)	0 (0.0)
Microscopy prevalence, n (%)			
< 5 years	87 (44.2)	43 (21.6)	7 (3.5)
5 years – 15 years	88 (51.5)	41 (24.7)	8 (9.8)
> 15 years	15 (11.3)	0 (0.0)	0 (0.0)
Total	190 (38.2)	84 (17.2)	15 (5.3)
qPCR prevalence, n (%) *			
< 5 years	35 (17.8)	40 (19.9)	16 (8.0)
5 years – 15 years	44 (26.2)	48 (28.9)	7 (8.5)
> 15 years	35 (26.3)	18 (14.5)	0 (0.0)
Total	114 (22.9)	106 (21.6)	23 (8.2)
Total prevalence (Microscopy or qPCR), n (%)			

< 5 years	122 (61.9)	83 (41.3)	23 (11.4)
5 years – 15 years	132 (78.9)	89 (53.6)	15 (18.8)
> 15 years	50 (37.6)	18 (14.5)	0 (0.00)
Total	304 (61.1)	190 (38.7)	38 (13.5)

***Samples negative for microscopy are further tested on rRNA qPCR.**

The total number of MASSIV/MATAMAL DBS samples used in the study was 1,278, out of which 48% (n=618) were female and 52% (n=660) were male. The sample size of individuals below five years of age in the intervention group was significantly lower compared to the rest of the groups. This discrepancy was because all qPCR positive samples from both clinical trials were included in this study and negative samples were matched per age group and treatment arm, and there were few qPCR positive samples from the trials particularly in the lowest age group of the intervention arm.

Table 2: MASSIV/MATAMAL samples characteristics.

	Control	Intervention
Sample size (n) *	768	540
Age strata, n (%)		
< 5 years	106 (13.80)	48 (8.89)
5 years – 15 years	290 (37.76)	178 (32.96)
> 15 years	372 (48.44)	314 (58.15)

***50% of samples across each age group for both, control, and intervention arm, were *P. falciparum varATS* qPCR positive.**

7.3.2 FIGHTMAL seroprevalence

The odds of being seropositive for the < 5 years age group at baseline were significant for all markers of exposure and STEVOR recombinants but *PfKE01_100005800* (Z-score: -0.622; p-value: 0.534), according to the logistic regression analysis summarised in Table 3.

There was a significant decrease in seropositivity for individuals above 15 years of age compared to the younger age groups for all markers of exposure, except for *PfAMA1* (Z-score: -1.904; p-value: 0.064). Conversely, there was a significant increase in seropositivity for few of the V2 STEVOR recombinants, including *PfKE01_100005800*, *PfIT_130077600*, *PfHB3_040028900* and *PfGA01_070034600*. The rest of the recombinants showed a significant increase of seropositivity for the 5 – 15 years group compared to < 5 years group, with no significant change in the oldest age group.

Sampling showed a positive effect on seropositivity 6 weeks post-treatment only for CSP (Z-score: 4.884; p-value: < 0.001) and STEVOR6_SC (Z-score: 2.892; p-value: 0.003) while a negative effect was observed for HSP40.Ag1 (Z-score: -3.454; p-value: < 0.001). Sampling at 16 weeks post treatment had a positive effect for all of the *PfSN01* strain STEVOR recombinants.

Interaction terms generally showed no significant effect in this dataset, indicating that changes in seropositivity were mostly driven by main effects of age or the time point of sampling.

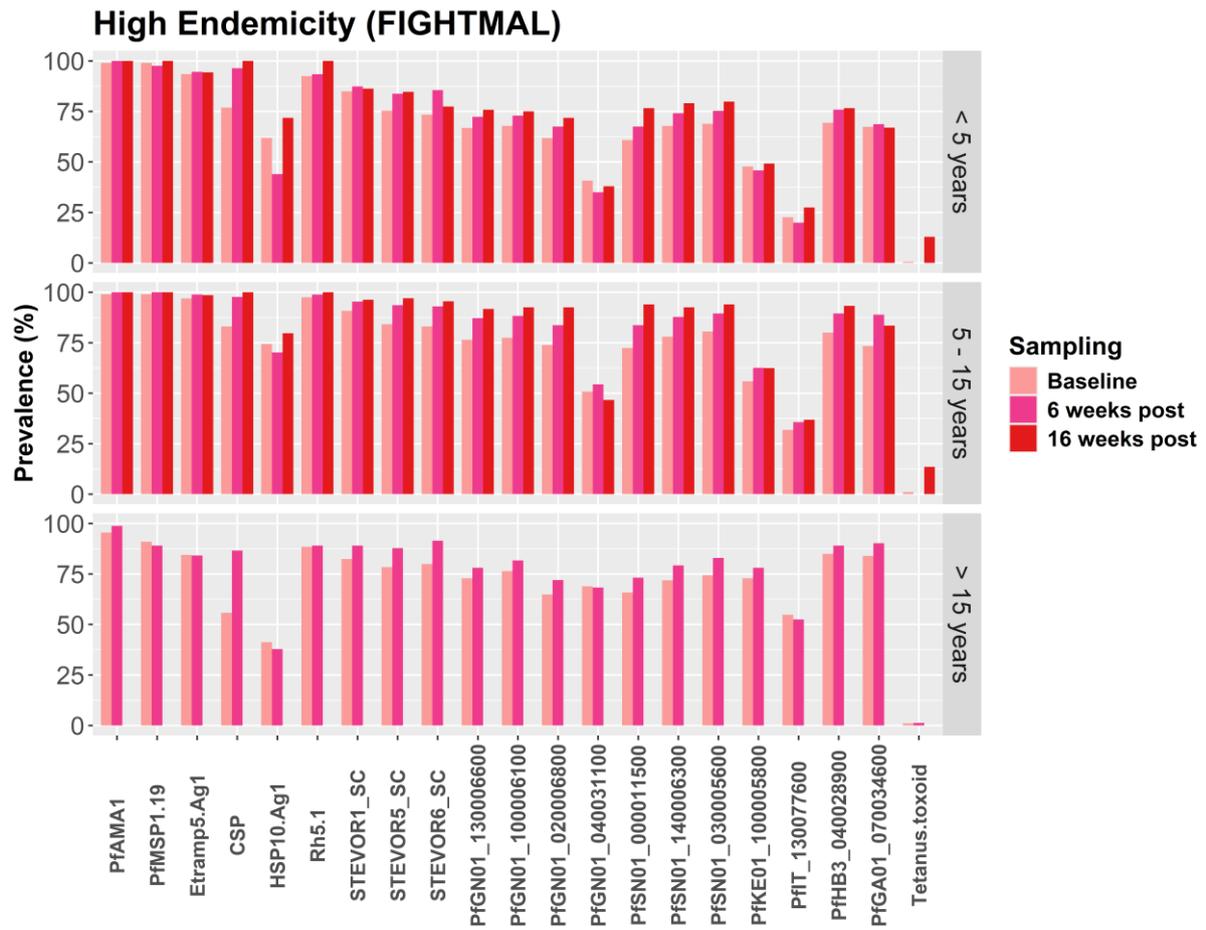


Figure 1: Percentage seropositive FIGHTMAL samples for baseline (prior to treatment) (salmon), 6 weeks after ACT treatment (pink) and 16 weeks after ACT treatment (red), stratified by age-group (< 5 years, 5-15 years and > 15 years). Recombinant antigens are shown on the x-axis. STEVOR antigens with the _SC suffix refer to semi-conserved antigen targets. STEVOR antigens *PfGN01_130006600* – *PfGA01_070034600* were all based on the large hypervariable loop.

Table3: Logistic regression model analysing the effect of age group, sampling time point and their interaction on seropositivity for each antigen for FIGHTMAL

samples, in data format: Z-score (p-value)

Antigen	Baseline (5 years)	5 – 15 years	> 15 years	6 weeks post	16 weeks post	5-15 years x 6 weeks post	> 15 years x 6 weeks post	5-15 years x 16 weeks post
PfAMA1	6.294 (< 0.001)	-0.019 (0.984)	-1.904 (0.064)	< 0.001 (0.997)	< 0.001 (0.997)	1.707 (1.000)	< 0.001 (0.999)	1.707 (1.000)
PfMSP1.19	6.294 (< 0.001)	-0.019 (0.984)	-2.955 (0.003)	-1.022 (0.307)	< 0.001 (0.999)	< 0.001 (0.999)	0.720 (0.471)	1.707 (1.000)
Etramp5.Ag1	9.039 (< 0.001)	1.536 (0.125)	-2.725 (0.006)	0.455 (0.649)	0.360 (0.718)	0.868 (0.385)	-0.422 (0.673)	0.679 (0.497)
CSP	6.965 (< 0.001)	1.499 (0.134)	-4.284 (< 0.001)	4.884 (< 0.001)	< 0.001 (0.999)	0.087 (0.929)	-0.900 (0.368)	< 0.001 (0.999)
HSP40.Ag1	3.216 (0.001)	2.604 (0.009)	-3.978 (< 0.001)	-3.454 (< 0.001)	2.051 (0.040)	1.653 (0.098)	1.957 (0.050)	-0.451 (0.652)
Rh5.1	9.098 (< 0.001)	2.108 (0.035)	-1.321 (0.035)	0.345 (0.730)	< 0.001 (0.999)	0.725 (0.469)	-0.155 (0.876)	< -0.001 (0.999)
STEVOR1_SC	8.504 (< 0.001)	1.723 (0.085)	-0.660 (0.509)	0.681 (0.496)	0.378 (0.705)	1.008 (0.313)	0.814 (0.416)	1.556 (0.119)
STEVOR5_SC	6.626 (< 0.001)	2.097 (0.036)	0.695 (0.487)	2.003 (0.045)	2.244 (0.025)	1.114 (0.265)	0.430 (0.667)	2.259 (0.024)
STEVOR6_SC	6.158 (< 0.001)	2.271 (0.023)	1.497 (0.134)	2.892 (0.003)	0.914 (0.361)	0.526 (0.599)	0.549 (0.582)	2.662 (0.008)
PfGN01_130006600	4.535 (< 0.001)	2.057 (0.039)	1.276 (0.202)	1.151 (0.249)	1.922 (0.055)	1.346 (0.178)	0.069 (0.945)	2.026 (0.043)
PfGN01_100006100	4.793 (< 0.001)	2.084 (0.037)	1.846 (0.065)	1.075 (0.283)	1.538 (0.124)	1.499 (0.134)	0.236 (0.814)	2.319 (0.020)
PfGN01_020006800	3.216 (0.001)	2.494 (0.013)	0.608 (0.543)	1.150 (0.250)	2.051 (0.040)	1.027 (0.304)	0.267 (0.789)	2.615 (0.009)
PfGN01_040031100	-2.541 (0.011)	1.961 (0.049)	5.416 (< 0.001)	-1.155 (0.248)	-0.557 (0.577)	1.319 (0.187)	0.716 (0.474)	-0.166 (0.868)
PfSN01_000011500	2.947 (0.003)	2.361 (0.018)	1.013 (0.311)	1.349 (0.177)	3.287 (0.001)	1.139 (0.254)	0.184 (0.854)	2.513 (0.012)
PfSN01_140006300	4.793 (< 0.001)	2.201 (0.028)	0.851 (0.395)	1.338 (0.181)	2.449 (0.014)	1.099 (0.272)	0.299 (0.765)	1.650 (0.099)
PfSN01_030005600	5.048 (< 0.001)	2.589 (0.009)	1.190 (0.234)	1.397 (0.163)	2.432 (0.015)	1.056 (0.291)	0.562 (0.574)	1.737 (0.082)
PfKE01_100005800	-0.622 (0.534)	1.586 (0.113)	4.923 (< 0.001)	-0.381 (0.703)	0.283 (0.777)	1.208 (0.227)	1.135 (0.256)	0.719 (0.472)
PfIT_130077600	-7.075 (< 0.001)	1.998 (0.046)	6.259 (< 0.001)	-0.649 (0.516)	1.077 (0.281)	1.012 (0.311)	0.214 (0.831)	-0.101 (0.919)
PfHB3_040028900	5.175 (< 0.001)	2.367 (0.018)	3.545 (< 0.001)	1.427 (0.154)	1.587 (0.113)	1.118 (0.264)	0.087 (0.931)	2.093 (0.036)
PfGA01_070034600	4.664 (< 0.001)	1.275 (0.202)	3.693 (< 0.001)	0.279 (0.780)	-0.083 (0.934)	2.799 (0.005)	1.328 (0.184)	1.857 (0.063)
Tetanus toxoid	-5.141 (< 0.001)	0.572 (0.567)	0.554 (0.579)	< -0.001 (0.999)	3.214 (0.001)	< -0.001 (0.999)	< -0.001 (1.00)	-0.513 (0.608)

* Highlighted p-value scores indicate significant change of seropositivity according to the set threshold value of 0.05, and positive or negative Z-score indicates the direction of the change. The “x” in the column names indicate the interaction between the variables listed before and after the “x” symbol.

7.3.3 MASSIV/MATAMAL seroprevalence

The odds of being seropositive for the < 5 years age group in the Control arm were significant for all antigens and STEVOR recombinants, according to the logistic regression model results summarised in Table 4. The oldest age group (> 15 years) showed strong, highly significant effect for most antigens, indicating increased seropositivity in older individuals. Age group 5 – 15 years also showed significant effects for many antigens, although the magnitude of the effect was lower compared to > 15 years group.

There was no significant difference in seropositivity for the majority of the markers of exposure between the intervention arms, except for the detected significant decreases in seroprevalence in the Intervention arm for CSP (Z-score: -2.579; p-value: 0.009) and Rh5.1 (Z-score: -2.957; p-value: 0.003). However, there was a significant increase in seropositivity for all semi-conserved STEVOR recombinants and the majority of the STEVOR V2 recombinants in the Intervention arm, with the exception of PfGN01_040031100 (Z-score: -2.313; p-value: 0.021) and PfIT_130077600 (Z-score: -3.398; p-value: < 0.001).

There was some evidence for the effect of the interaction between age and intervention arm, although not as significant as the individual effect of the two variables on seroprevalence.

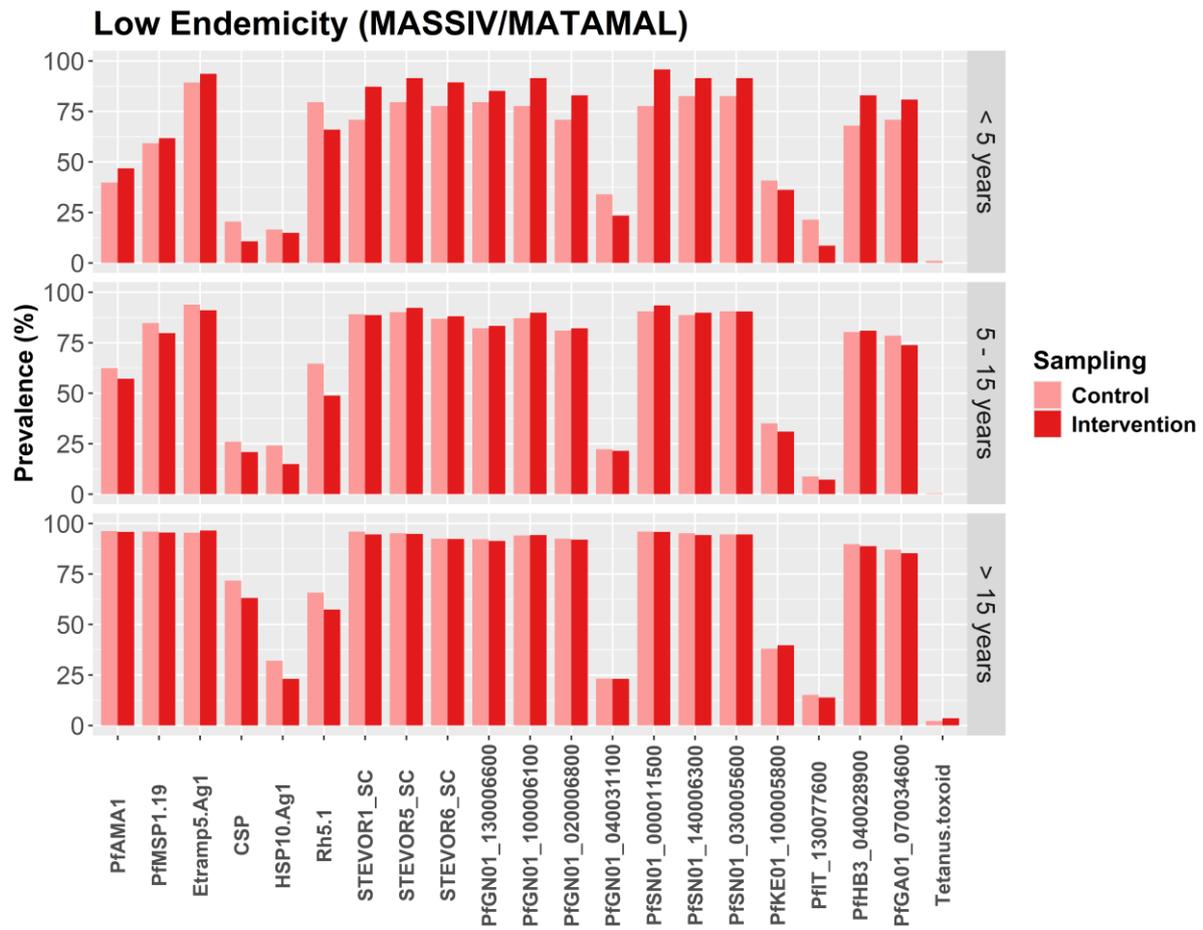


Figure 2: Percentage seropositive MASSIV/MATAMAL samples for the Control (salmon) and Intervention (red) treatment arms, stratified by age-group (< 5 years, 5-15 years and > 15 years). Recombinant antigens are shown on the x-axis. STEVOR antigens with the _SC suffix refer to semi-conserved antigen targets. STEVOR antigens *PfGN01_130006600* – *PfGA01_070034600* were all based on the large hypervariable loop.

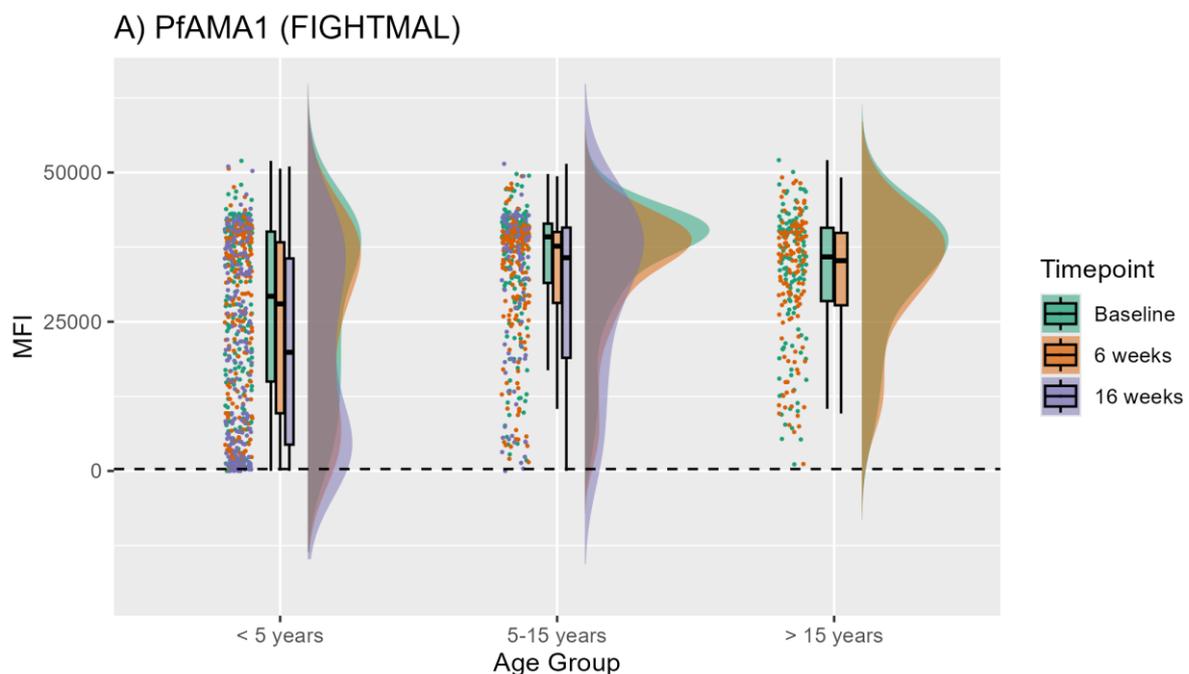
Table4: Logistic regression model analysing the effect of age group, intervention arm and their interaction on seropositivity for each antigen for MASSIV/MATAMAL samples, in data format: Z-score (p-value)

Antigen	Control (5 years)	5 – 15 years	> 15 years	Intervention	5-15 years x Intervention	> 15 years x Intervention
PfAMA1	-2.783 (0.005)	4.357 (< 0.001)	8.916 (< 0.001)	1.373 (0.170)	-1.707 (0.012)	-0.686 (0.493)
PfMSP1.19	2.521 (0.012)	5.336 (< 0.001)	7.022 (< 0.001)	0.493 (0.622)	-1.286 (0.199)	-0.385 (0.699)
Etramp5.Ag1	9.018 (< 0.001)	1.547 (0.122)	2.171 (0.029)	1.480 (0.139)	-1.744 (0.081)	-0.445 (0.656)
CSP	-7.545 (< 0.001)	1.271 (0.204)	9.461 (< 0.001)	-2.579 (0.009)	1.252 (0.210)	1.012 (0.311)
HSP40.Ag1	-8.273 (< 0.001)	1.824 (0.068)	3.478 (< 0.001)	-0.430 (0.667)	-1.221 (0.222)	-0.906 (0.365)
Rh5.1	7.545 (< 0.001)	-3.222 (0.001)	-2.994 (0.003)	-2.957 (0.003)	0.163 (0.870)	1.084 (0.279)
STEVOR1_SC	5.555 (< 0.001)	4.269 (< 0.001)	5.659 (< 0.001)	3.818 (< 0.001)	-2.519 (0.012)	-2.408 (0.016)
STEVOR5_SC	7.545 (< 0.001)	2.804 (0.005)	4.206 (< 0.001)	3.194 (0.001)	-1.545 (0.122)	-1.882 (0.059)
STEVOR6_SC	7.137 (< 0.001)	2.316 (0.021)	3.861 (< 0.001)	3.004 (0.003)	-1.798 (0.072)	-1.855 (0.064)
PfGN01_130006600	7.545 (< 0.001)	0.619 (0.536)	3.395 (< 0.001)	1.397 (0.163)	-0.767 (0.443)	-1.061 (0.289)
PfGN01_100006100	7.137 (< 0.001)	2.415 (0.016)	4.286 (< 0.001)	3.596 (< 0.001)	-1.916 (0.055)	-2.039 (0.041)
PfGN01_020006800	5.555 (< 0.001)	2.294 (0.022)	5.074 (< 0.001)	2.766 (0.006)	-1.697 (0.089)	-1.655 (0.098)
PfGN01_040031100	-4.325 (< 0.001)	-2.519 (0.012)	-2.313 (0.021)	-2.519 (0.024)	1.393 (0.164)	1.537 (0.124)
PfSN01_000011500	7.137 (< 0.001)	3.324 (< 0.001)	4.701 (< 0.001)	4.662 (< 0.001)	-2.656 (0.008)	-2.898 (0.004)
PfSN01_140006300	8.104 (< 0.001)	1.695 (0.090)	3.660 (< 0.001)	2.543 (0.011)	-1.503 (0.133)	-1.787 (0.074)
PfSN01_030005600	8.104 (< 0.001)	2.242 (0.025)	3.503 (< 0.001)	2.543 (0.011)	-1.733 (0.0883)	-1.495 (0.135)
PfKE01_100005800	-2.521 (0.012)	-1.149 (0.250)	-0.551 (0.581)	-0.919 (0.358)	0.032 (0.974)	0.896 (0.358)
PfIT_130077600	-7.344 (< 0.001)	-3.327 (< 0.001)	-1.571 (0.116)	-3.398 (< 0.001)	1.716 (0.086)	2.242 (0.025)
PfHB3_040028900	4.824 (< 0.001)	2.716 (0.006)	4.958 (< 0.001)	3.348 (< 0.001)	-2.195 (0.028)	-2.249 (0.024)
PfGA01_070034600	5.555 (< 0.001)	1.692 (0.091)	3.775 (< 0.001)	2.253 (0.024)	-2.347 (0.019)	-1.821 (0.024)
Tetanus toxoid	-6.235 (< 0.001)	-0.695 (0.487)	0.905 (0.365)	< -0.001 (0.999)	< 0.001 (0.999)	< 0.001 (0.999)

* Highlighted p-value scores indicate significant change of seropositivity according to the set threshold value of 0.05, and positive or negative Z-score indicates the direction of the change. The “x” in the column names indicate the interaction between the variables listed before and after the “x” symbol..

7.3.4 Recombinants seroreactivity profiles

There was a marked difference between the seroreactivity profiles of FIGHTMAL samples versus the MASSIV/MATAMAL samples to all validated markers of seroincidence, with higher antibody titres (MFI values) in FIGHTMAL samples in all age groups across all sampling time points, as demonstrated in the raincloud plots for *Pf*AMA1 in Figure 3. The seroreactivity profile figures to the rest of the validated markers can be found in Supplementary Figures 1.1-1.5. However, the antibody titre profiles to the STEVOR recombinants were not as different between the two studies, more pronounced for the STEVOR V2 library recombinants, as demonstrated for *Pf*GA01_070034600 in Figure 5 (Supplementary Figures 1.8-1.17) compared to the semi-conserved recombinants, exemplified for STEVOR1_SC in Figure 4 (Supplementary Figure 1.6-1.7).



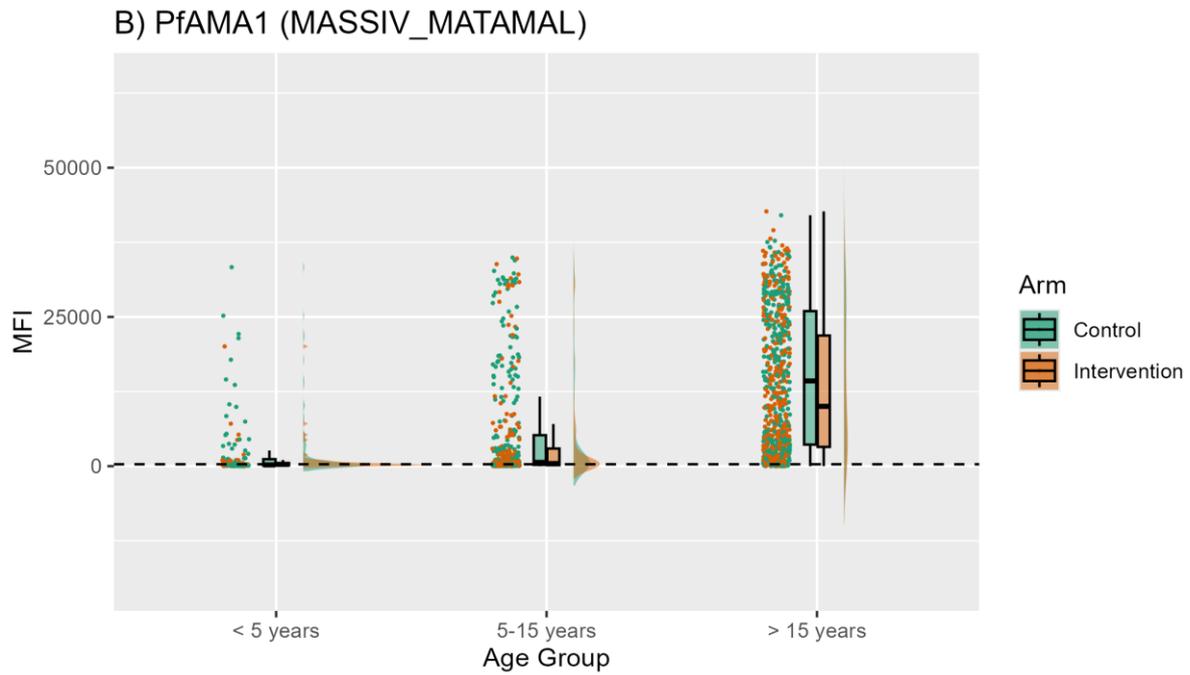
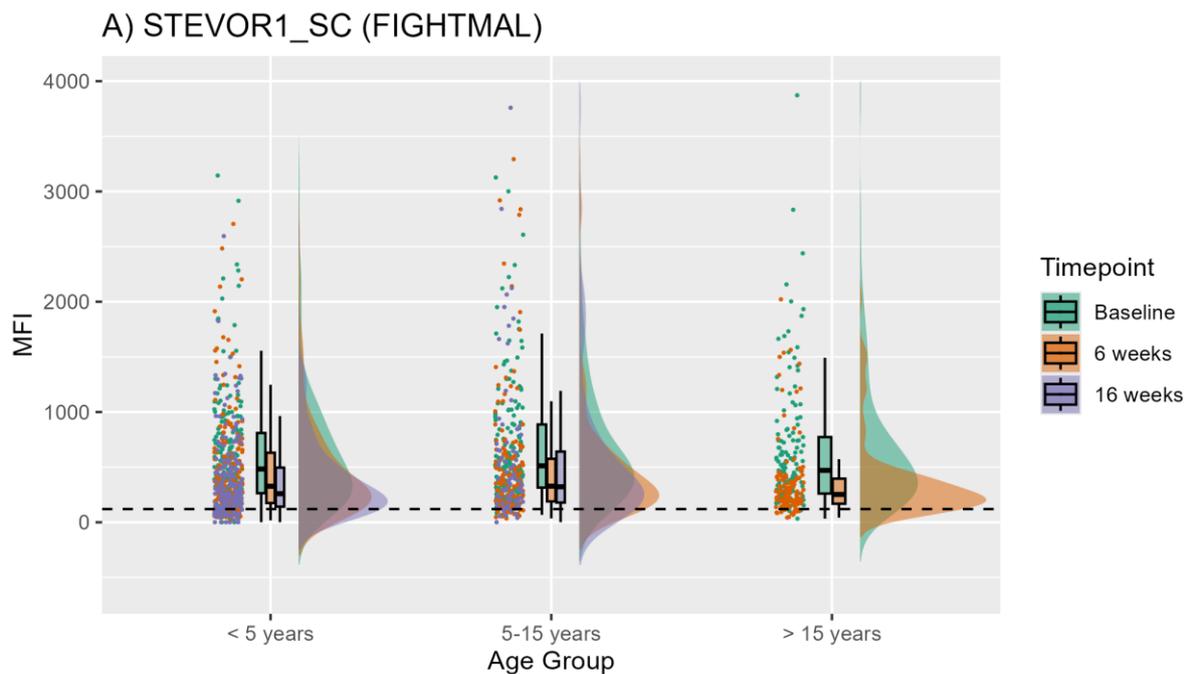


Figure 3: *PfAMA1* raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.



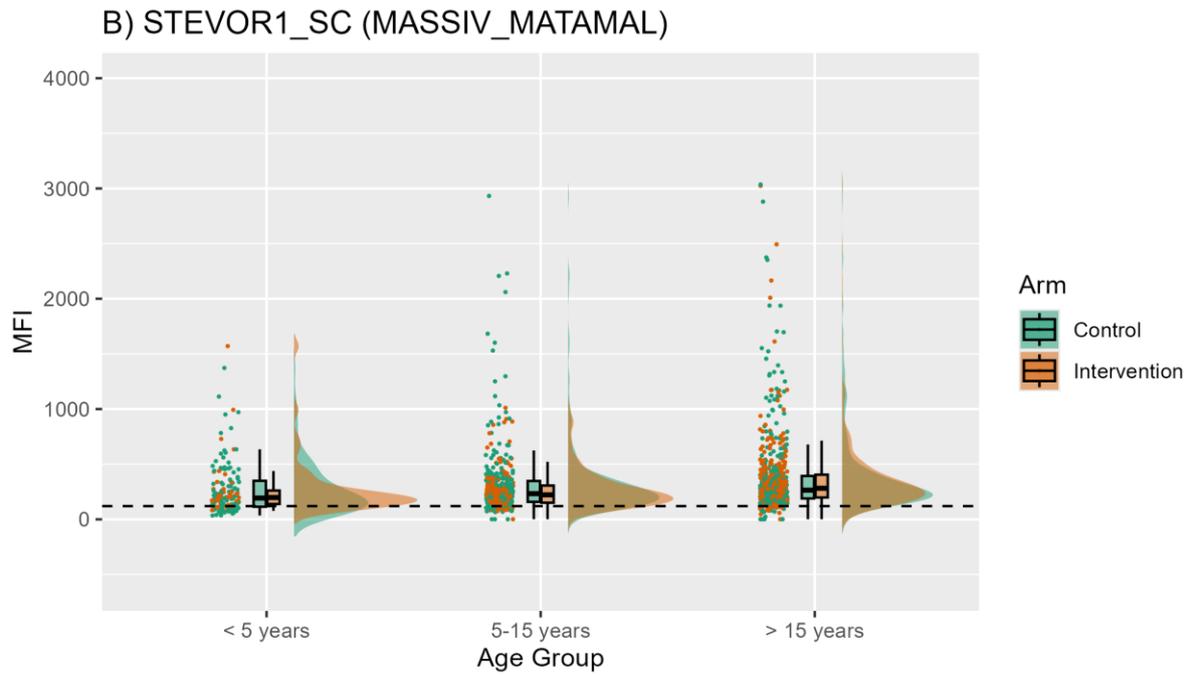
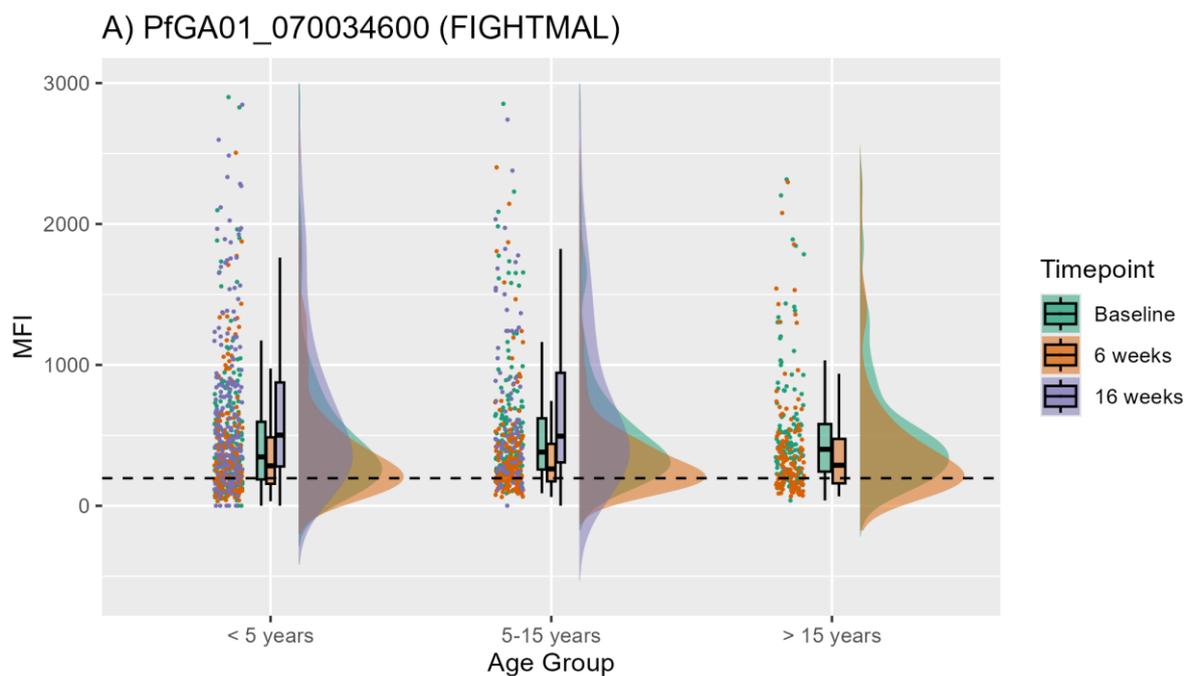


Figure 4: STEVOR1_SC raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.



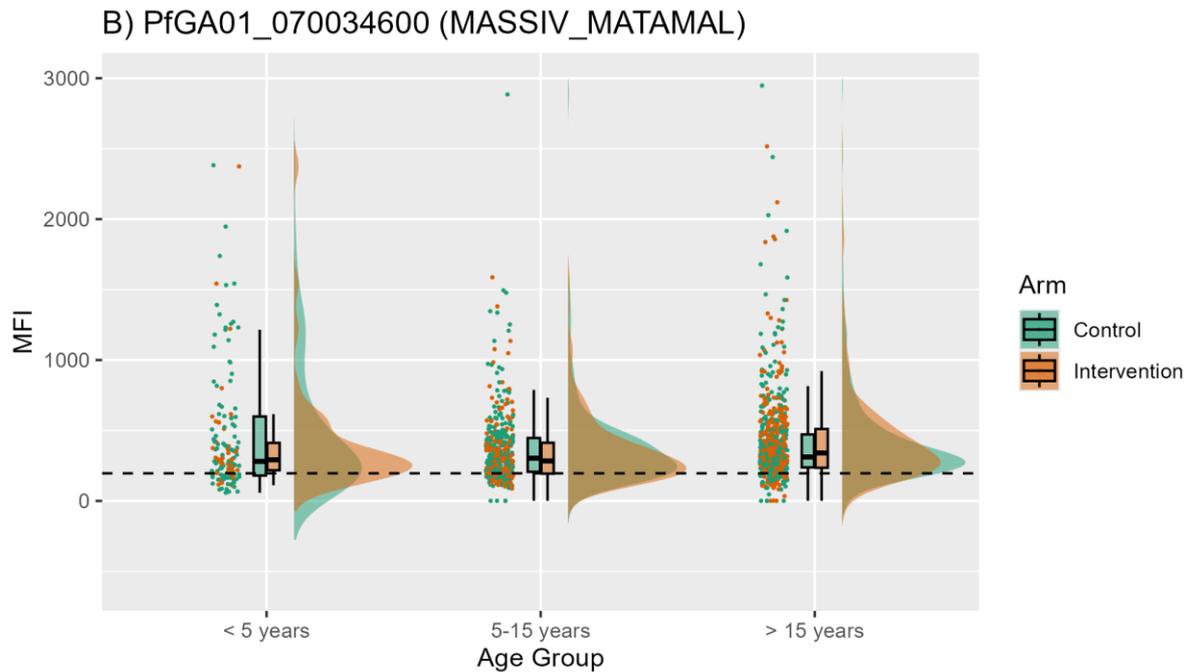


Figure 5: *PfGA01_070034600* raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.

7.3.5 Breadth of responses to STEVOR V2 recombinants

A total of 85.4% (n=1,084) of all FIGHTMAL samples used in the study were found to be seropositive to at least one of the STEVOR V2 recombinants. These represent 90.6% (n=451) of baseline samples, 79.6% (n=391) of 6 weeks post ACT treatment samples and 86.1% (n=242) of 16 weeks post ACT treatment samples. A total of 36.5% (n=396) of total seropositive FIGHTMAL samples were seropositive to all 11 STEVOR V2 recombinant antigens: 35.0% (n=158) of baseline samples, 24.5% (n=96) of 6 weeks post samples, and 58.7% (n=142) of 16 weeks post samples, making the most of the samples in the 10-11 antigen group illustrated in Figure 6.

Consequently, a total of 14.7% (n=186) of all samples from the FIGHTMAL study were not seropositive to any of the 11 STEVOR V2 recombinants. Those represent 9.4% (n=47) of baseline samples, 20.4% (n=100) of 6 weeks post ACT treatment samples, and 13.9% (n=39) of 16 weeks post ACT treatment samples. This data is not included in Figure 6, as it reduces the resolution of breadth of responses data falling under the rest of the antigen categories. An adapted version including the breadth of responses to 0 antigens can be found in Supplementary Figure 2.

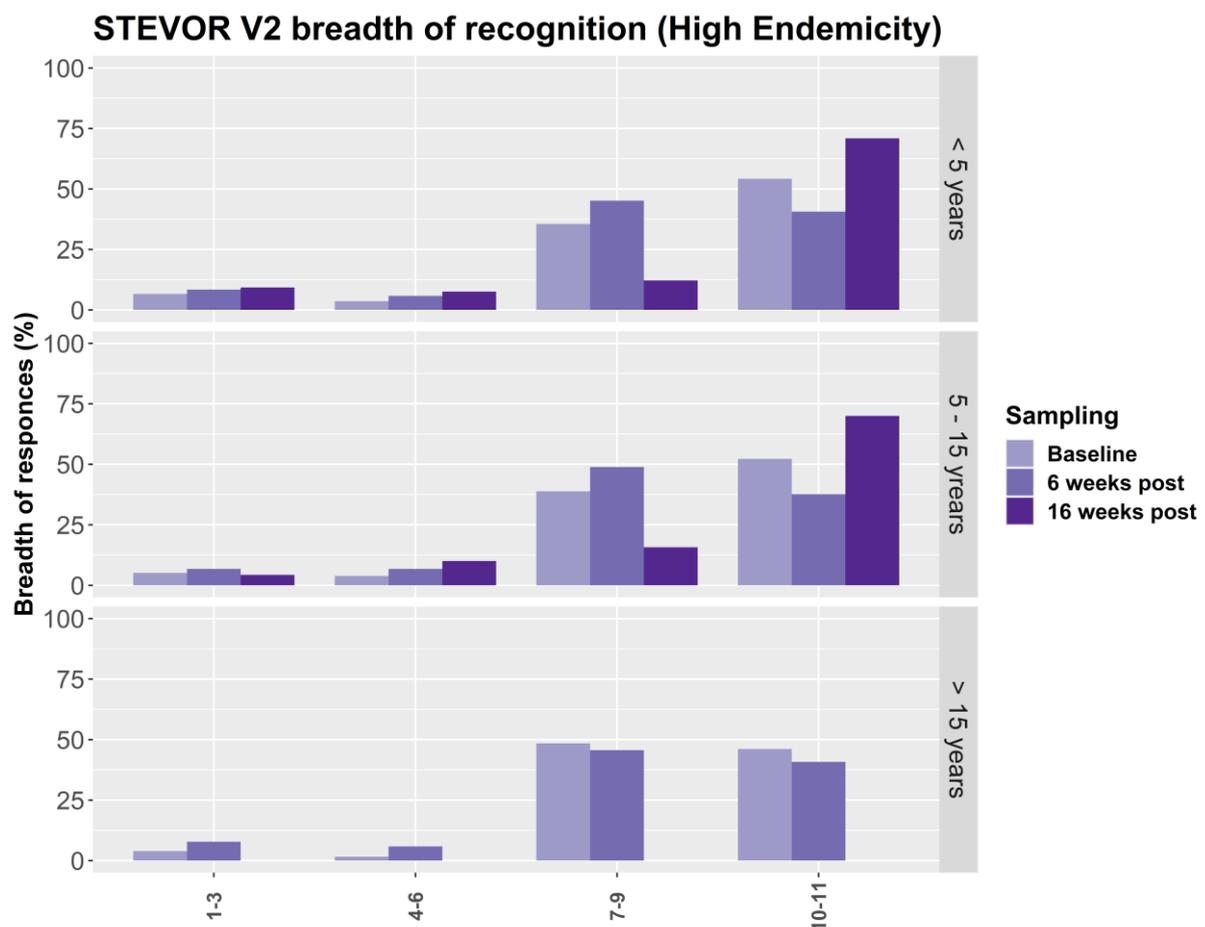


Figure 6: FIGHTMAL population (n=1,084) prevalence of the breadth of antibody responses to the STEVOR V2 recombinant antigens according to four categories (1-3, 4-6, 7-9, and 10-11 recombinant antigens). Breadth of responses prevalence is stratified by age group (< 5 years, 5-15 years and > 15

years) and sampling timepoint: Baseline (grey), 6 weeks post treatment (light purple) and 16 weeks post treatment (dark purple), indicated in the legend.

A total of 94.6% (n=1,209) of all MASSIV/MATAMAL samples were seropositive to at least one of the STEVOR V2 recombinant antigens: 93.6% (n=702) of the control arm and 96.1% (n=507) of the intervention arm samples, (Figure 7). Most of the seropositive MASSIV/MATAMAL samples of 43.8% (n=560) were seropositive for eight STEVOR V2 recombinants: 44.4% (n=312) of control samples and 48.9% (n=248) of intervention samples, making the majority of the samples in the 7-9 recombinant antigens group (Figure 7). Almost all samples seropositive to eight V2 recombinants were systematically seronegative to three members of the library: *PfGN01_040031100*, *PfIT01_130077600*, and *PfKE01_100005800*.

Consequently, a total of 5.4% (n=69) of the MASSIV/MATAMAL samples did not cross the seropositivity threshold for any of the STEVOR V2 recombinant antigens, representing 6.7% (n=50) of samples from the control arm and 3.8% (n=20) from the intervention arm. This data was not included in Figure 7, as it reduces the resolution of breadth of responses data falling under the rest of the antigen categories. An adapted version including the breadth of responses to 0 antigens can be found as Supplementary Figure 3.

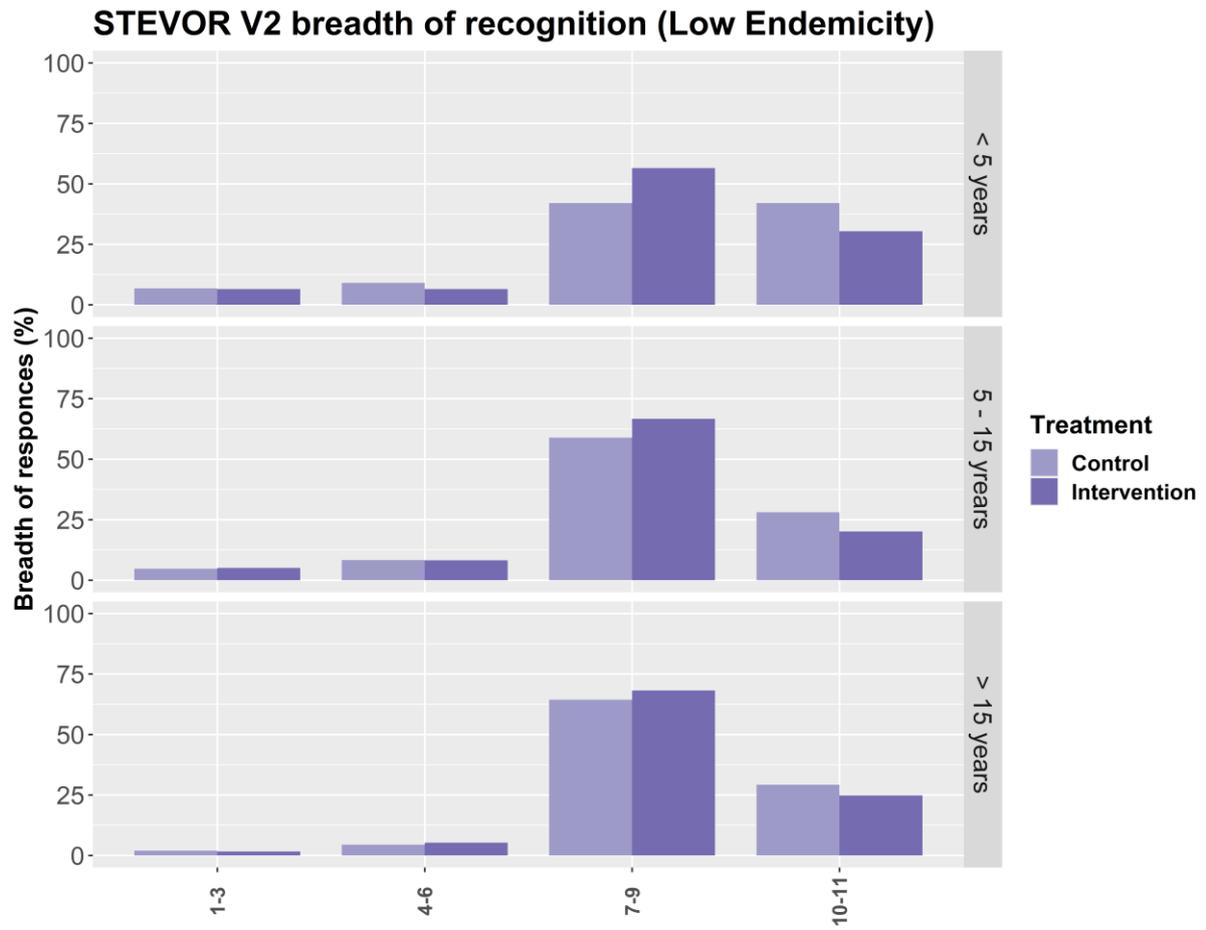


Figure 7: MASSIV/MATAMAL population (n=1,209) prevalence of breadth of antibody responses to the STEVROR V2 recombinant antigens according to four categories (1-3, 4-6, 7-9, and 10-11 recombinant antigens). Breadth of responses prevalence is stratified by age group (< 5 years, 5-15 years and > 15 years) and intervention arms: Control (grey) and Intervention (purple), indicated in the legend.

7.4 Discussion

This study aimed to investigate the impact of the *P. falciparum* contrasting endemicity levels in different settings on the breadth of antibody responses to the STEVOR protein family, defined as the number of STEVOR hypervariable domain recombinant antigens that individuals are seropositive to, out of the total 11 geographically diverse recombinants tested (21). The populations used in the study included participants from high (FIGHTMAL) and low (MASSIV/MATAMAL) malaria endemicity settings. The sample sizes of the two investigated populations were matched, and stratified by age groups, according to their assumed malaria immunological profiles (27),(28). Seroprevalence to validated markers of seroincidence was used as a control for the data analysis and results interpretation. There were no significant differences in seroprevalence across the age groups for the validated markers of seroincidence, but for HSP40.Ag1 and Rh5.1 in FIGHTMAL and PfAMA1 and PfMSP1.19 in MASSIV/MATAMAL. Although these recombinants are regarded as markers of malaria exposure, with age and exposure dependent acquisition of antibodies, the lack of differences in the high endemicity setting (FIGHTMAL) is likely to be since the area is holoendemic for malaria at the time of sample collection (31). The highest *P. falciparum* prevalence was recorded above 70% in children under five, thus an immune response to markers such as PfAMA1 and PfMSP1.19, but not to the sporozoite protein CSP, HSP40 or RG5.1, is expected to be already high in this age group (32),(33). Seroprevalence to these markers was comparably lower in the MASSIV/MATAMAL sample set, but the lack of significant increase in seroprevalence with age can be attributed to the fact that 50 % of the sub-selected samples for this study from the original studies were *P. falciparum varATS* qPCR positive across each age groups for both, the control, and intervention arms, not representing the true *P. falciparum* prevalence in the settings. Furthermore, these results were supported by the serological outcomes from both MASSIV and MATAMAL trials, showing no significant increase of

antibody acquisition to the markers associated with age or intervention arm (Kositz C *et al* unpublished), (Hutchins H *et al.* unpublished). However, we found a significant decrease in CSP and Rh5.1 seroprevalence in the intervention arm compared to the control arm, where there was a significant increase in seroprevalence for all semi-conserved and most variable domain STEVOR recombinants. These results suggest that the intervention influenced the seropositivity, but the direction of influence was not consistent between markers of exposure and the STEVOR recombinants. Since antigens such as CSP and Rh5.1 are associated with exposure to the malaria parasite during the sporozoite and blood-stage of the infection, the reduction of transmission and infection rates in the intervention arm can explain the decrease of seropositivity (34). On the other hand STEVORs are variable surface antigens expressed during the intraerythrocytic stages of the parasites and increase seroprevalence to the recombinants may indicate a continues low-level infections, not entirely cleared by the intervention (35).

In the FIGHTMAL study, the high prevalence of malaria may explain the observed high seropositivity to all STEVOR recombinants. The study samples were seropositive to all markers of exposure, even in children under 5 years of age, which is consistent with the holoendemicity of malaria in the region, as outlined above (31). However, a notable observation is that the overall recorded seroprevalence to the STEVOR recombinants was considerably higher compared to the markers of exposure, even in the low endemicity samples for the under-5 age group—an unusual finding. The high seropositivity to a large repertoire of STEVORs even in low endemicity settings in children could indicate a heightened or more durable immune response to these antigens compared to markers of exposure, or may reflect the parasite's ability to express these antigens across various stages of infection, making them more immunogenic or detectable over time, an observation not previously demonstrated for STEVORs(35). Individuals' malaria status historical data was not available for the study and

only qPCR results at the time of sample collection was available which provides only a snapshot of malaria infection at this particular time. Previous studies comparing seropositivity to STEVOR variants with other VSA families have yielded varying conclusions. While some have reported broader immune responses to STEVORs compared to *PfEMP1* due to conserved immunogenic domains, others have shown the opposite, with STEVOR seroprevalence being lower than that of *PfEMP1* and the RIFIN family (36), (37). Furthermore, antigenic variability and cross-reactivity among VSAs have been highlighted differently across studies, ranging from observed cross-reactivity across heterogeneous STEVOR variants to weak cross-reactivity of antibodies to *PfEMP1* (37), (38). However, studies on STEVORs generally do not emphasize their suggested role in chronic parasitaemia but instead provide nuanced insights into their protective potential and antigenic variability, underscoring the complex immunological landscape of STEVORs and their potential implications for malaria control strategies (37), (38).

Another possible explanation for this discrepancy could be the threshold used to determine seropositivity. The use of serum from malaria-naive individuals in the UK, who have never been exposed to malaria, to establish the threshold may not be entirely appropriate. These individuals lack any prior exposure to the parasite, which might lead to an underestimated threshold that does not accurately reflect low-level exposure in endemic regions, as well as may not account for low levels of cross-reactivity or background immune activity in endemic populations. This could result in an artificial inflation of seropositivity in individuals with minimal or past exposure. Using alternative methods to derive seropositivity thresholds, such as utilising samples from endemic areas with known exposure histories and longitudinal data, was not feasible for this study due to limited knowledge about STEVOR variability and exposure to variants. Another potential approach is the application of mathematical or machine

learning techniques, such as finite mixture models, which determine positive and negative populations for each antigen relative to the studied data (39).

Notably, the differences in seroprevalence profiles between the two studies for the validated markers is also observed in the antibody titre profiles (MFI levels), with higher MFI values recorded for individuals from the FIGHTMAL study across all ages, as compared to those from MASSIV/MATAMAL, due to higher exposure to the infection, as these recombinant antigens are regarded as markers of *P. falciparum* infection exposure with age and exposure dependent antibody acquisition (22),(28),(40). However, the antibody titre profiles to the STEVOR recombinants were not as different between the two studies, particularly the antibody titre profiles to the STEVOR V2 library recombinants. This suggests that the difference in endemicity level and subsequently exposure to the infection do not influence the level of antibody responses to members of the STEVOR protein family. This is a controversial result to already published literature using peptide arrays of STEVOR semi-conserved and hypervariable domains, showing that antibody levels against both STEVOR domains are positively correlated with age, exposure and clinical outcome of the infection (20). However, the results in the mentioned study are based on only six STEVOR variants from the 3D7 reference strain, which might not be the best approach for investigating the seroreactivity to STEVORs in clinical samples (21),(41).

The study then investigated if the endemicity levels influenced the breadth of responses to the STEVOR V2 library, hypothesising that higher endemicity levels will increase the breadth of responses. Aligning with the hypothesis most of the samples from the lower endemicity setting showed lower breadth of responses, with most of the MASSIV/MATAMAL samples having a breadth of responses to eight out of the 11 recombinants and were systematically seronegative to *PfGN01_04003100*, *PfKE01_100005800*, and *PfIT01_130077600*, compared to the majority of FIGHTMAL individuals being seropositive to all 11 STEVOR V2 recombinants.

Interestingly, these variants were the only V2 recombinants to which there was no significant increase in seroprevalence with age in the FIGHTMAL samples too, suggesting no differences in levels of exposure to them with time in the high endemicity setting. These variants are from a Guinean and Kenyan clinical isolates, and a Cambodian laboratory strain, respectively. Thus, there was no pattern of geographical association between them, however, it appears that individuals from the MASSIV/MATAMAL sample set were not exposed to these variants and seroreactivity to them in the rest of the samples could be attributed to cross reactivity. Although MASSIV/MATAMAL samples present with lower breadth of responses compared to the FIGHTMAL samples, there was a substantial number of samples seropositive to all tested STEVOR V2 recombinants. The controlled human malaria infection (CHMI) trial by Turner *et al.* in 2011 demonstrated that even a single brief exposure to *P. falciparum* in children generates a broad repertoire of antibodies against PfEMP1 domains. These antibodies were also found to be cross-reactive to PfEMP1 domains from parasites with different genomes (16). These breadth of responses were also shown to be none uniform amongst individuals, suggesting that differences in the acquisition of repertoire of antibodies to VSAs are more likely due to individual differences in the ability to induce immune response (16). Thus, a newly formed hypothesis that antibodies to variable STEVOR domains are generated quickly even in low endemicity settings, characterised by low *P. falciparum* exposure in children is made, and the differences in the breadth of responses were more likely due to individual host immune variations. Moreover, there was a high level of cross-reactivity of antibodies to VSAs demonstrated in a CHMI trial showing the generation of cross-reactive antibodies to more than five different parasite's genomes after a single specific infection, could also explain the high breadth of responses to the specifically selected STEVOR V2 recombinants coming from various geographical clinical isolates and laboratory strains (16),(42).

Moreover, there was no pattern of antibody acquisition to variants following other variants from a preliminary network analysis which is not presented in the study since the network did not give any insight into the exploration of the STEVOR V2 breadth of response. The lack of pattern is also observed for other VSA variants in the CHMI studies (16),(42).

The first two STEVOR V2 variants that individuals become seropositive to were *PfGA01_070034600* for both tested populations, and *PfSN01_030005600* for FIGHTMAL and *PfSN01_000011500* for MASSIV/MATAMAL. For the individuals with low breadth of responses, antibodies against STEVOR variants to which individuals become seropositive first seem to compete with each other, as those seropositive to one of the variants are always seronegative to the other and vice versa. Although previous work on antibody responses to other VSAs demonstrated that the number of serorecognised variants in individuals is dependent on individual immune characteristics and the sequence of variant recognition is random (16),(42).

7.5 Conclusion

This study provides further insights into seroreactivity against *P. falciparum* STEVOR proteins using a panel of 11 STEVOR V2 recombinant antigens based on STEVOR variants from geographically diverse, non-reference strain sequences.

This study does not demonstrate that high malaria endemicity, with relatively greater exposure to *P. falciparum* would result in individuals having a wider breadth of responses to the STEVOR hypervariable domain panel. As a result, a new hypothesis was formed, that high breadth of responses to STEVOR V2 variants is expected even after a single infection in children, which may reflect the parasite's ability to express these antigens across various stages of infection, making them more immunogenic or detectable over time, or as a result of degree of cross-reactivity of antibodies to STEVOR variants

However, there were some differences, where individuals from lower *P. falciparum* endemicity settings presented with lower breadth of responses, whilst antibody levels of reactivity to the panel was not related to the endemicity. Additionally, three recombinants were found to be serodominant despite the endemicity levels, and three recombinants were found to be the least serorecognised in low endemicity settings, however those findings were not strongly supported to be significant.

The question remains of why is the STEVOR hypervariability important for the parasite, considering the quick immune response to all variants tested, despite the setting's endemicity level and levels of infection exposure. Further work testing the panel on individuals with different clinical malaria outcome should be done to elucidate the importance of STEVORs in disease pathology, already suggested in literature.

The findings underscore the need for refining serological thresholds and advancing our understanding of STEVOR immunogenicity to accurately interpret seroprevalence patterns and their implications in malaria infection immunity.

7.6 Limitations

This study utilised residual samples from historical studies, resulting in the absence of samples representing individuals above 15 years of age in the 16 weeks post ACT treatment from the FIGHTMAL study, due to serum unavailability. Consequently, the interpretation of results was based on the other two age groups (“< 5 years” and “5 – 15 years”). Moreover, the longitudinal samples from the FIGHTMAL study for the 16 weeks post treatment sampling were lower than those of the other sampling points, potentially influencing the conclusions made from the percentage seropositivity analysis.

Pooling samples from two distinct population studies is suboptimal, particularly in studying human immune responses, which is the case of the MASSIV/MATAMAL sample set. However, this decision was based on the geographical proximity of the two populations, exhibiting similar *P. falciparum* endemicity and malaria transmission patterns, and almost identically designed interventional studies, with sampling at the same time of the year, aimed at achieving a more robust sample size for statistical significance. This aggregated sample set represents a setting of low malaria endemicity, although the sample set was chosen to have all cases from the original studies and matched controls, resulting in 50% *P. falciparum varATS* qPCR positivity across all age groups in both treatment arms. Nevertheless, PCR positivity provides only a snapshot of the malaria situation in the region, compare to antibody responses, thus these samples remain relevant for studying antibody responses in low endemicity settings. Sample sub-selection from the original sample sets was necessary due to resource constraints for conducting Luminex analysis with large target number.

Although the demographics and interventions of the two studies are similar if not identical, dissection of the MFI responses per study reveals significant differences in MFIs for all markers of exposure and STEVOR recombinants. These differences are illustrated in the series of

raincloud plots in Supplementary Figures 5.1-5.21, as determined by the Mann-Whitney non-parametric test for significance. However, these differences are less pronounced when comparing populations seropositivity according to the calculated threshold, shown in Supplementary Figure 4, which has a p-value of 0.03, indicating only a weak association of significant differences. Therefore, all comparative analysis between FIGHTMAL and MASSIV/MATAMAL was done looking at seropositivity rather than as a quantitative analysis using continuous MFI data.

Despite the presented findings, the study has notable limitations, and the results should be regarded as associative rather than conclusive. Ideally, the analysis would have been conducted separately for the two populations; however, the small sample sizes posed a challenge. A better approach would have been to focus on one trial and test the antigens against a larger sample size, even if many were malaria-negative, rather than pooling the two populations and sub-selecting. Unfortunately, resource constraints made this approach infeasible. To validate the conclusions drawn in this study, further research with larger sample sizes from Sub-Saharan African regions characterized by low malaria endemicity is necessary. Finally, the study compares antibody responses between two types of samples: serum (FIGHTMAL) and DBS (MASSIV/MATAMAL). To adjust for this difference as much as possible, samples were diluted to the same dilution factor in the same elution buffer prior processing them on the Luminex platform. However, while a 1/400 dilution of serum and a 1/400 dilution of DBS eluate may theoretically seem equivalent, significant limitations arise from differences in sample composition, analyte recovery, and assay compatibility. Serum is a purified liquid matrix, while DBS eluates contain additional components like haemoglobin and cell debris that can interfere with assays. Variability in analyte recovery from DBS, influenced by elution efficiency and haematocrit levels, may result in inconsistent concentrations compared to serum. Furthermore, analytes may degrade differently in DBS. The discussed factors make it difficult

to directly compare results between the two sample types without robust standardization. In this study, these limitations were unavoidable due to the nature of the available samples. While previous studies have shown no significant differences in seropositivity when using paired serum and DBS samples, the conclusions of this study should be interpreted as indicative rather than definitive (43), (44). The differences in sample types inherently limit the ability to fully compare the breadth of antibody responses across the tested populations. By measuring total IgG concentrations in both serum and DBS samples, antigen-specific antibody levels could potentially be normalized relative to total IgG, for instance by expressing responses as a ratio of antigen-specific IgG to total IgG.

7.7 Declarations

7.7.1 Competing interests.

All authors declare that the research was conducted in the absence of any financial, personal, or professional relationships that could be construed to have influenced the work.

7.7.2 Authors contributions

HV was responsible for the design, acquisition and analysis, original samples collection, laboratory work, interpretation of results and drafting of the work. KT was involved in the design of the work. KT and AL were involved in the interpretation of the results. HBM was involved in the laboratory work. HH, ES, ED, MN, CD, UA and AL were involved in original samples and data collection. AL and KT were responsible for securing funds for the study. All authors have approved the submitted version and have agreed to be personally accountable for their own contributions of this work.

7.8 List of abbreviations:

- PfEMP1: *Plasmodium falciparum* Erythrocyte Membrane Protein 1
- STEVOR: Sub-Telomeric Variable Open Reading frame family
- WHO: World Health Organisation
- CSP: Circumsporotite Protein
- PfAMA1: *Plasmodium falciparum* Apical Membrane Antigen 1
- PfMSP1.19: *Plasmodium falciparum* Merozoite Surface Protein 1
- Etramp5.Ag1: Early Transcribed Membrane Protein 5
- HSP40.Ag1: Heat-Shock Protein 40
- VSA: Variable Surface Antigens
- RIFIN: Repetitive Interspersed Family
- DBL: Duffy-Like Binding domain
- DBS: Dry Blood Spots
- MDA: Mass Drug Administration
- ACT: Artemisinin Combination Therapy
- DP: Dihydroartemisinin-piperaquine
- IVM: Ivermectin
- EC50: Half Maximum Effect Concentration
- SC: Semi-Conserved domain
- V2: Large hypervariable domain
- UKHSA: United Kingdom Health Security Agency
- MFI: Median Fluorescence Intensity
- SD: Standard Deviation
- Rh5.1: Reticulocyte-Binding protein Homologue 5

7.9 References

1. Dattoo MS, Natama MH, Somé A, Traoré O, Rouamba T, Bellamy D, et al. Efficacy of a low-dose candidate malaria vaccine, R21 in adjuvant Matrix-M, with seasonal administration to children in Burkina Faso: a randomised controlled trial. *www.thelancet.com* [Internet]. 2021 [cited 2023 Jun 10];397. Available from: <https://doi.org/10.1016/>
2. RTS, Partnership SCT. Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: Final results of a phase 3, individually randomised, controlled trial. *Lancet* [Internet]. 2015 Jul 4 [cited 2020 Sep 16];386(9988):31–45. Available from: </pmc/articles/PMC5626001/?report=abstract>
3. Miller LH, Baruch DI, Marsh K, Doumbo OK. The pathogenic basis of malaria. *Nature* [Internet]. 2002 Feb 7 [cited 2021 Nov 17];415(6872):673–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/11832955/>
4. Arnold KB, Chung AW. Prospects from systems serology research. *Immunology*. 2018 Mar 1;153(3):279–89.
5. Dieng CC, Ford CT, Lerch A, Doniou D, Vegesna K, Janies D, et al. Genetic variations of *Plasmodium falciparum* circumsporozoite protein and the impact on interactions with human immunoproteins and malaria vaccine efficacy. *Infect Genet Evol*. 2023 Jun 1;110:105418.
6. Malkin EM, Diemert DJ, McArthur JH, Perreault JR, Miles AP, Giersing BK, et al. Phase 1 Clinical Trial of Apical Membrane Antigen 1: an Asexual Blood-Stage Vaccine for *Plasmodium falciparum* Malaria. *Infect Immun* [Internet]. 2005 Jun [cited 2023 May 5];73(6):3677. Available from: </pmc/articles/PMC1111886/>

7. Blank A, Fürle K, Jäschke A, Mikus G, Lehmann M, Hüsing J, et al. Immunization with full-length *Plasmodium falciparum* merozoite surface protein 1 is safe and elicits functional cytophilic antibodies in a randomized first-in-human trial. *npj Vaccines* 2020 51 [Internet]. 2020 Jan 31 [cited 2023 May 4];5(1):1–15. Available from: <https://www.nature.com/articles/s41541-020-0160-2>
8. van den Hoogen LL, Stresman G, Présomé J, Romilus I, Mondélus G, Elismé T, et al. Selection of Antibody Responses Associated With *Plasmodium falciparum* Infections in the Context of Malaria Elimination. *Front Immunol* [Internet]. 2020 May 15 [cited 2024 Apr 18];11:928. Available from: </pmc/articles/PMC7243477/>
9. Wu L, Mwesigwa J, Affara M, Bah M, Correa S, Hall T, et al. Antibody responses to a suite of novel serological markers for malaria surveillance demonstrate strong correlation with clinical and parasitological infection across seasons and transmission settings in The Gambia. *BMC Med* [Internet]. 2020 Sep 25 [cited 2023 Mar 13];18(1):304. Available from: </pmc/articles/PMC7517687/>
10. Scherf A, Lopez-Rubio JJ, Riviere L. Antigenic Variation in *Plasmodium falciparum*. *Annu Rev Microbiol* [Internet]. 2008 Oct 11 [cited 2019 May 20];62(1):445–70. Available from: <http://www.annualreviews.org/doi/10.1146/annurev.micro.61.080706.093134>
11. Niang M, Bei AK, Madnani KG, Pelly S, Dankwa S, Kanjee U, et al. STEVOR is a *plasmodium falciparum* erythrocyte binding protein that mediates merozoite invasion and rosetting. *Cell Host Microbe* [Internet]. 2014 Jul 9 [cited 2022 Feb 22];16(1):81–93. Available from: <http://www.cell.com/article/S1931312814002182/fulltext>
12. Lavazec C, Sanyal S, Templeton TJ. Hypervariability within the Rifin, Stevor and PfmC-2TM superfamilies in *plasmodium falciparum*. *Nucleic Acids Res* [Internet].

- 2006 Dec [cited 2020 Oct 5];34(22):6696–707. Available from:
[/pmc/articles/PMC1751529/?report=abstract](#)
13. Chaudhry S, Singh V. A systematic review on genetic diversity of var gene DBL1 α domain from different geographical regions in Plasmodium falciparum isolates. *Infect Genet Evol.* 2021 Nov 1;95:105049.
 14. Barry AE, Trieu A, Fowkes FJI, Pablo J, Kalantari-Dehaghi M, Jasinskas A, et al. The Stability and Complexity of Antibody Responses to the Major Surface Antigen of Plasmodium falciparum Are Associated with Age in a Malaria Endemic Area. *Mol Cell Proteomics* [Internet]. 2011 [cited 2024 Apr 3];10(11). Available from:
[/pmc/articles/PMC3226400/](#)
 15. Tessema SK, Nakajima R, Jasinskas A, Monk SL, Lekieffre L, Lin E, et al. Protective Immunity against Severe Malaria in Children Is Associated with a Limited Repertoire of Antibodies to Conserved PfEMP1 Variants. *Cell Host Microbe* [Internet]. 2019 Nov 13 [cited 2024 Apr 18];26(5):579-590.e5. Available from:
<https://pubmed.ncbi.nlm.nih.gov/31726028/>
 16. Turner L, Wang CW, Lavstsen T, Mwakalinga SB, Sauerwein RW, Hermsen CC, et al. Antibodies against PfEMP1, RIFIN, MSP3 and GLURP Are Acquired during Controlled Plasmodium falciparum Malaria Infections in Naïve Volunteers. *PLoS One* [Internet]. 2011 [cited 2024 May 2];6(12):29025. Available from:
[/pmc/articles/PMC3236238/](#)
 17. Rono J, Osier FHA, Olsson D, Montgomery S, Mhoja L, Rooth I, et al. Breadth of Anti-Merozoite Antibody Responses Is Associated With the Genetic Diversity of Asymptomatic Plasmodium falciparum Infections and Protection Against Clinical Malaria. *Clin Infect Dis* [Internet]. 2013 Nov 15 [cited 2024 Apr 4];57(10):1409–16.

Available from: <https://dx.doi.org/10.1093/cid/cit556>

18. Kanoi BN, Nagaoka H, White MT, Morita M, Palacpac NMQ, Ntege EH, et al. Global Repertoire of Human Antibodies Against Plasmodium falciparum RIFINs, SURFINs, and STEVORs in a Malaria Exposed Population. *Front Immunol* [Internet]. 2020 May 12 [cited 2023 Apr 11];11. Available from: <https://pubmed.ncbi.nlm.nih.gov/32477363/>
19. Andersson A, Kudva R, Magoulopoulou A, Lejarre Q, Lara P, Xu P, et al. Membrane integration and topology of RIFIN and STEVOR proteins of the Plasmodium falciparum parasite. *FEBS J* [Internet]. 2020 Jul 1 [cited 2023 Mar 13];287(13):2744–62. Available from: <https://pubmed.ncbi.nlm.nih.gov/31821735/>
20. Zhou AE, Berry AA, Bailey JA, Pike A, Dara A, Agrawal S, et al. Antibodies to Peptides in Semiconserved Domains of RIFINs and STEVORs Correlate with Malaria Exposure. *mSphere*. 2019 Mar 20;4(2).
21. Vasileva H, Benavente ED, Last A. In-silico classification and antigen library expression of Plasmodium falciparum STEVOR hypervariable infected erythrocyte surface-expressed multivariant protein family. 2023 Nov 14 [cited 2024 Mar 19]; Available from: <https://www.researchsquare.com>
22. Drakeley CJ, Corran PH, Coleman PG, Tongren JE, McDonald SLR, Carneiro I, et al. Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. *Proc Natl Acad Sci U S A* [Internet]. 2005 Apr 5 [cited 2020 Sep 29];102(14):5108–13. Available from: www.pnas.org/cgi/doi/10.1073/pnas.0408725102
23. Proietti C, Verra F, Bretscher MT, Stone W, Kanoi BN, Balikagala B, et al. Influence of infection on malaria-specific antibody dynamics in a cohort exposed to intense

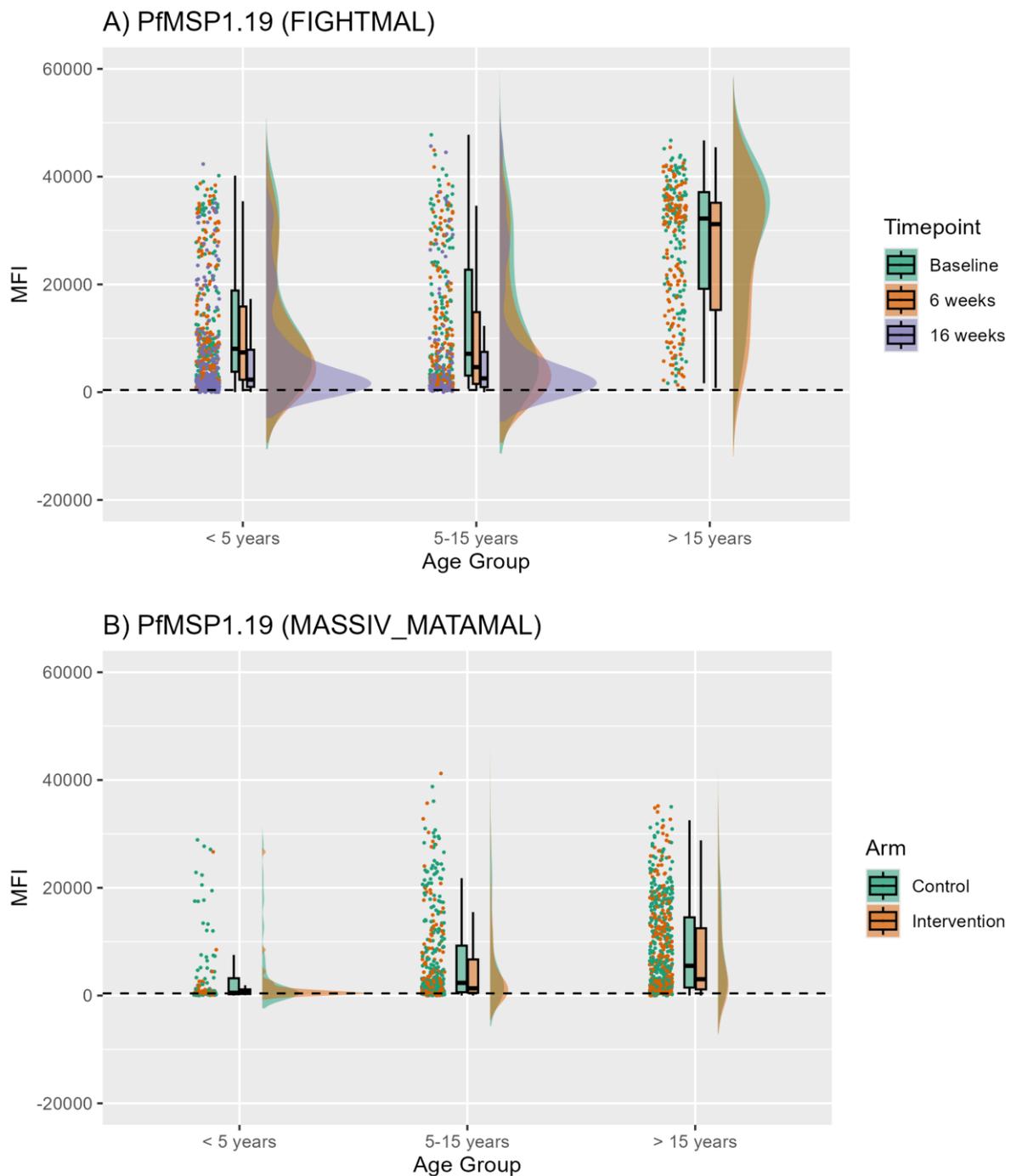
- malaria transmission in northern Uganda. *Parasite Immunol* [Internet]. 2013 May 1 [cited 2024 Mar 8];35(5–6):164–73. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1111/pim.12031>
24. Dabira ED, Soumare HM, Lindsay SW, Conteh B, Ceesay F, Bradley J, et al. Mass Drug Administration With High-Dose Ivermectin and Dihydroartemisinin-Piperaquine for Malaria Elimination in an Area of Low Transmission With High Coverage of Malaria Control Interventions: Protocol for the MASSIV Cluster Randomized Clinical Trial. *JMIR Res Protoc* [Internet]. 2020 Nov 1 [cited 2024 Mar 17];9(11). Available from: </pmc/articles/PMC7714640/>
 25. Hutchins H, Bradley J, Pretorius E, Teixeira Da Silva E, Vasileva H, Jones RT, et al. Protocol: Protocol for a cluster randomised placebo-controlled trial of adjunctive ivermectin mass drug administration for malaria control on the Bijagós Archipelago of Guinea-Bissau: the MATAMAL trial. *BMJ Open* [Internet]. 2023 Jul 7 [cited 2023 Sep 27];13(7):72347. Available from: </pmc/articles/PMC10335573/>
 26. Dabira ED, Soumare HM, Conteh B, Ceesay F, Ndiath MO, Bradley J, et al. Mass drug administration of ivermectin and dihydroartemisinin-piperaquine against malaria in settings with high coverage of standard control interventions: a cluster-randomised controlled trial in The Gambia. *Lancet Infect Dis* [Internet]. 2022 Apr 1 [cited 2024 Apr 24];22(4):519–28. Available from: <https://pubmed.ncbi.nlm.nih.gov/34919831/>
 27. Ranjha R, Singh K, Baharia RK, Mohan M, Anvikar AR, Bharti PK. Age-specific malaria vulnerability and transmission reservoir among children. *Glob Pediatr*. 2023 Dec 1;6:100085.
 28. Gonzales SJ, Reyes RA, Braddom AE, Batugedara G, Bol S, Bunnik EM. Naturally Acquired Humoral Immunity Against *Plasmodium falciparum* Malaria. *Front Immunol*

- [Internet]. 2020 Oct 29 [cited 2024 Apr 23];11:594653. Available from:
www.frontiersin.org
29. Wu L, Hall T, Ssewanyana I, Oulton T, Patterson C, Vasileva H, et al. Optimisation and standardisation of a multiplex immunoassay of diverse *Plasmodium falciparum* antigens to assess changes in malaria transmission using sero-epidemiology. Wellcome Open Res [Internet]. 2020 Apr 23 [cited 2020 Oct 15];4:26. Available from:
<https://doi.org/10.12688/wellcomeopenres.14950.1>
 30. Vasileva H, Chopo-Pizarro A, Ooko M, Dumont E, Wu L, Ssewanyana I, et al. Novel hypervariable erythrocyte surface expressed recombinant proteins show promise as serological markers of exposure to *Plasmodium falciparum* infection. bioRxiv [Internet]. 2023 Nov 2 [cited 2023 Nov 7];2023.10.31.564916. Available from:
<https://www.biorxiv.org/content/10.1101/2023.10.31.564916v1>
 31. Helb DA, Tetteh KKA, Felgner PL, Skinner J, Hubbard A, Arinaitwe E, et al. Novel serologic biomarkers provide accurate estimates of recent *Plasmodium falciparum* exposure for individuals and communities. 2015;
 32. Tassi Yunga S, Siriwardhana C, Fouda GG, Bobbili N, Sama G, Chen JJ, et al. Characterization of the primary antibody response to *Plasmodium falciparum* antigens in infants living in a malaria-endemic area. *Malar J* [Internet]. 2022 Dec 1 [cited 2024 May 1];21(1):346. Available from: [/pmc/articles/PMC9675181/](https://pubmed.ncbi.nlm.nih.gov/3675181/)
 33. Leonard CM, Uhomoibhi P, Abubakar A, Ogunniyi A, Mba N, Greby SM, et al. Dynamics of IgG antibody response against *Plasmodium* antigens among Nigerian infants and young children. *Front Immunol*. 2023 Aug 24;14:1208822.
 34. Ssewanyana I, Rek J, Rodriguez I, Wu L, Arinaitwe E, Nankabirwa JI, et al. Impact of a Rapid Decline in Malaria Transmission on Antimalarial IgG Subclasses and Avidity.

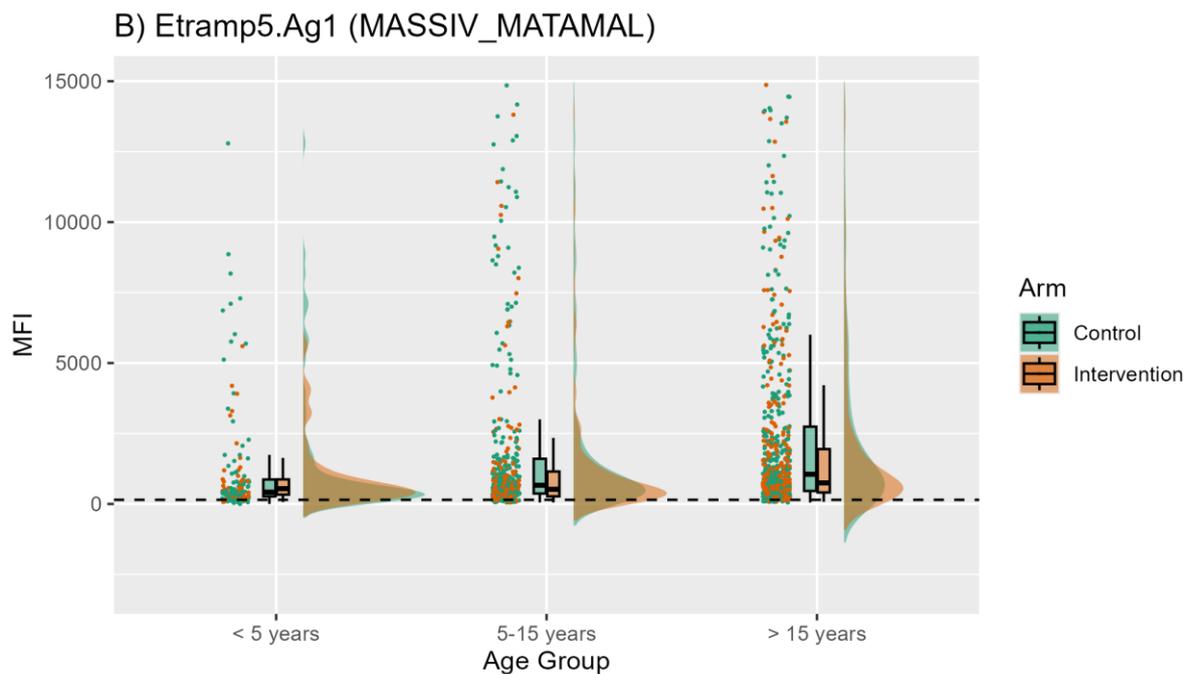
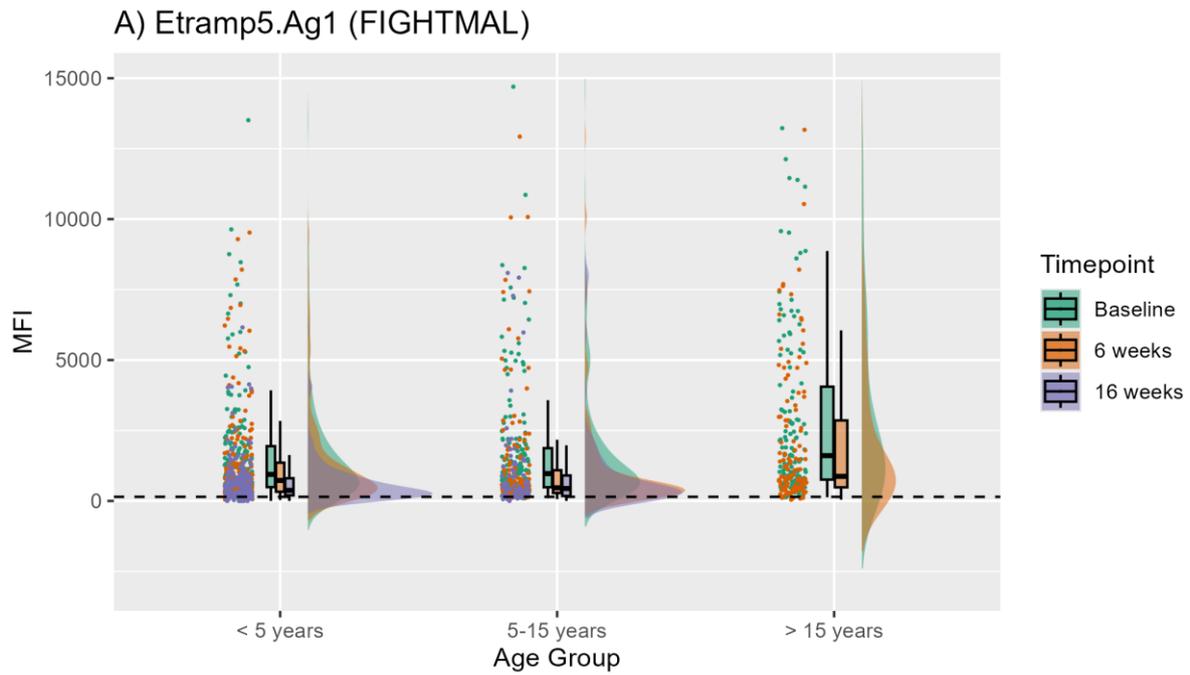
- Front Immunol. 2021 Jan 27;11:3308.
35. Wichers JS, Scholz JAM, Strauss J, Witt S, Lill A, Ehnold LI, et al. Dissecting the gene expression, localization, membrane topology, and function of the plasmodium falciparum STEVOR protein family. *MBio*. 2019 Aug 27;10(4).
 36. McLean FE, Azasi Y, Sutherland C, Toboh E, Ansong D, Agbenyega T, et al. Detection of naturally acquired, strain-transcending antibodies against rosetting Plasmodium falciparum strains in humans. *Infect Immun* [Internet]. 2024 Jul 1 [cited 2024 Dec 3];92(7):e0001524. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11238554/>
 37. Kanoi BN, Nagaoka H, White MT, Morita M, Palacpac NMQ, Ntege EH, et al. Global Repertoire of Human Antibodies Against Plasmodium falciparum RIFINs, SURFINs, and STEVORs in a Malaria Exposed Population. *Front Immunol* [Internet]. 2020 May 12 [cited 2023 Aug 24];11. Available from: <https://pubmed.ncbi.nlm.nih.gov/32477363/>
 38. Quintana MDP, Ch'ng JH, Moll K, Zandian A, Nilsson P, Idris ZM, et al. Antibodies in children with malaria to PfEMP1, RIFIN and SURFIN expressed at the Plasmodium falciparum parasitized red blood cell surface. *Sci Reports* 2018 81 [Internet]. 2018 Feb 19 [cited 2024 Dec 3];8(1):1–14. Available from: <https://www.nature.com/articles/s41598-018-21026-4>
 39. McLachlan GJ. Model-Based Clustering. *Compr Chemom*. 2009;2:655–81.
 40. Rodriguez-Barraquer I, Arinaitwe E, Jagannathan P, Kanya MR, Rosenthal PJ, Reik J, et al. Quantification of anti-parasite and anti-disease immunity to malaria as a function of age and exposure. *Elife* [Internet]. 2018 Jul 25 [cited 2024 Apr 24];7. Available from: <https://pubmed.ncbi.nlm.nih.gov/30044224/>

41. Blythe JE, Xue YY, Kuss C, Bozdech Z, Holder AA, Marsh K, et al. Plasmodium falciparum STEVOR proteins are highly expressed in patient isolates and located in the surface membranes of infected red blood cells and the apical tips of merozoites. *Infect Immun*. 2008 Jul;76(7):3329–36.
42. Antibody recognition of heterologous variant surface antigens after a single Plasmodium falciparum infection in previously naive adults - PubMed [Internet]. [cited 2024 May 2]. Available from: <https://pubmed.ncbi.nlm.nih.gov/17488905/>
43. Daag JV, Ylade M, Jadi R, Adams C, Cuachin AM, Alpay R, et al. Performance of Dried Blood Spots Compared with Serum Samples for Measuring Dengue Seroprevalence in a Cohort of Children in Cebu, Philippines. *Am J Trop Med Hyg* [Internet]. 2021 Jan 6 [cited 2024 Nov 27];104(1):130–5. Available from: <https://pubmed.ncbi.nlm.nih.gov/33146119/>
44. Meyers E, Heytens S, Formukong A, Vercruyssen H, De Sutter A, Geens T, et al. Comparison of Dried Blood Spots and Venous Blood for the Detection of SARS-CoV-2 Antibodies in a Population of Nursing Home Residents. *Microbiol Spectr* [Internet]. 2021 Oct 31 [cited 2024 Nov 27];9(2). Available from: <https://journals.asm.org/doi/10.1128/Spectrum.00178-21>

7.10 Supplementary data

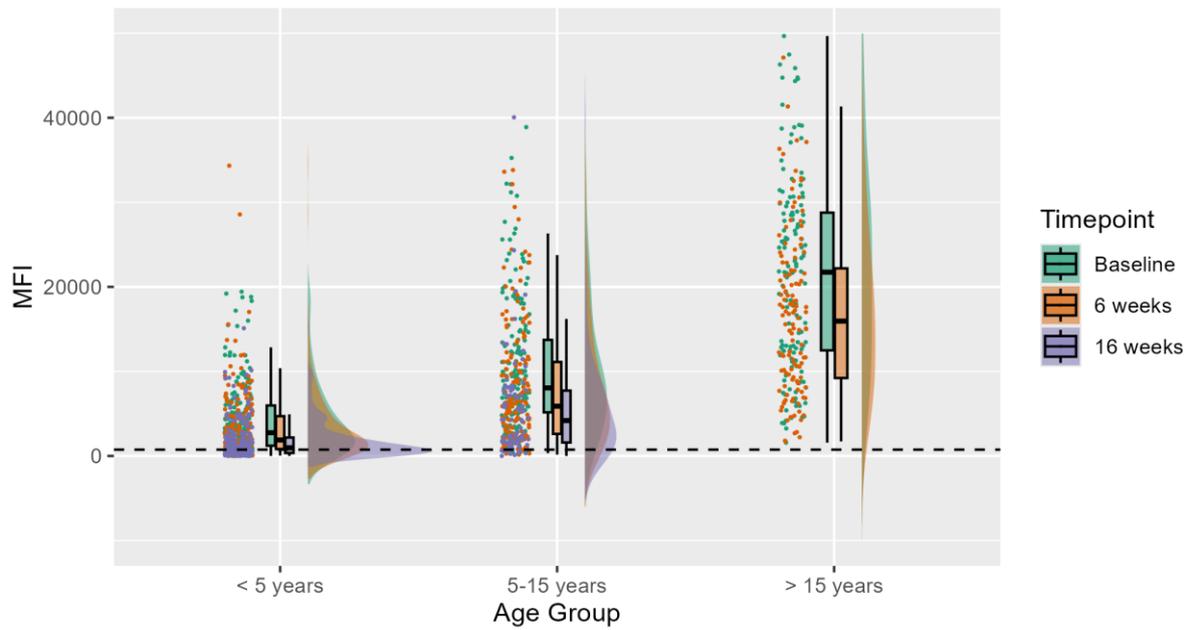


Supplementary Figure 1.1: *PfMSP1.19* raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.

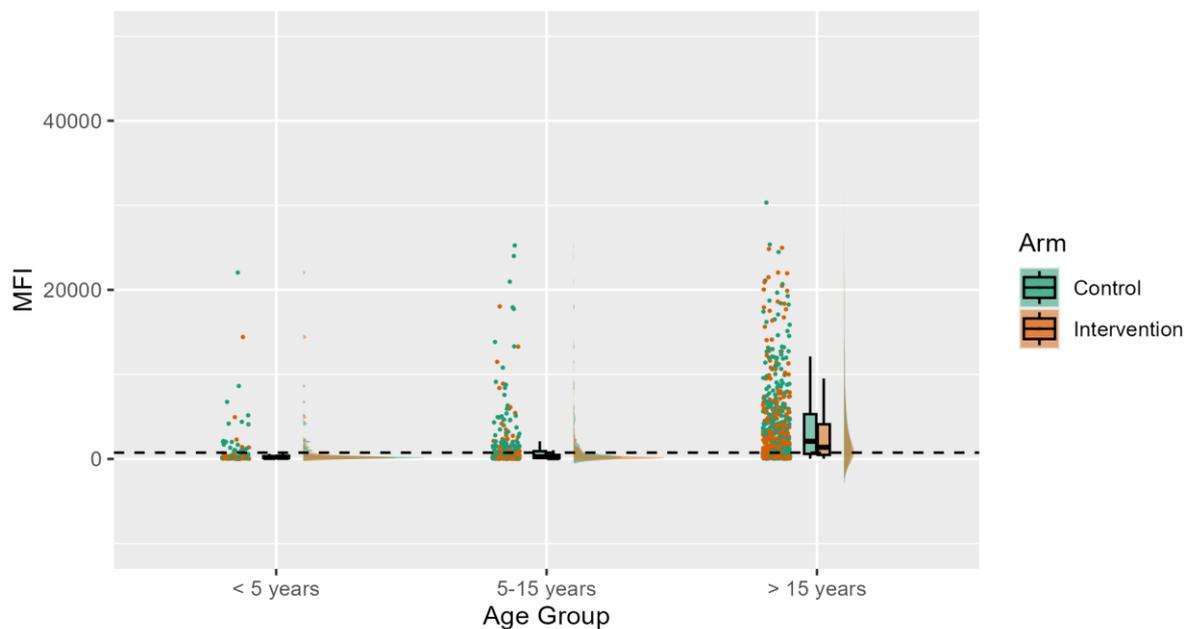


Supplementary Figure 1.2: Etramp5.Ag1 raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.

A) CSP (FIGHTMAL)

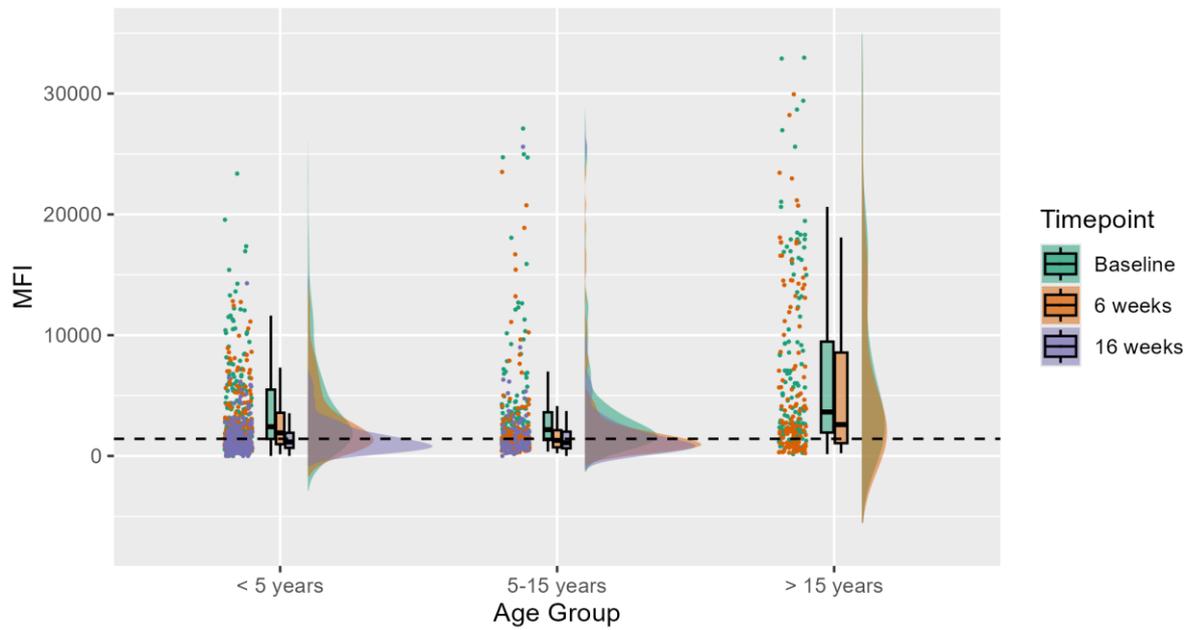


B) CSP (MASSIV_MATAMAL)

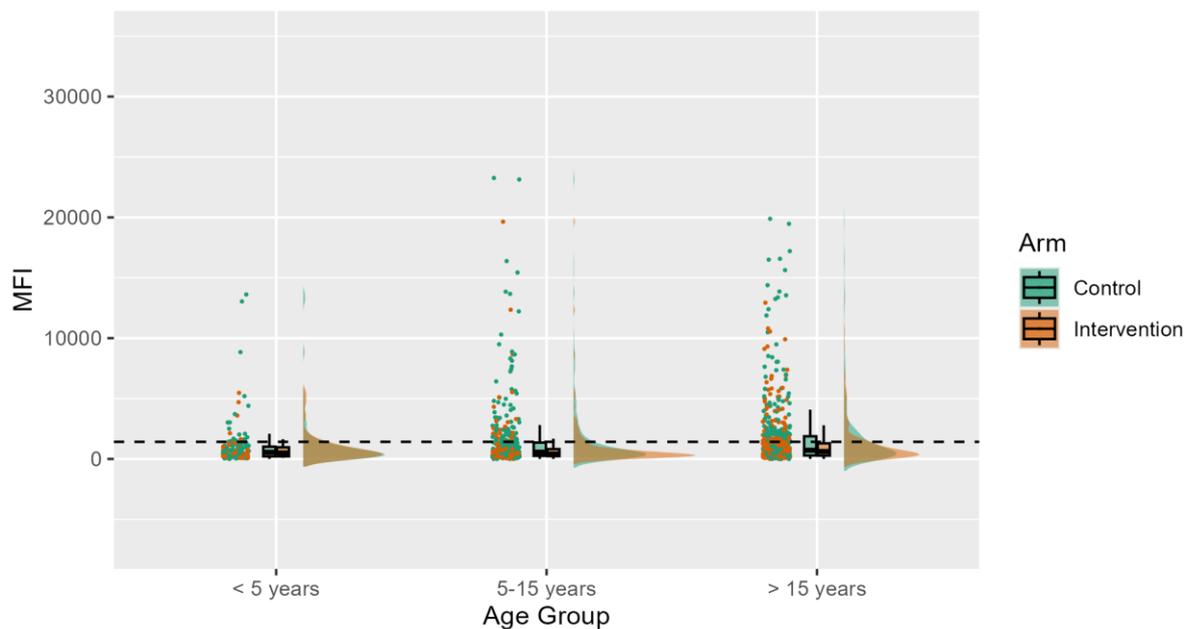


Supplementary Figure 1.3: CSP raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.

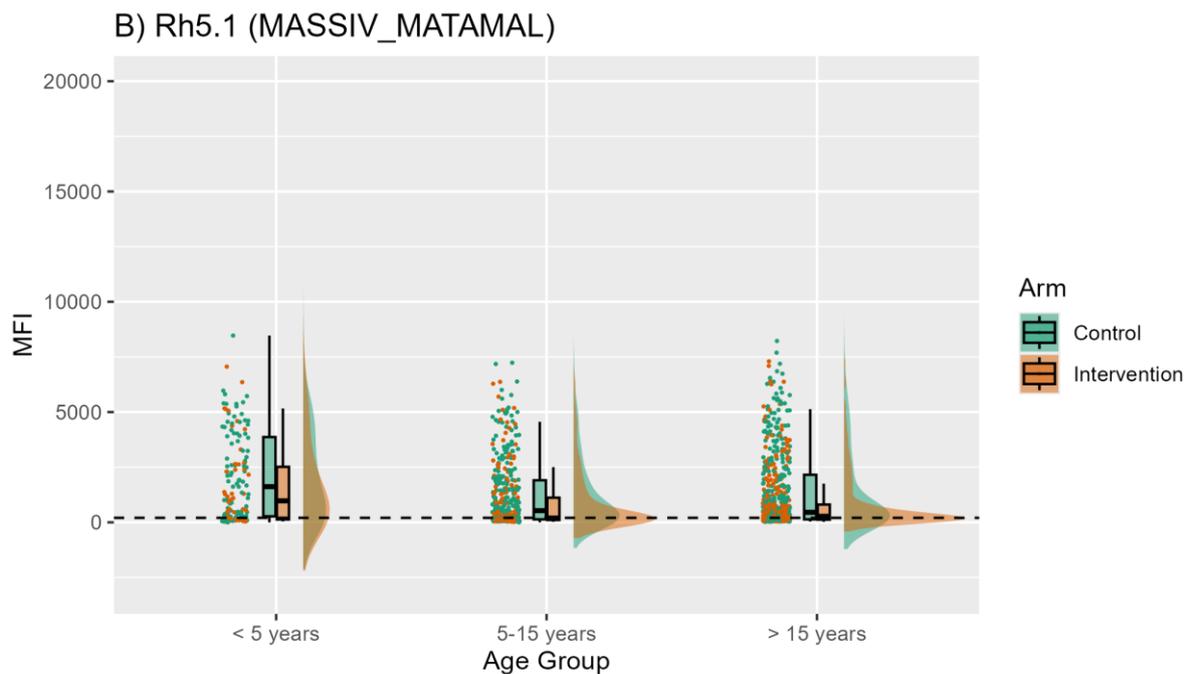
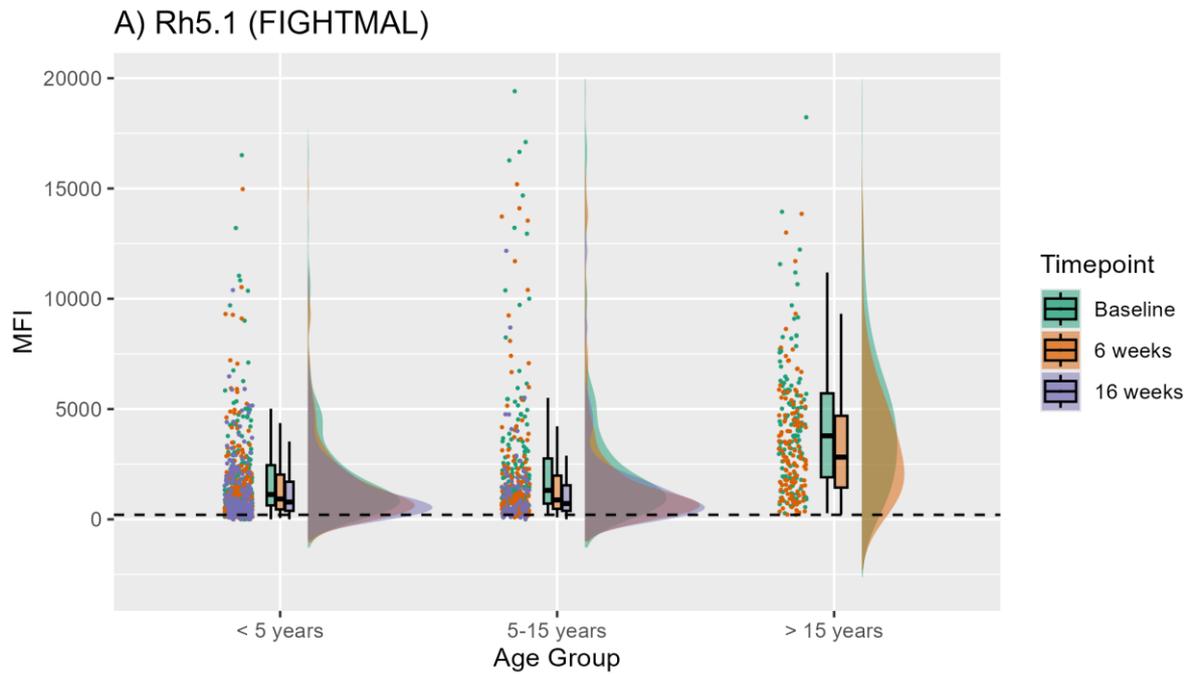
A) HSP40.Ag1 (FIGHTMAL)



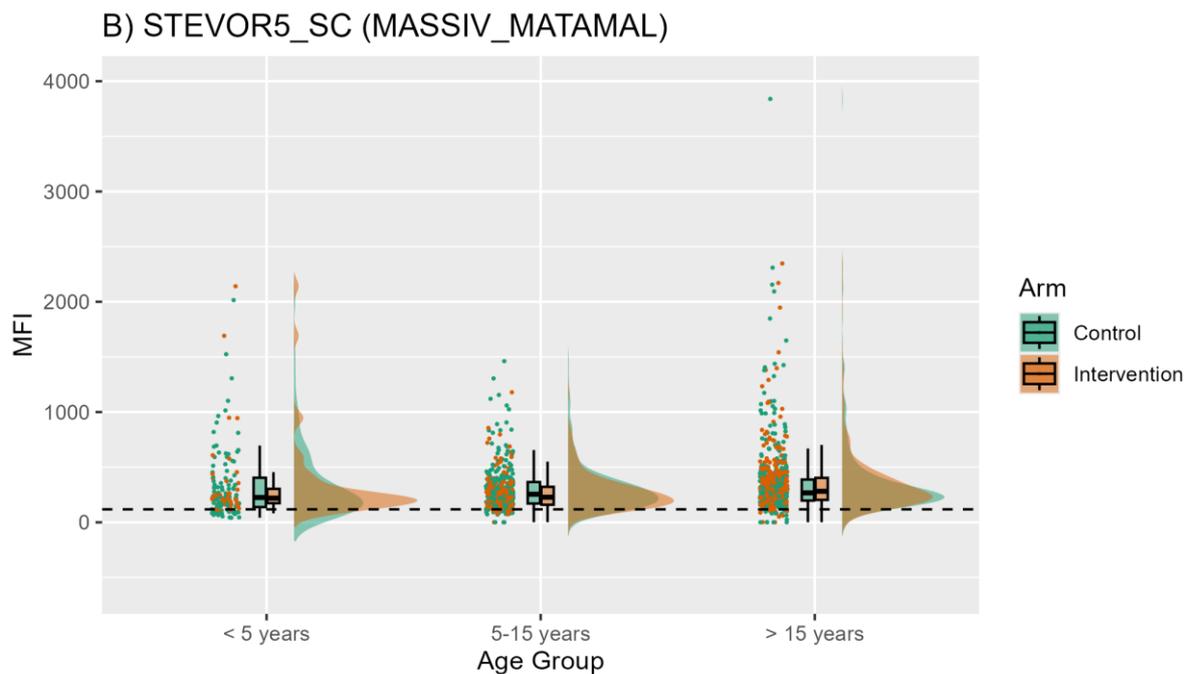
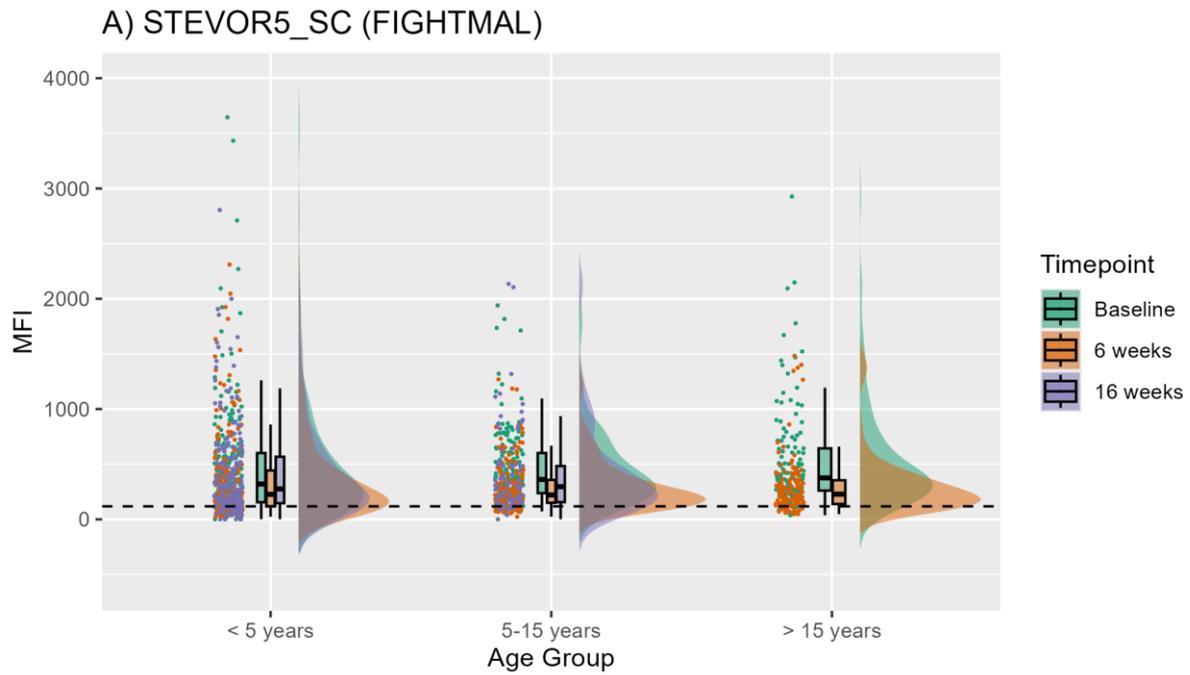
B) HSP40.Ag1 (MASSIV_MATAMAL)



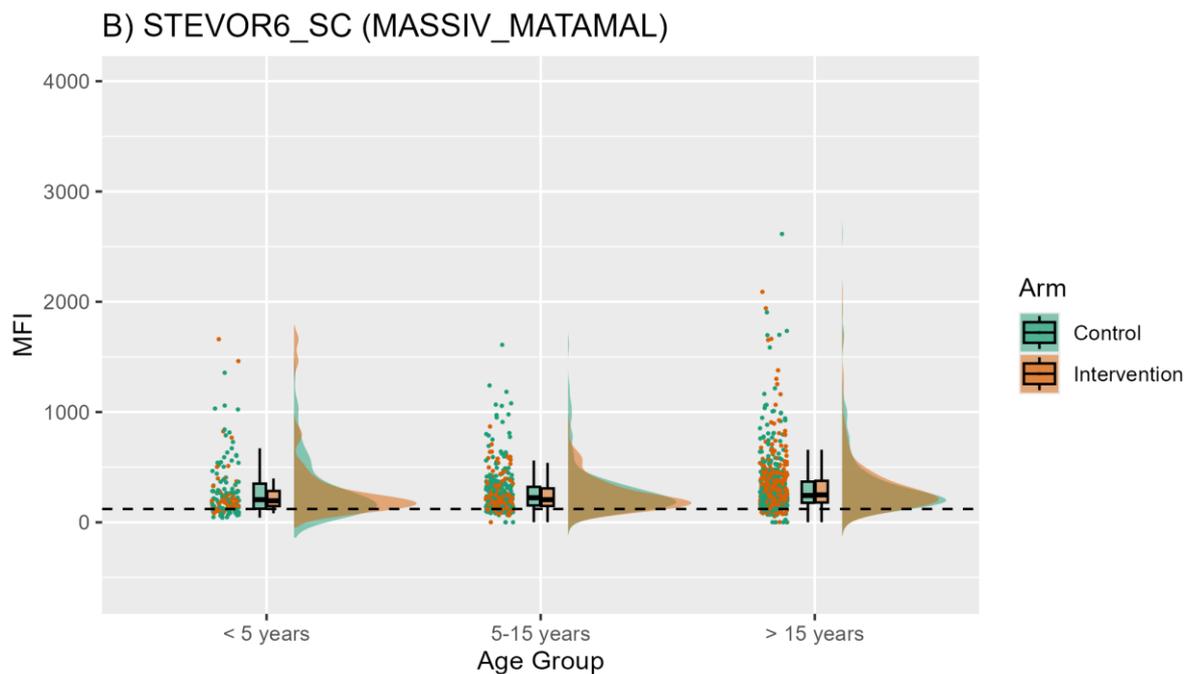
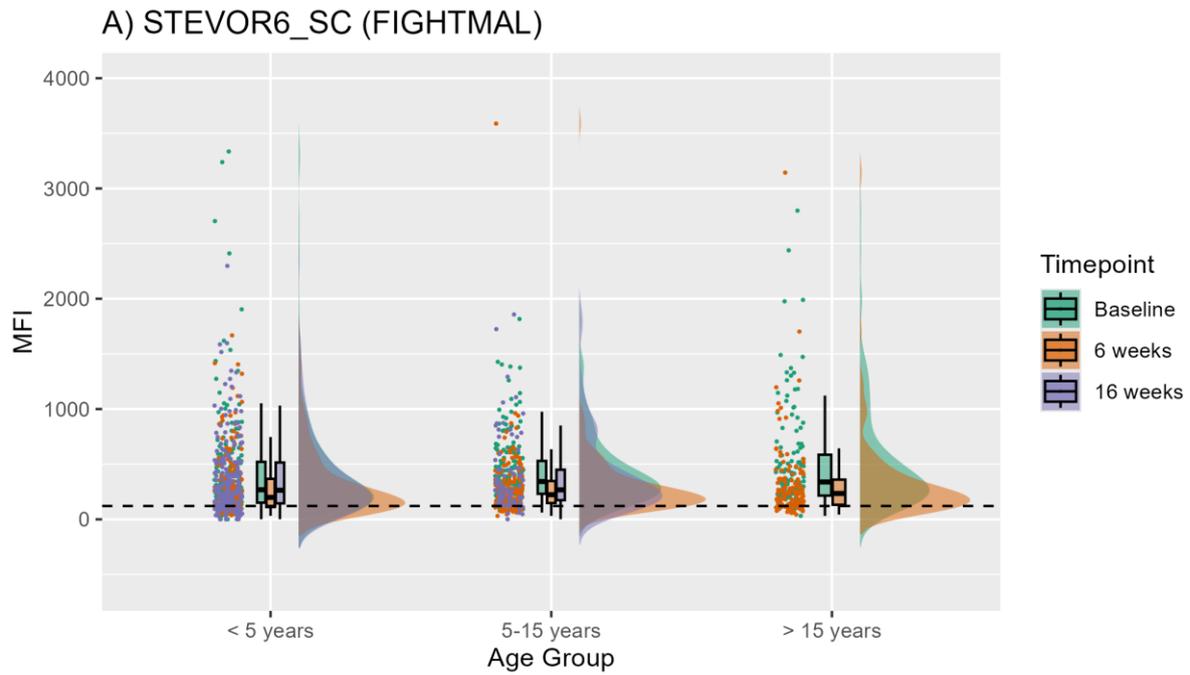
Supplementary Figure 1.4: HSP40.Ag1 raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.



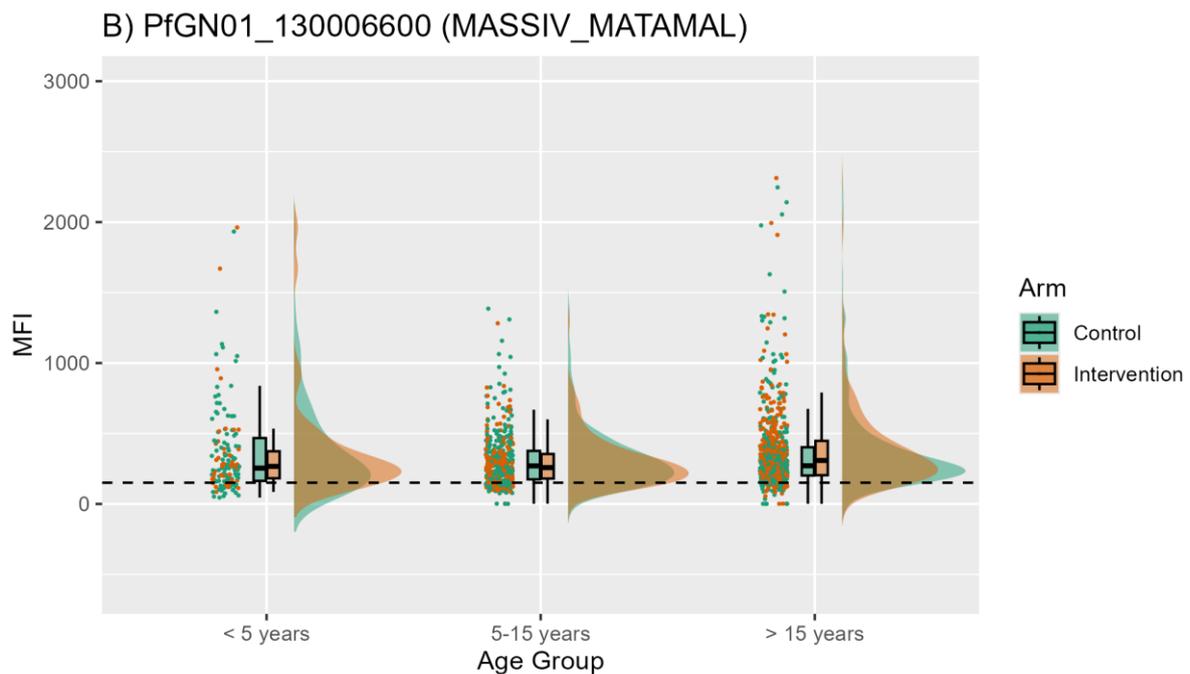
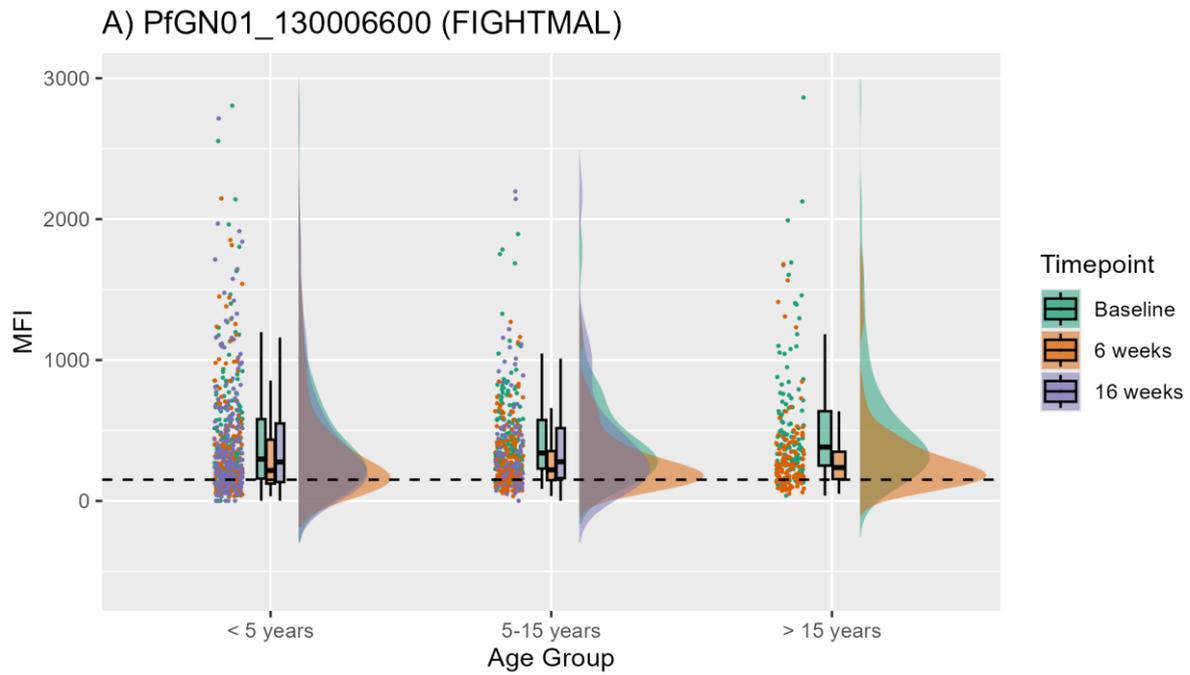
Supplementary Figure 1.5: Rh5.1 raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.



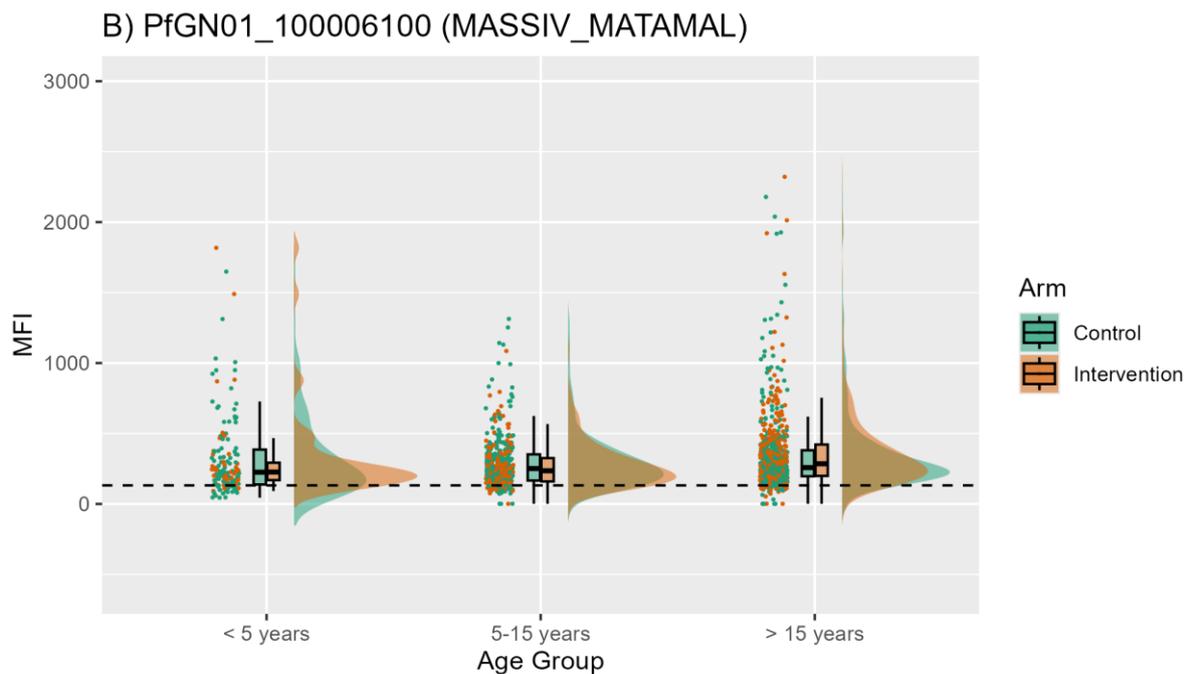
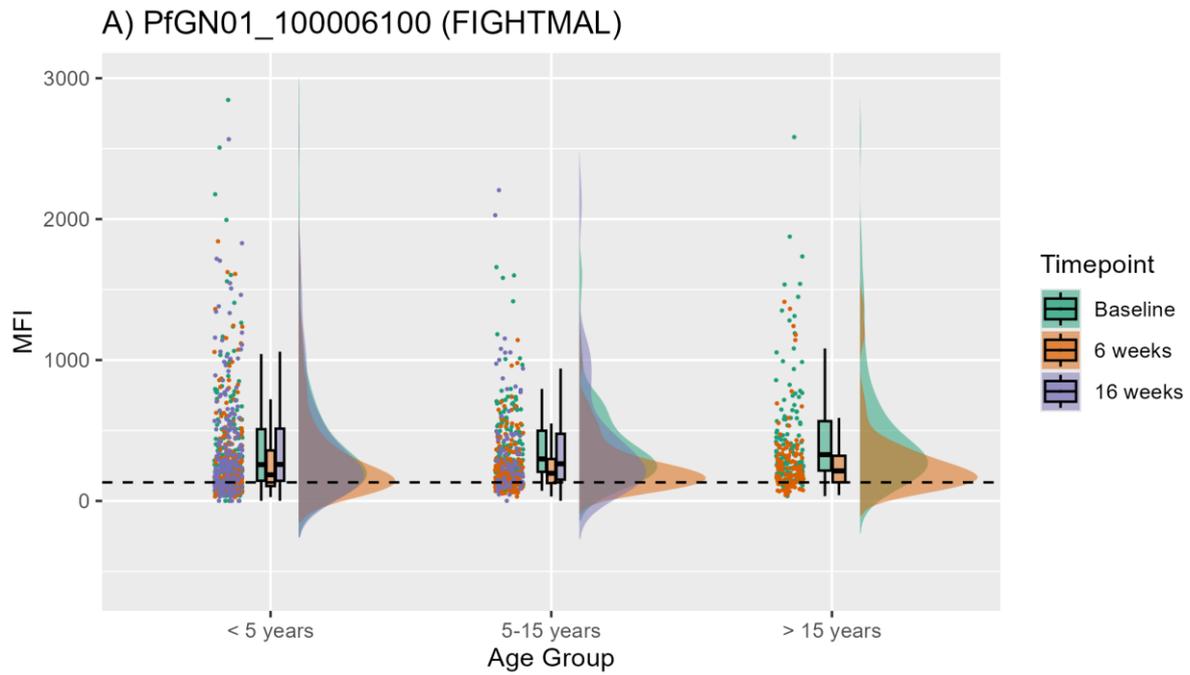
Supplementary Figure 1.6: STEVOR5_SC raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.



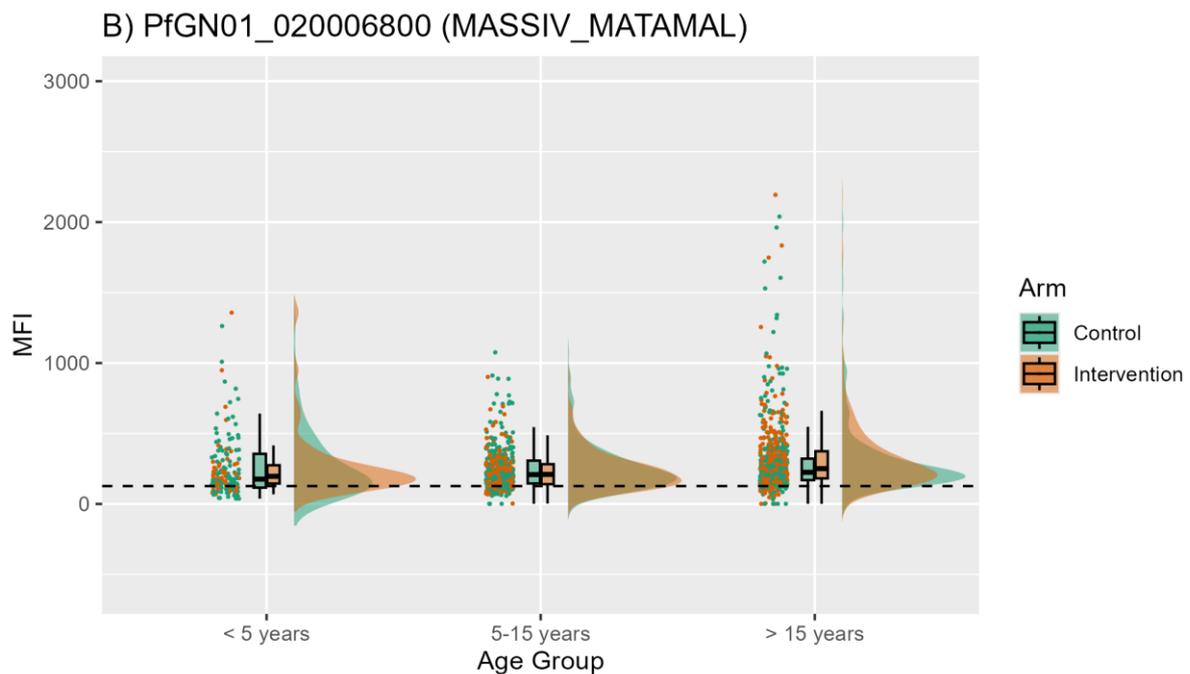
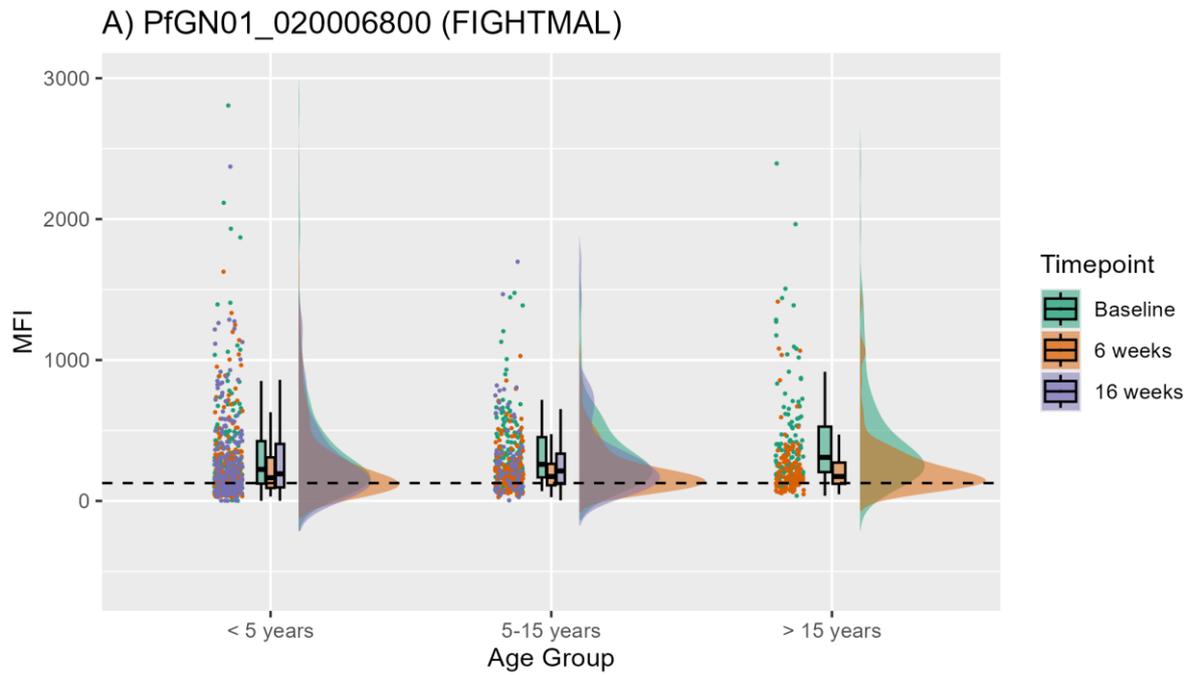
Supplementary Figure 1.7: STEVOR6_SC raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.



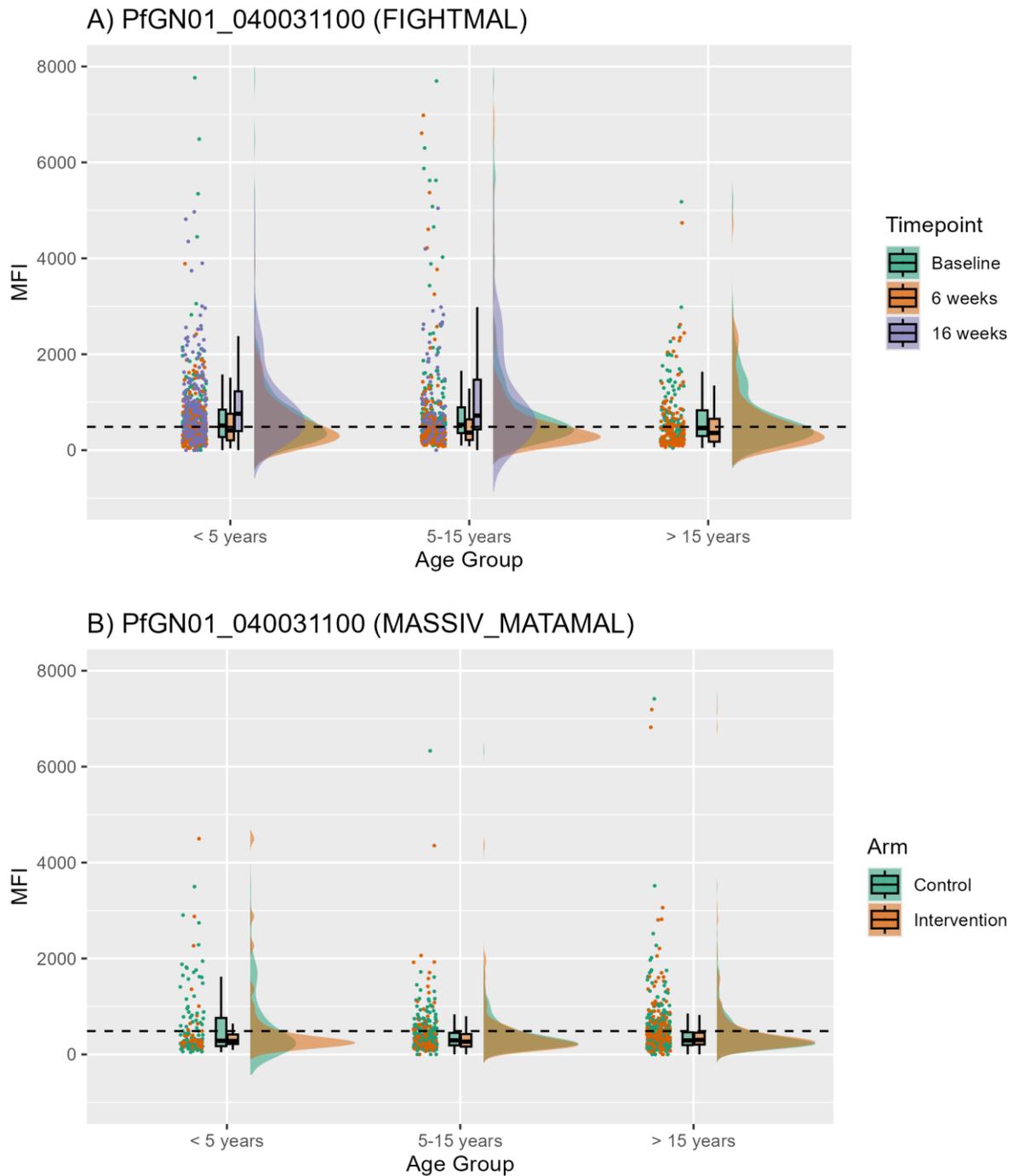
Supplementary Figure 1.8: *PfGN01_130006600* raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.



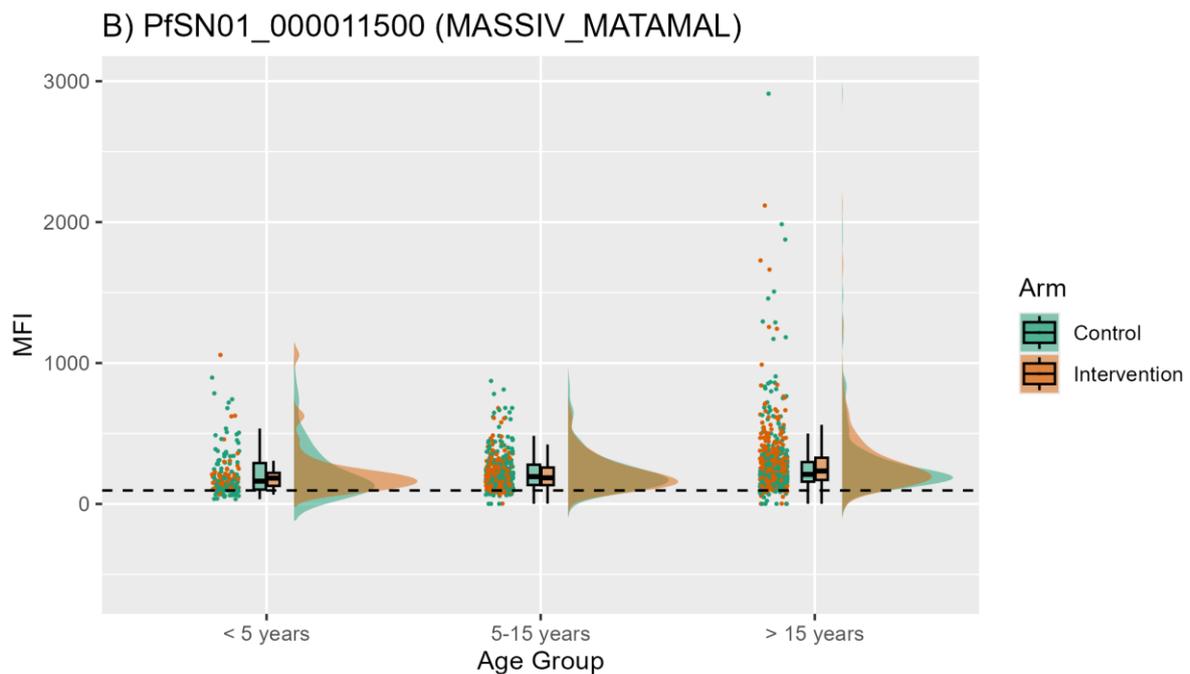
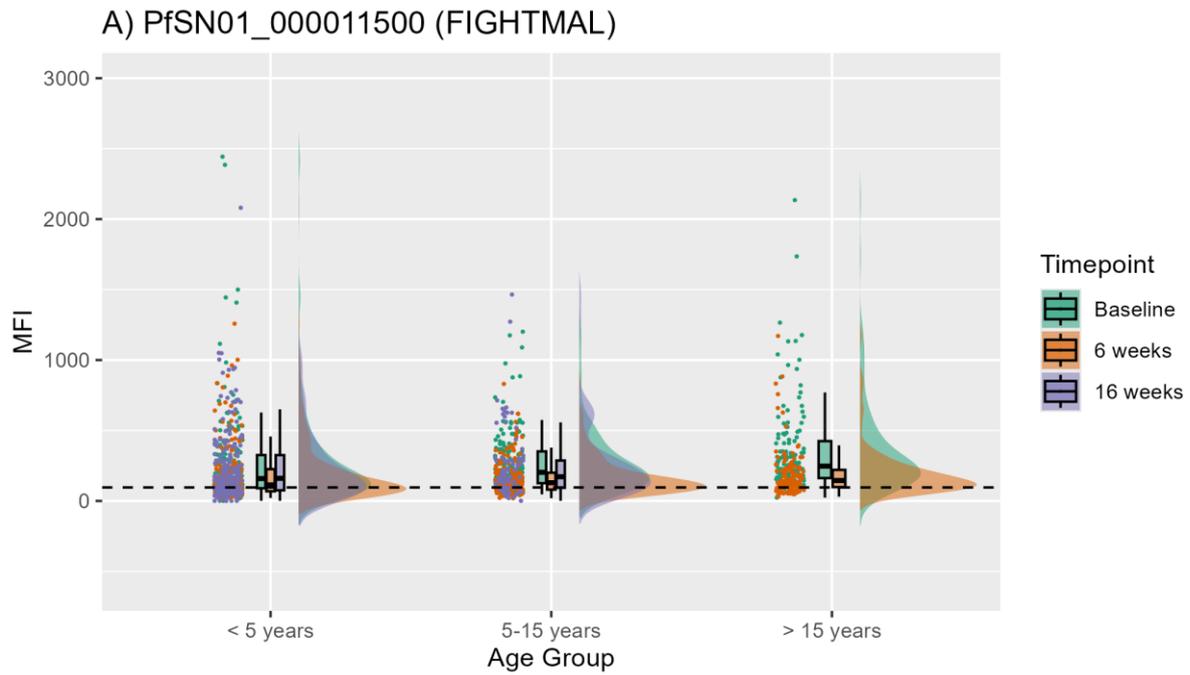
Supplementary Figure 1.9: *PfGN01_100006100* raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.



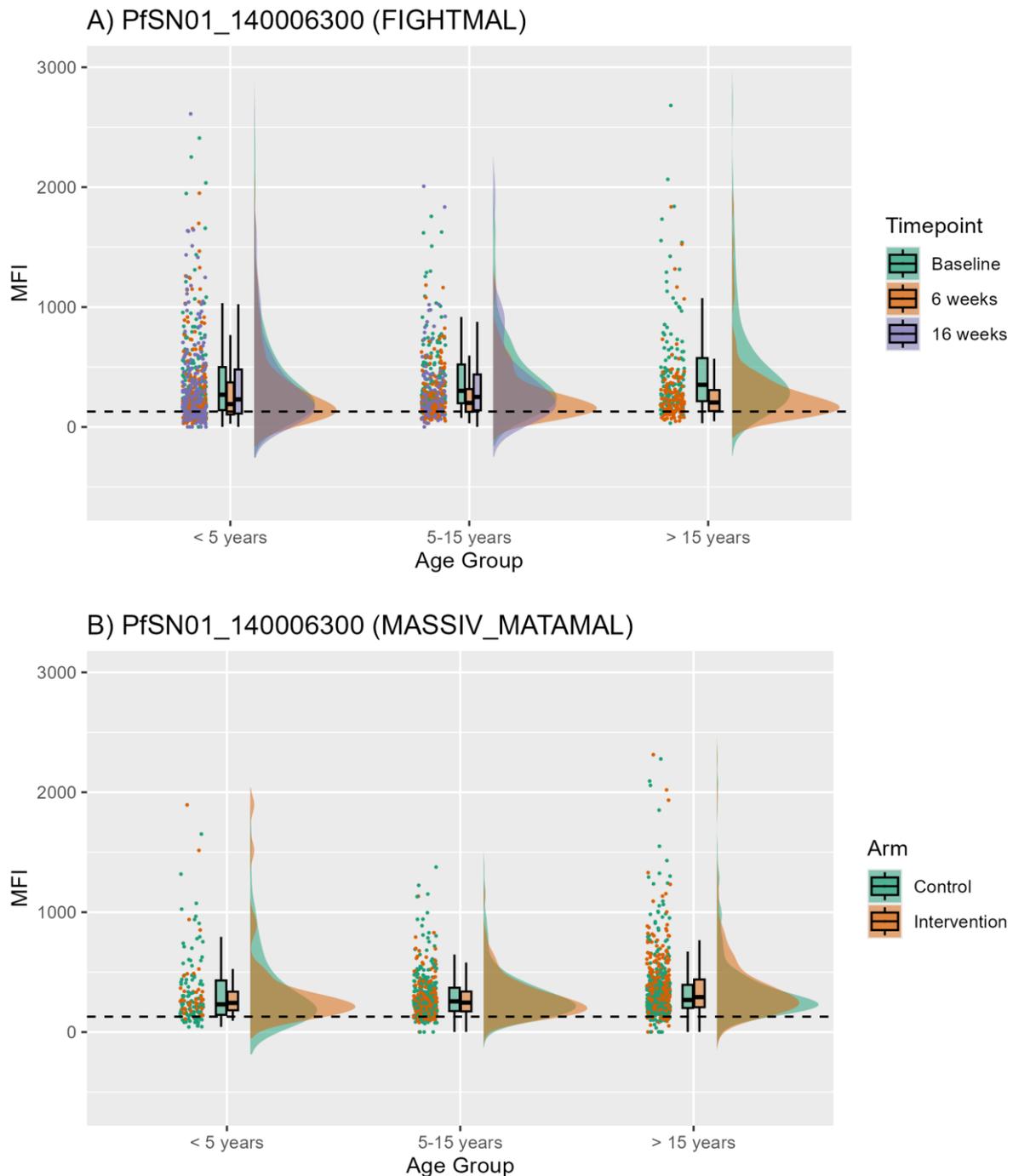
Supplementary Figure 1.10: *PfGN01_020006800* raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.



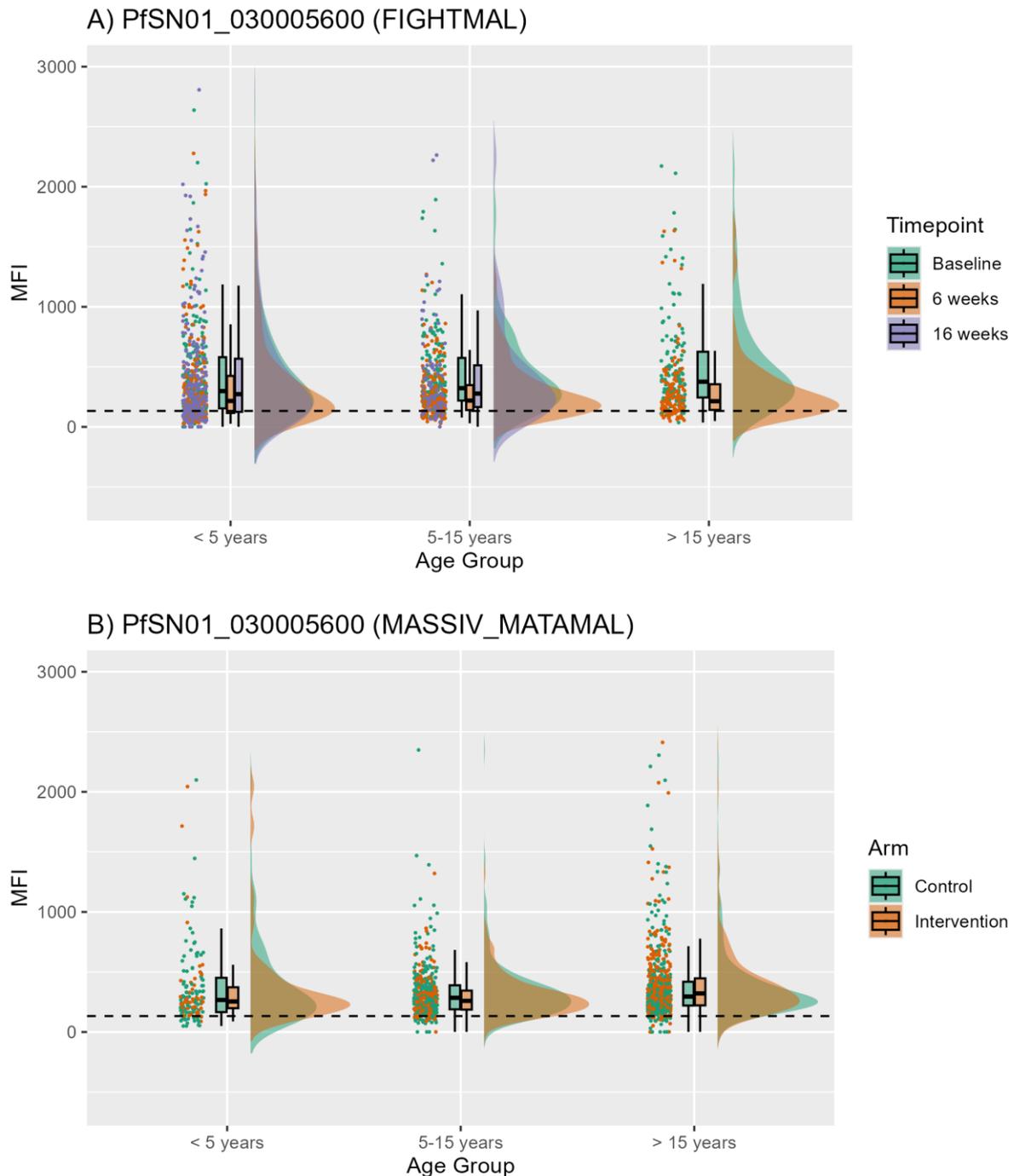
Supplementary Figure 1.11: PfGN01_040031100 raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.



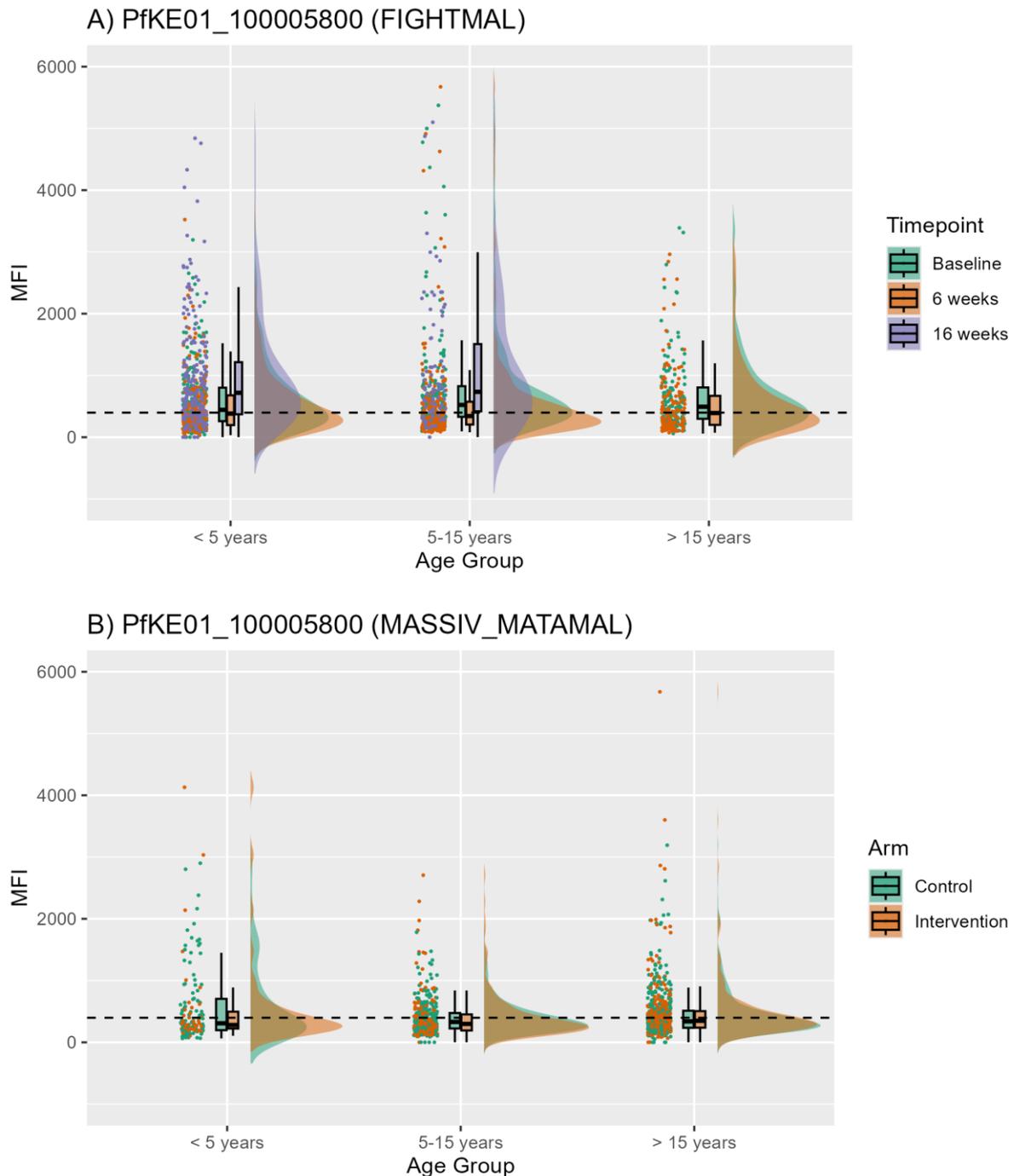
Supplementary Figure 1.12: *Pf*SN01_000011500 raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.



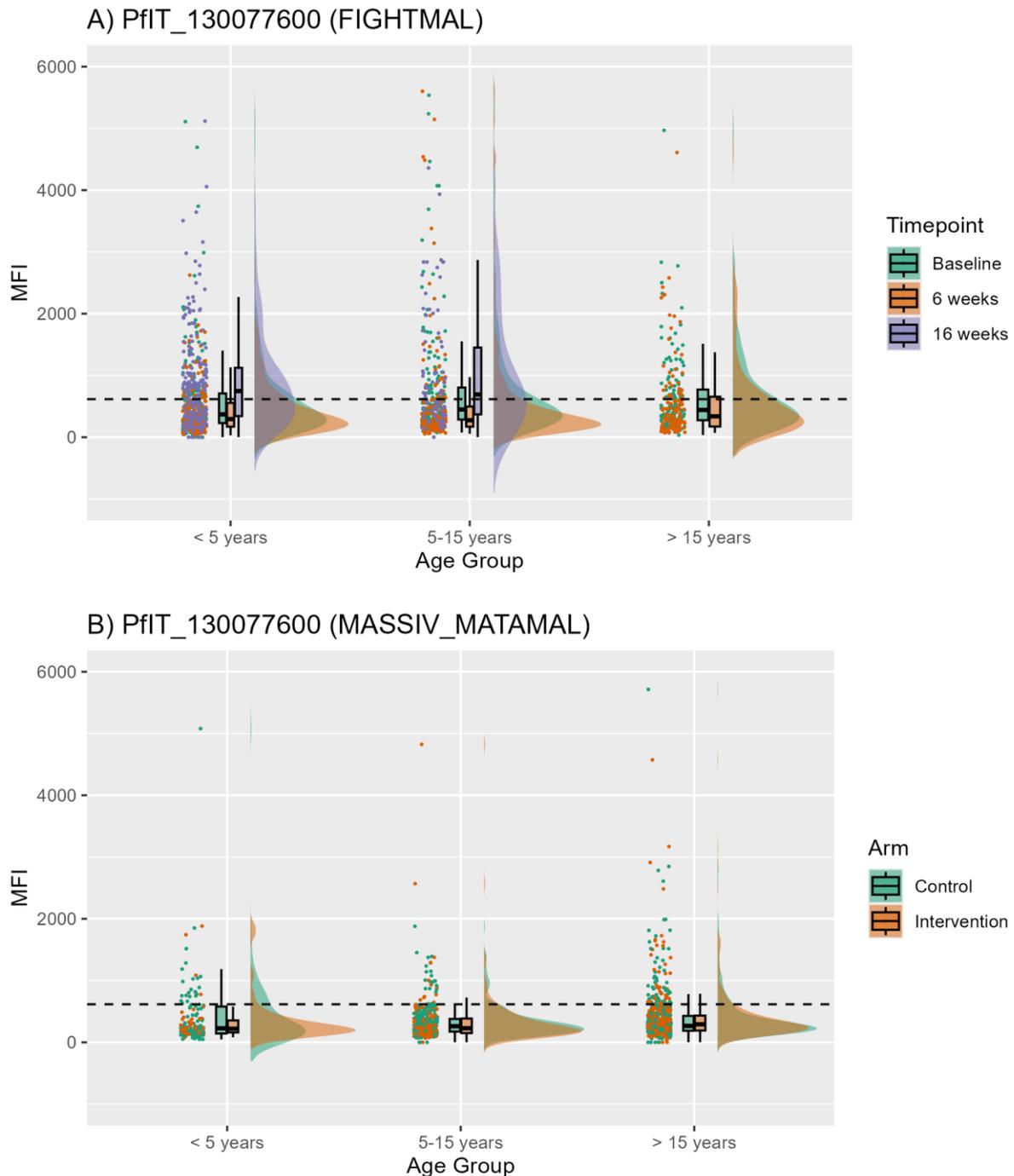
Supplementary Figure 1.13: PfSN01_140006300 raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.



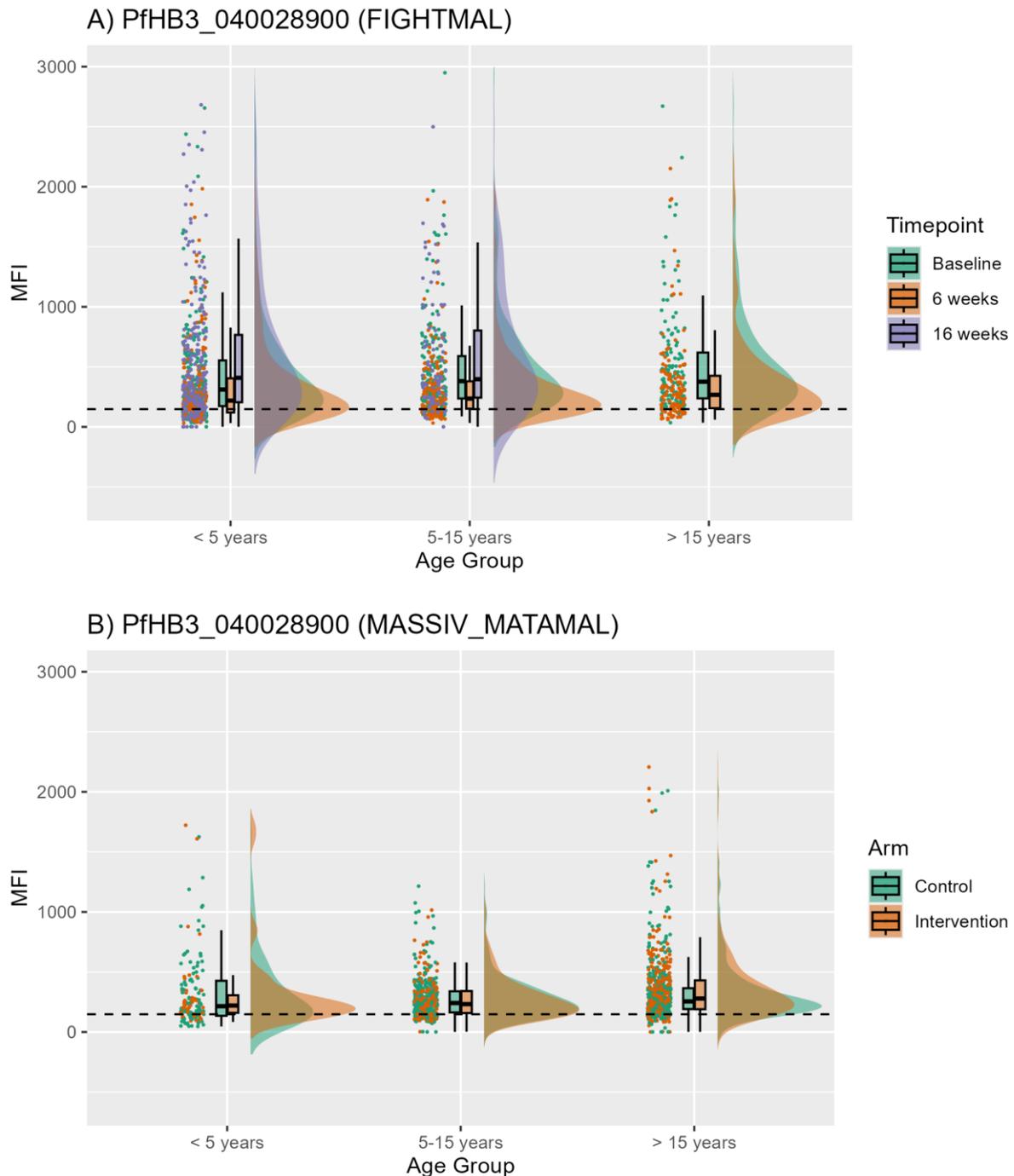
Supplementary Figure 1.14: PfSN01_030005600 raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.



Supplementary Figure 1.15: PfKE01_100005800 raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.

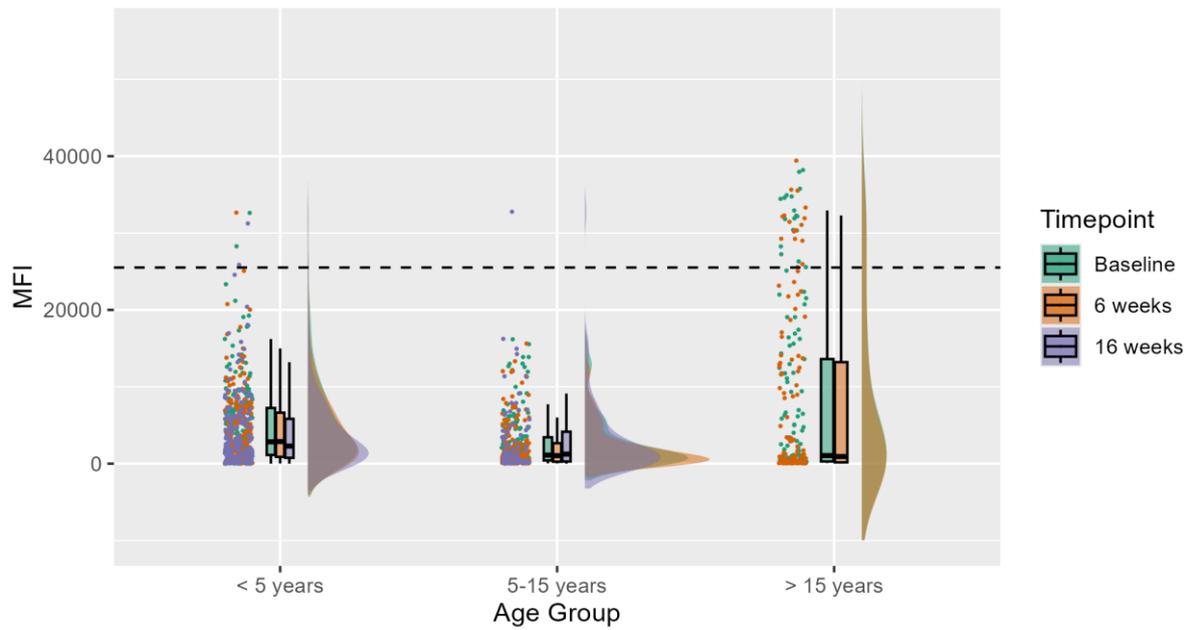


Supplementary Figure 1.16: *PfIT_130077600* raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.

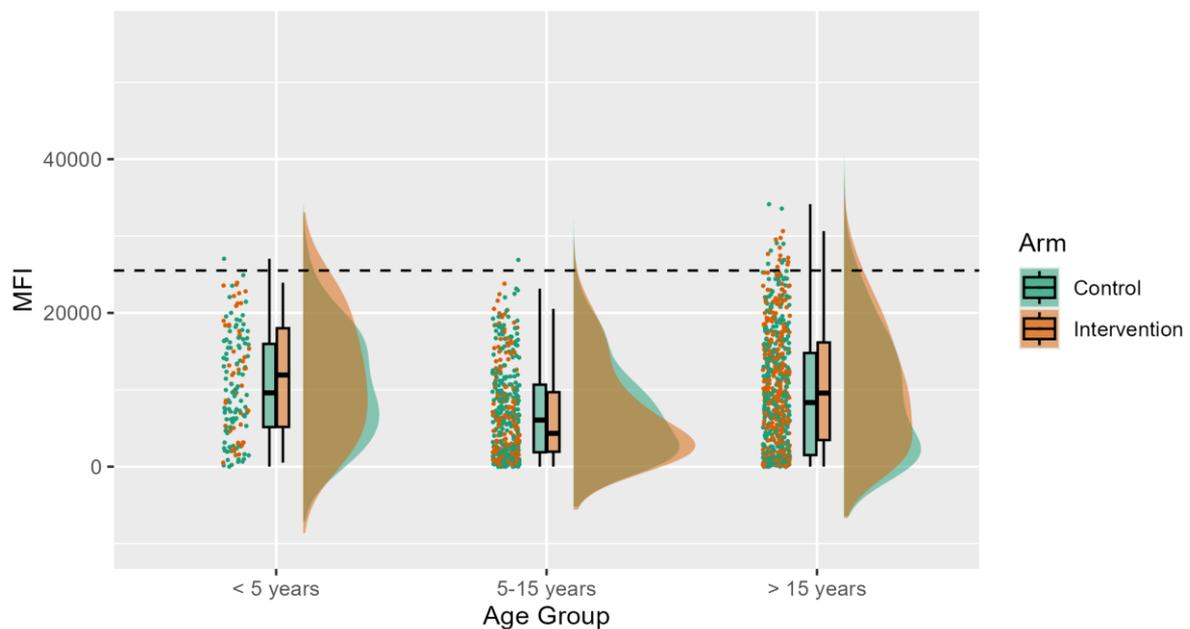


Supplementary Figure 1.17: PfHB3_040028900 raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.

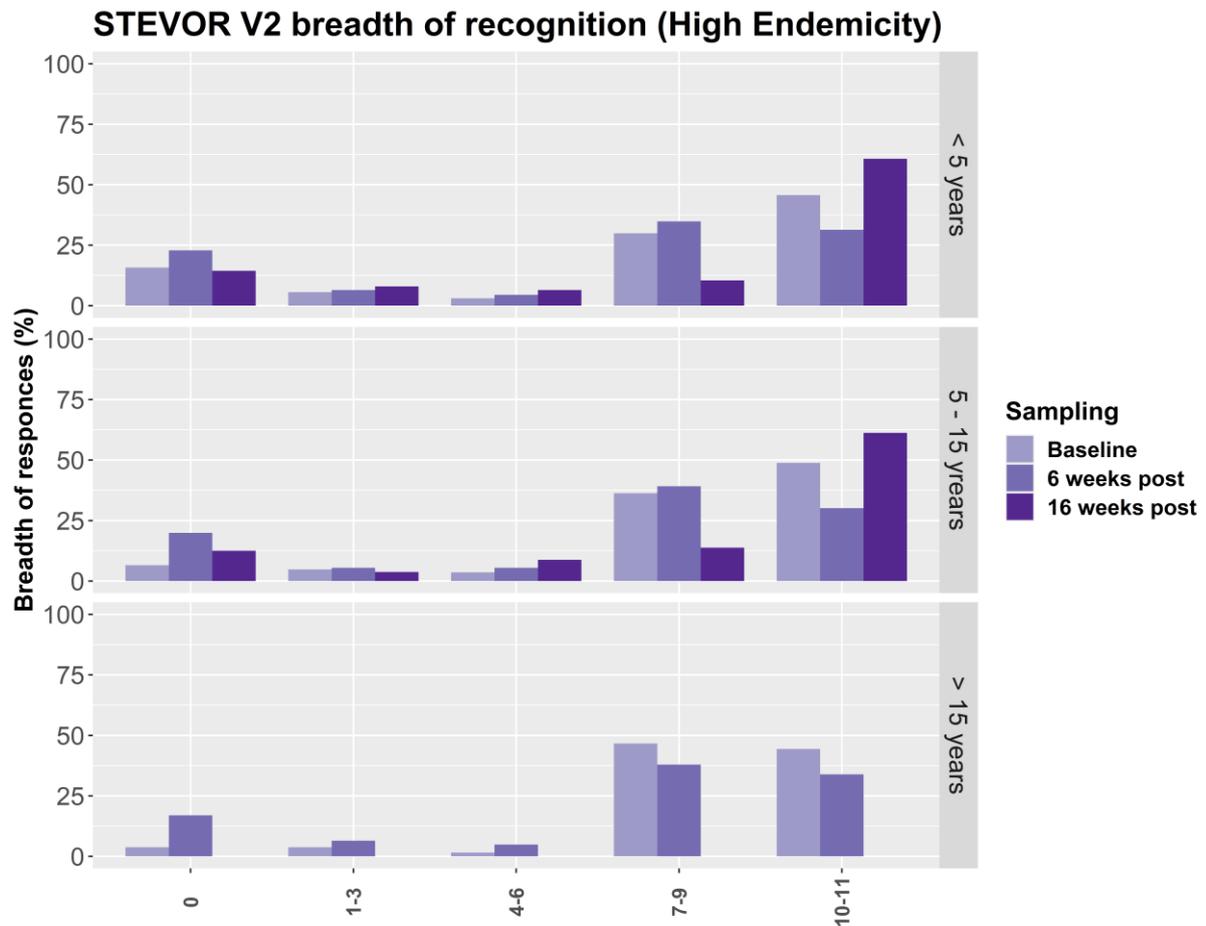
A) Tetanus.toxoid (FIGHTMAL)



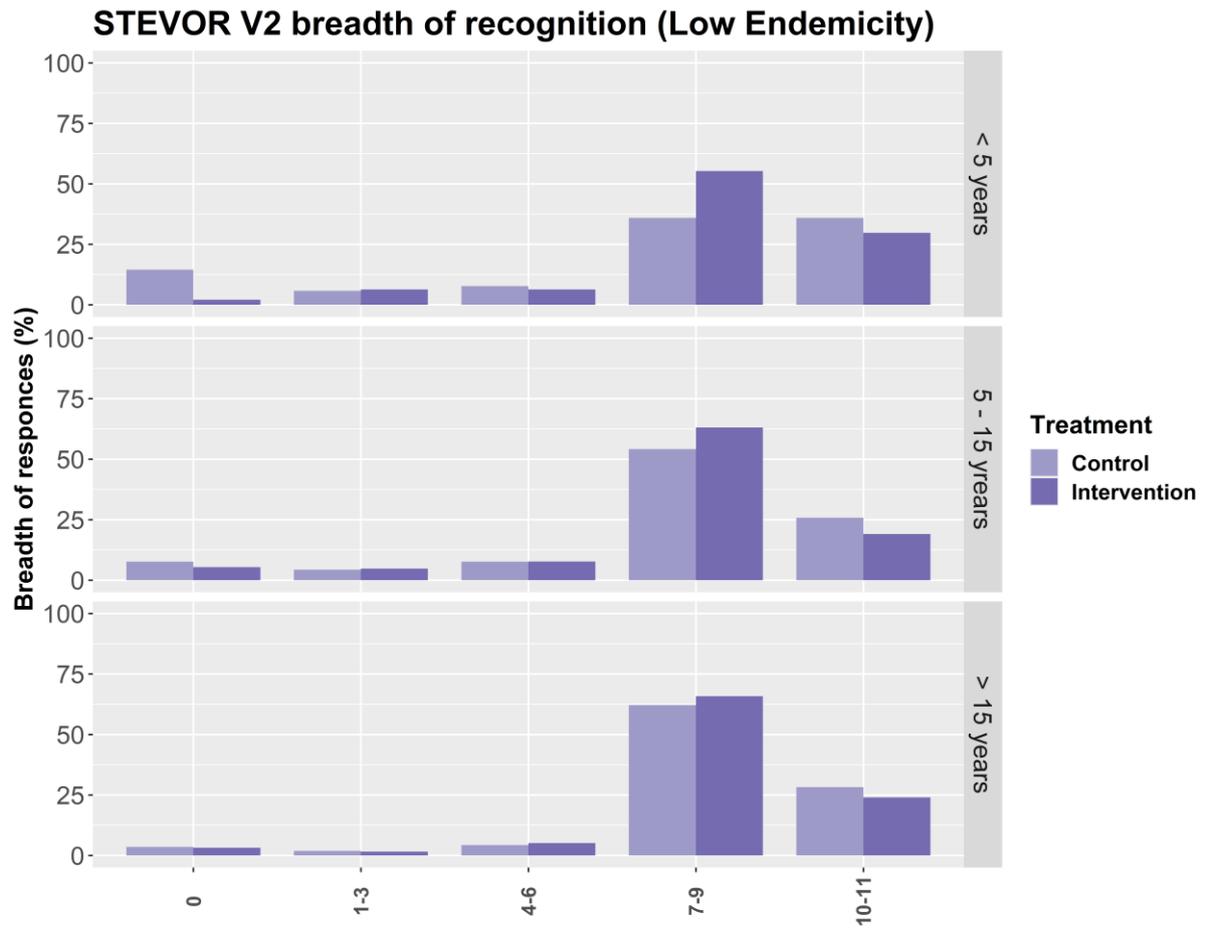
B) Tetanus.toxoid (MASSIV_MATAMAL)



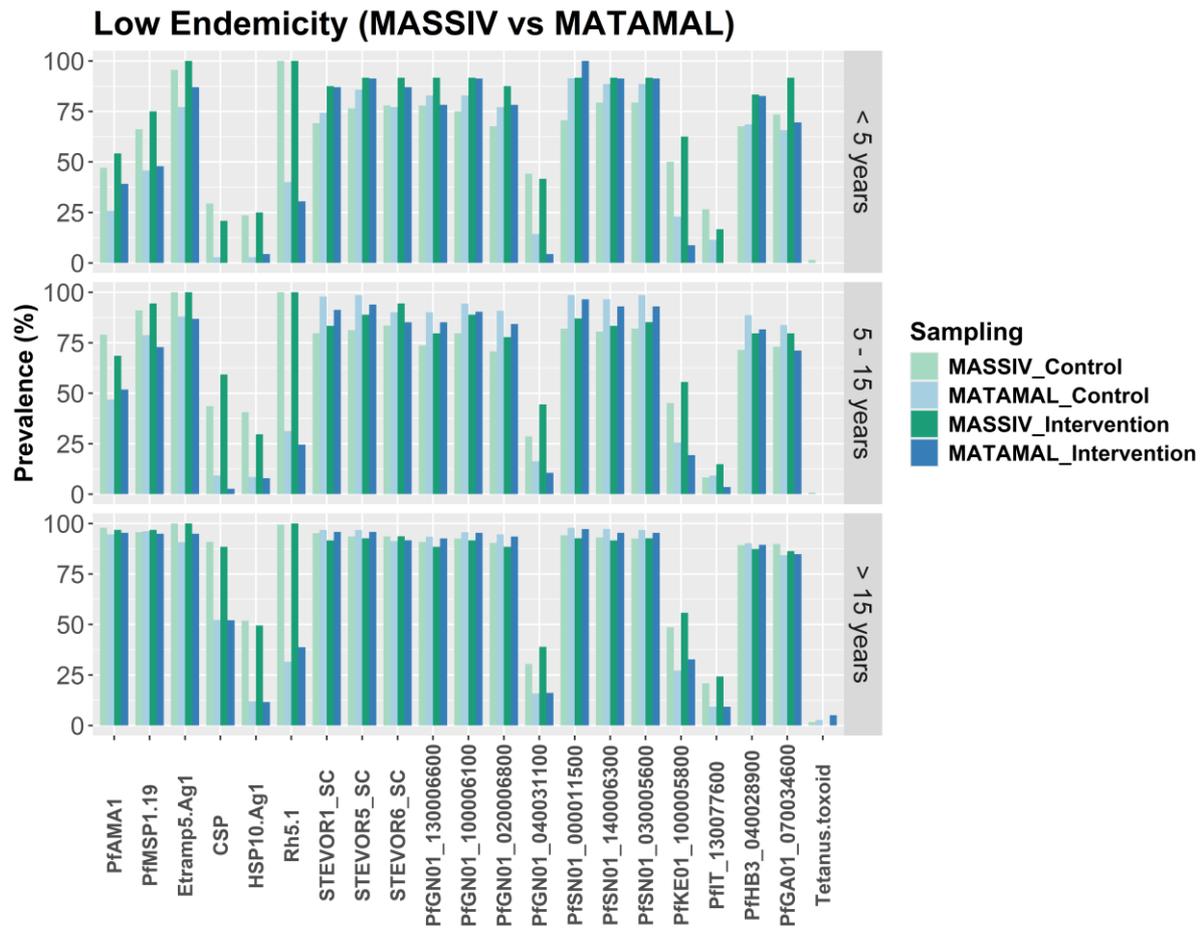
Supplementary Figure 1.18: Tetanus.toxoid raincloud plots of cleaned MFI data for all A) FIGHTMAL samples and all B) MASSIV/MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per A) Timepoint of sampling and B) Treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. All samples above the dashed line are perceived seropositive to the antigen.



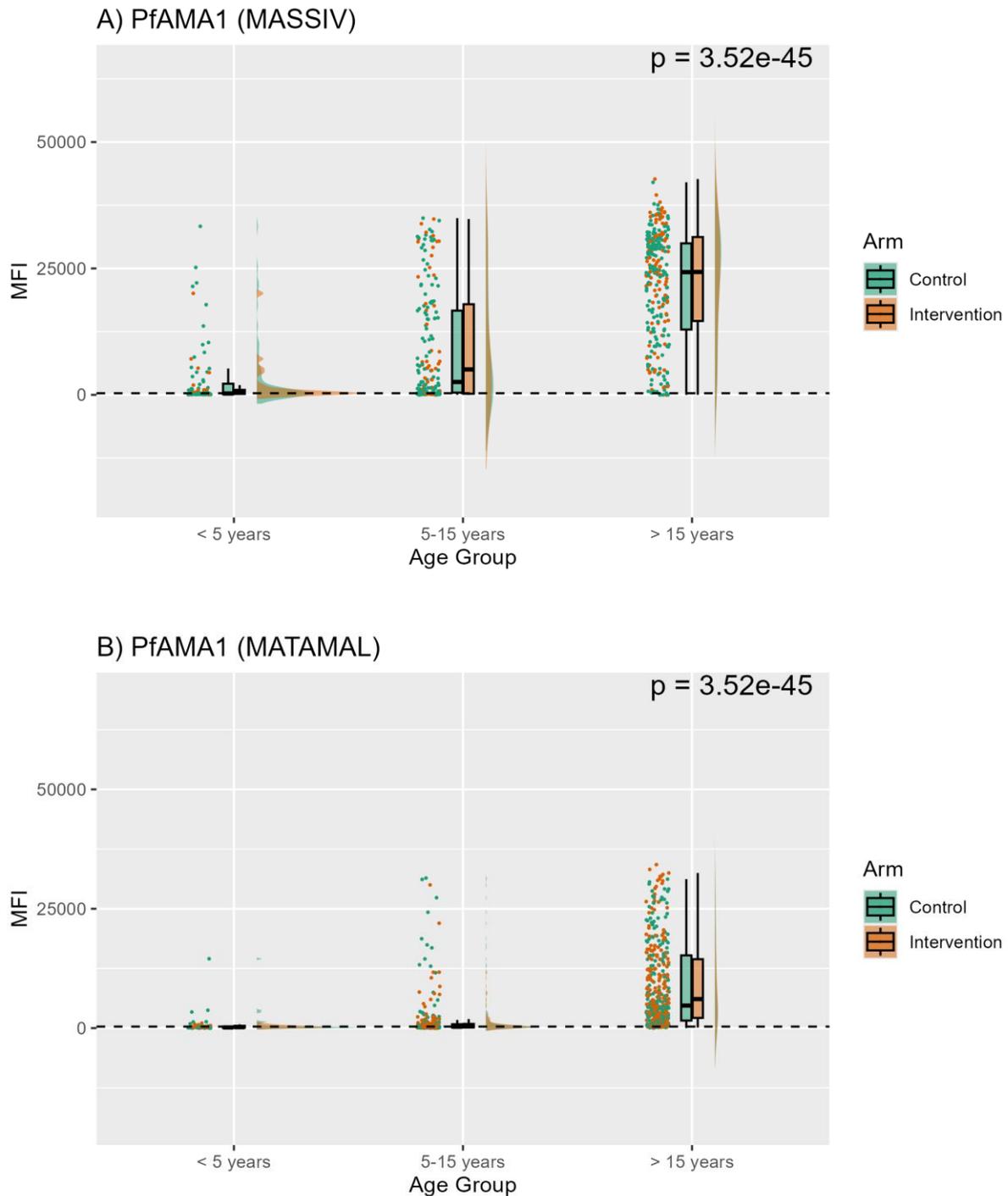
Supplementary Figure 2: FIGHTMAL population prevalence of the breadth of antibody responses to the STEVOR V2 recombinant antigens according to five categories (0, 1-3, 4-6, 7-9, and 10-11 recombinant antigens). Breadth of responses prevalence is stratified by age group (< 5 years, 5-15 years and > 15 years) and sampling timepoint: Baseline (grey), 6 weeks post treatment (light purple) and 16 weeks post treatment (dark purple), indicated in the legend.



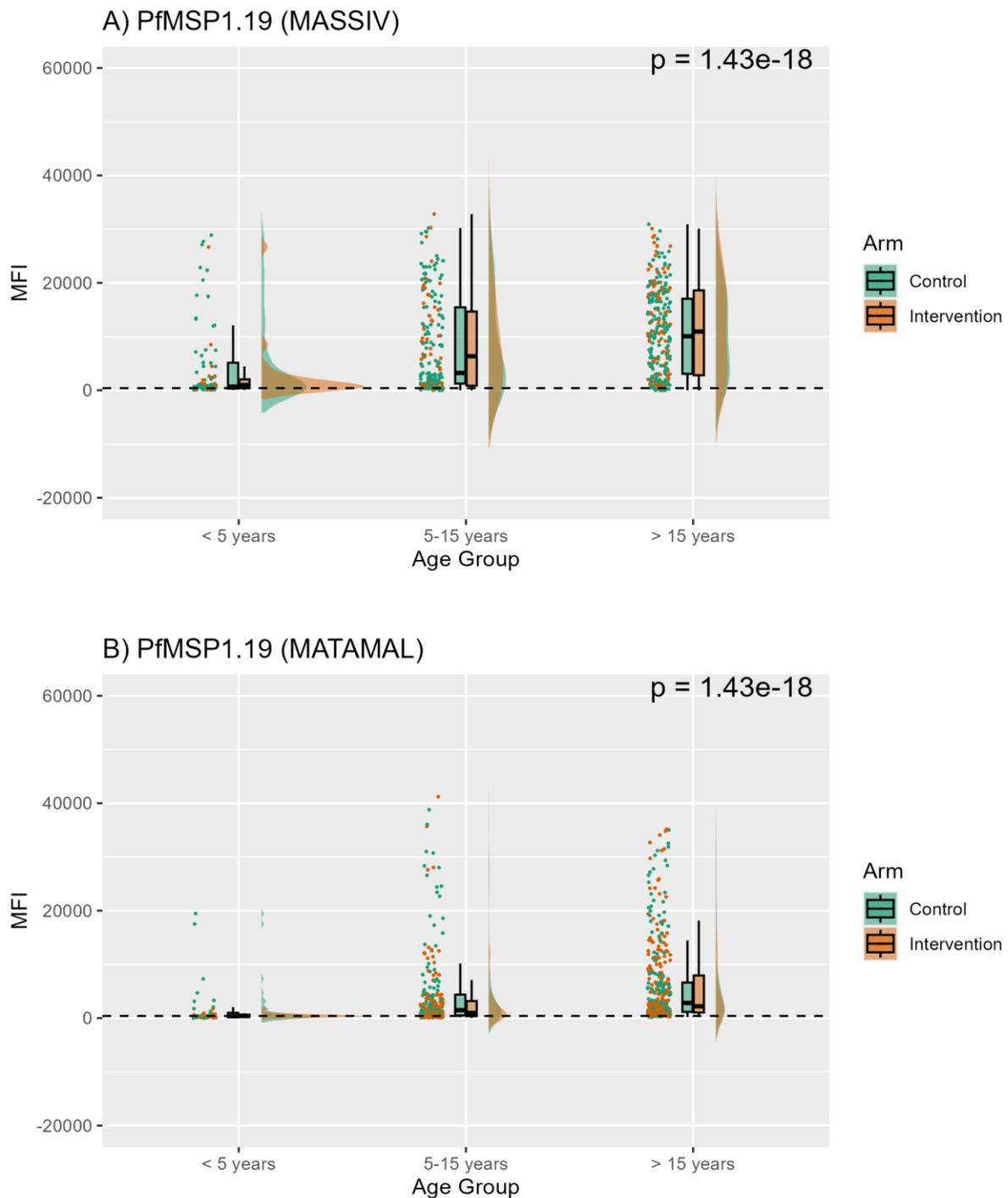
Supplementary Figure 3: MASSIV/MATAMAL pool population prevalence of breadth of antibody responses to the STEVOR V2 recombinant antigens according to five categories (0, 1-3, 4-6, 7-9, and 10-11 recombinant antigens). Breadth of responses prevalence is stratified by age group (< 5 years, 5-15 years and > 15 years) and intervention arms: Control (grey) and Intervention (purple), indicated in the legend.



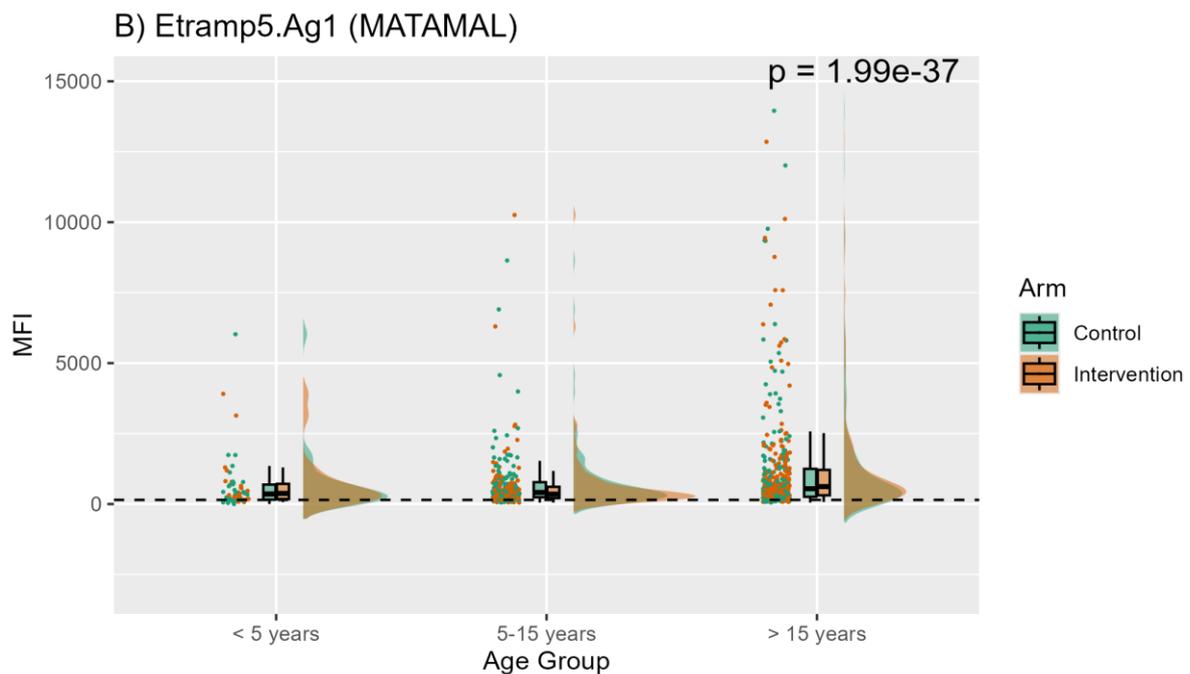
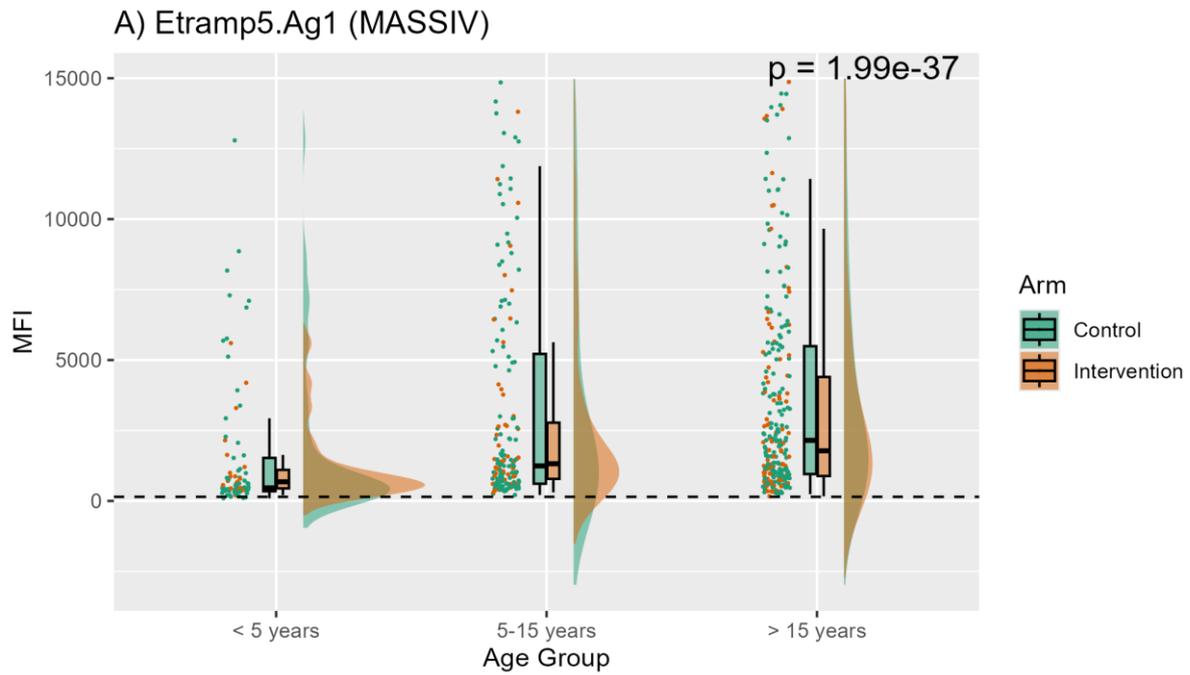
Supplementary Figure 4: Percentage seropositive MASSIV versus MATAMAL samples for the Control (light green and light blue, respectively) and Intervention (green and blue, respectively) treatment arms, stratified by age-group (< 5 years, 5-15 years and > 15 years). Recombinant antigens are shown on the x-axis. STEVOR antigens with the _SC suffix refer to semi-conserved antigen targets. STEVOR antigens *PfGN01_130006600* – *PfGA01_070034600* were all based on the large hypervariable loop. There was weak evidence for significant differences of seropositivity values between MASSIV and MATAMAL with a p-value of 0.03 according to Man-Whitney non-parametric test.



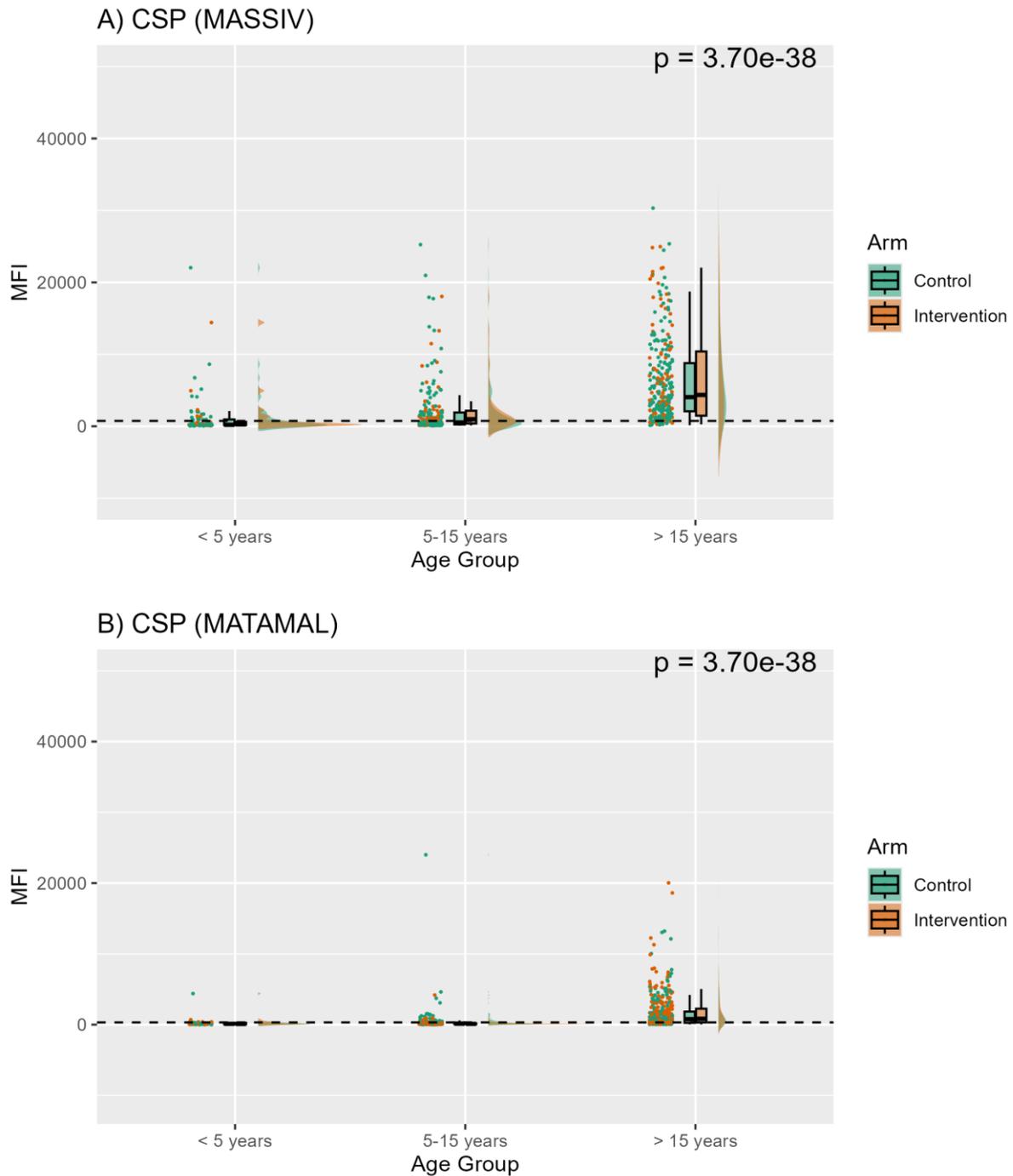
Supplementary Figure 5.1: PfAMA1 raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.



Supplementary Figure 5.2: PfMSP1.9 raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.

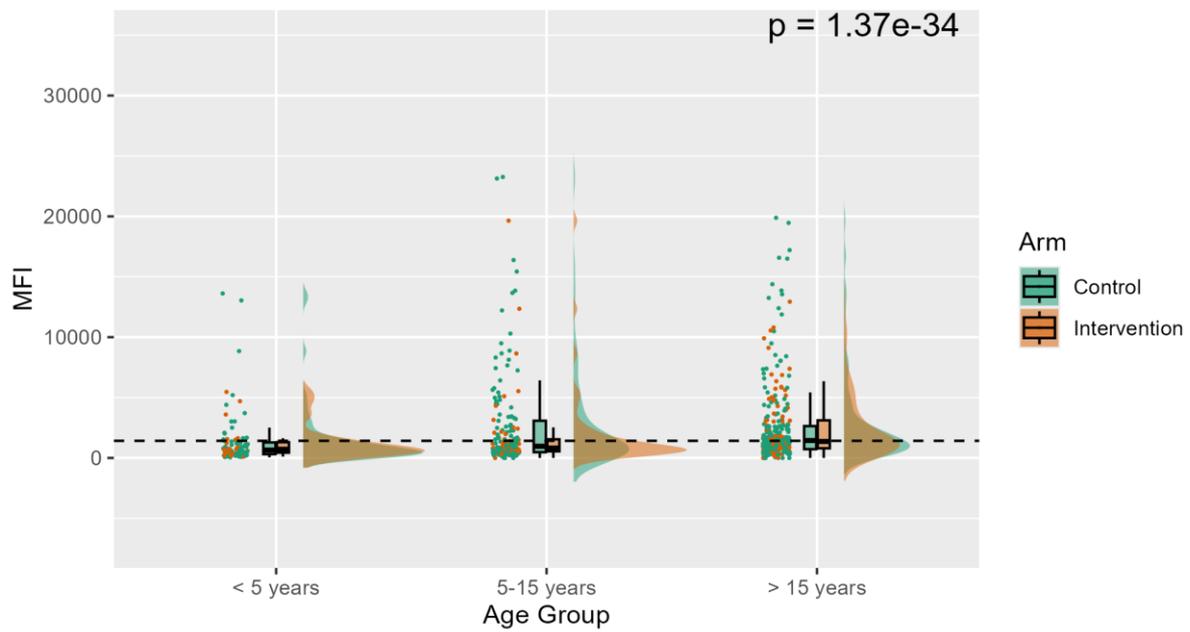


Supplementary Figure 5.3: *Pf*Etramp5.Ag1 raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.

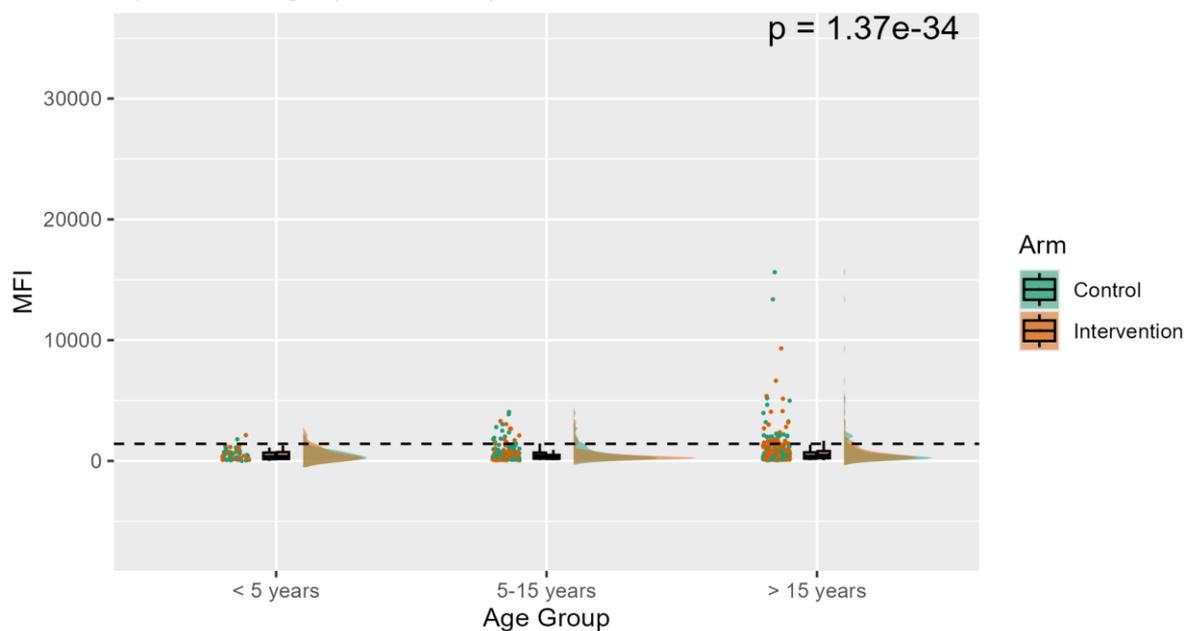


Supplementary Figure 5.4: *Pf*CSP raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.

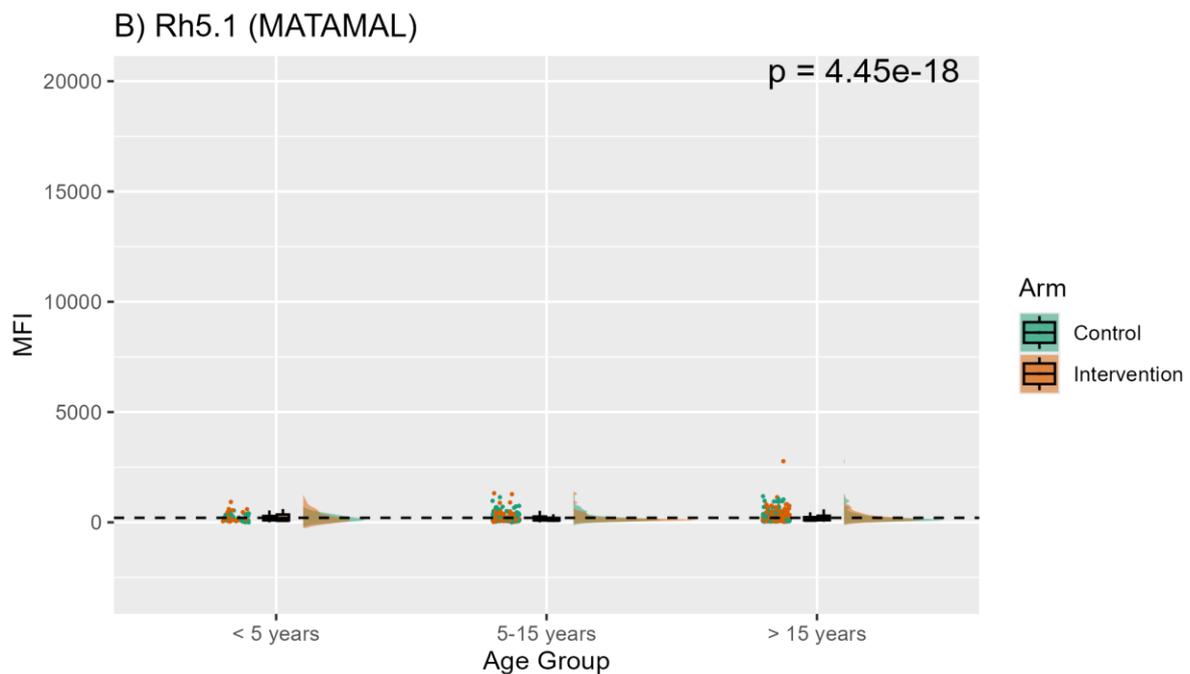
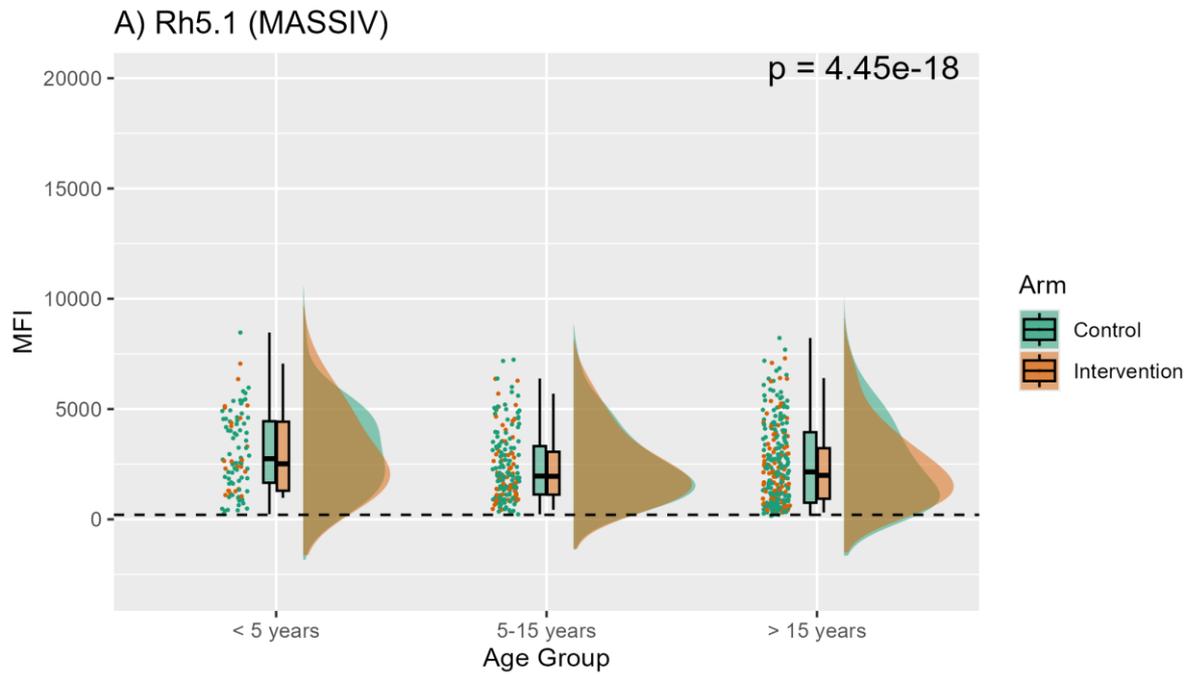
A) HSP40.Ag1 (MASSIV)



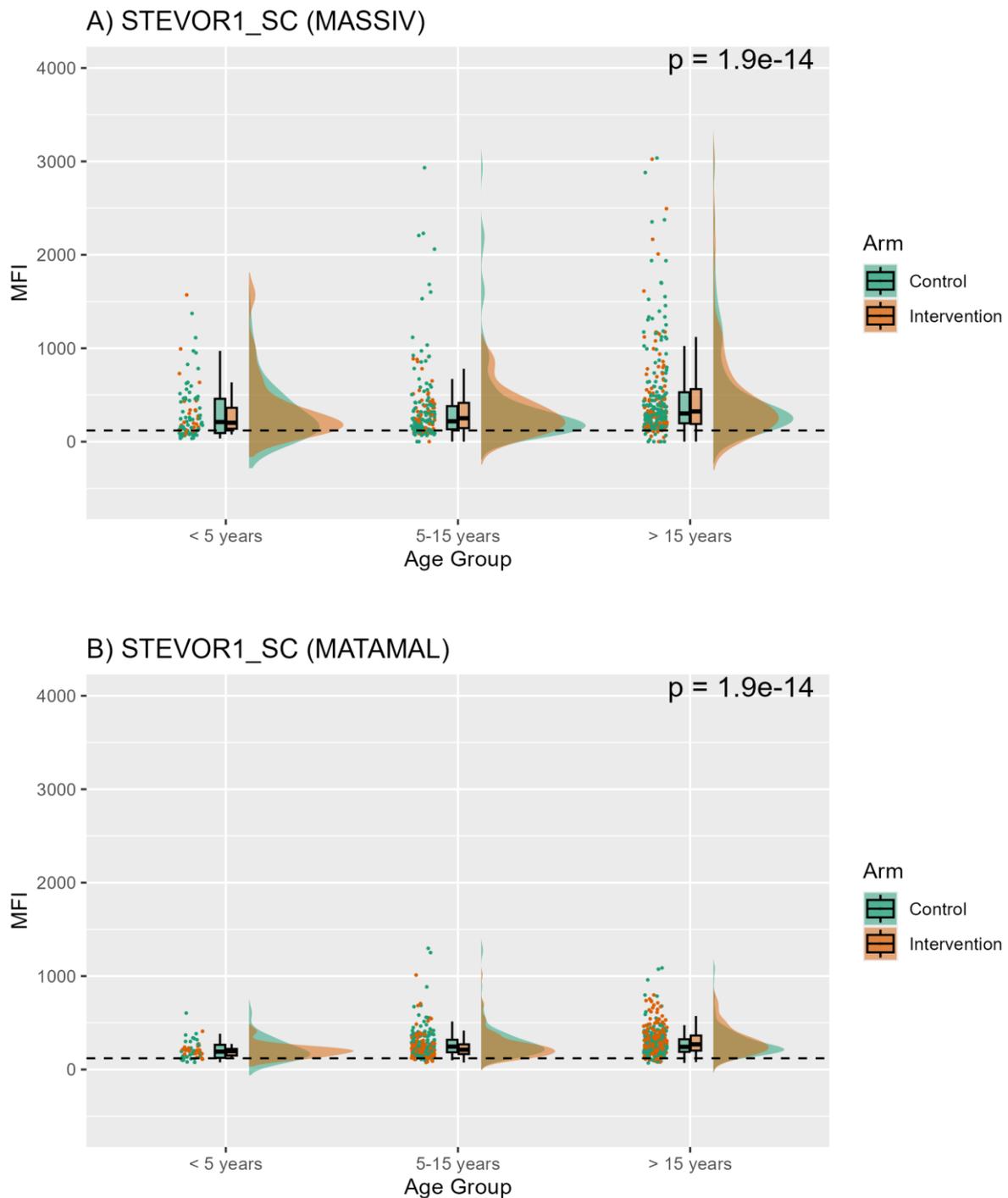
B) HSP40.Ag1 (MATAMAL)



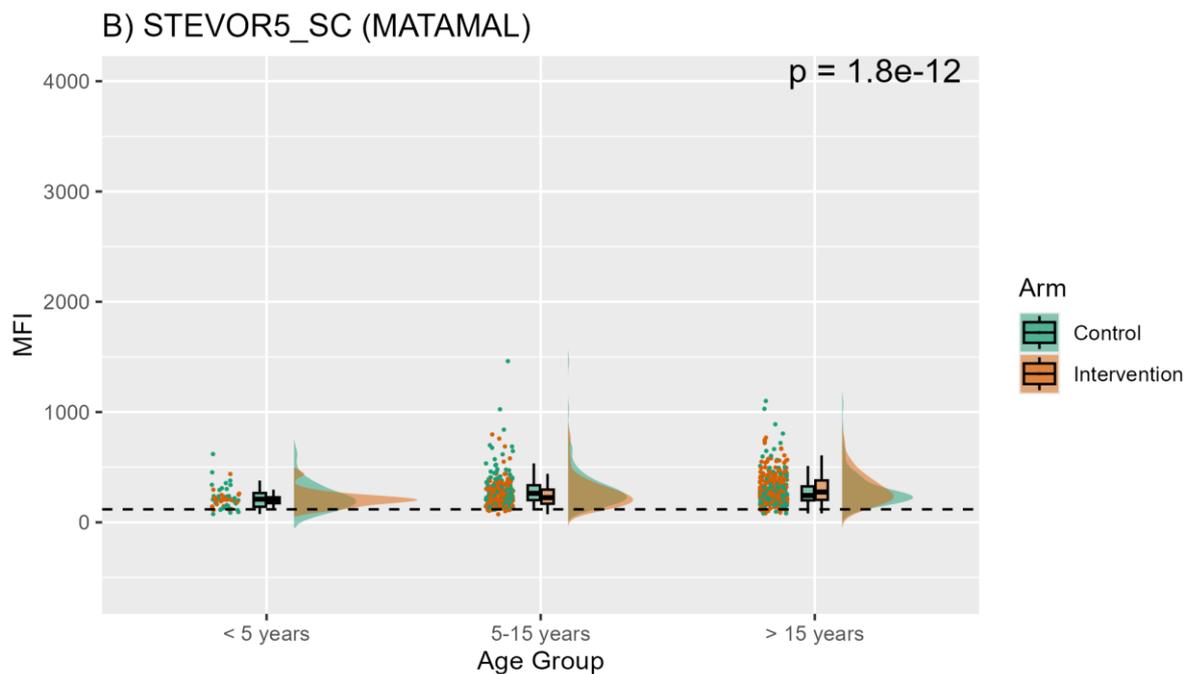
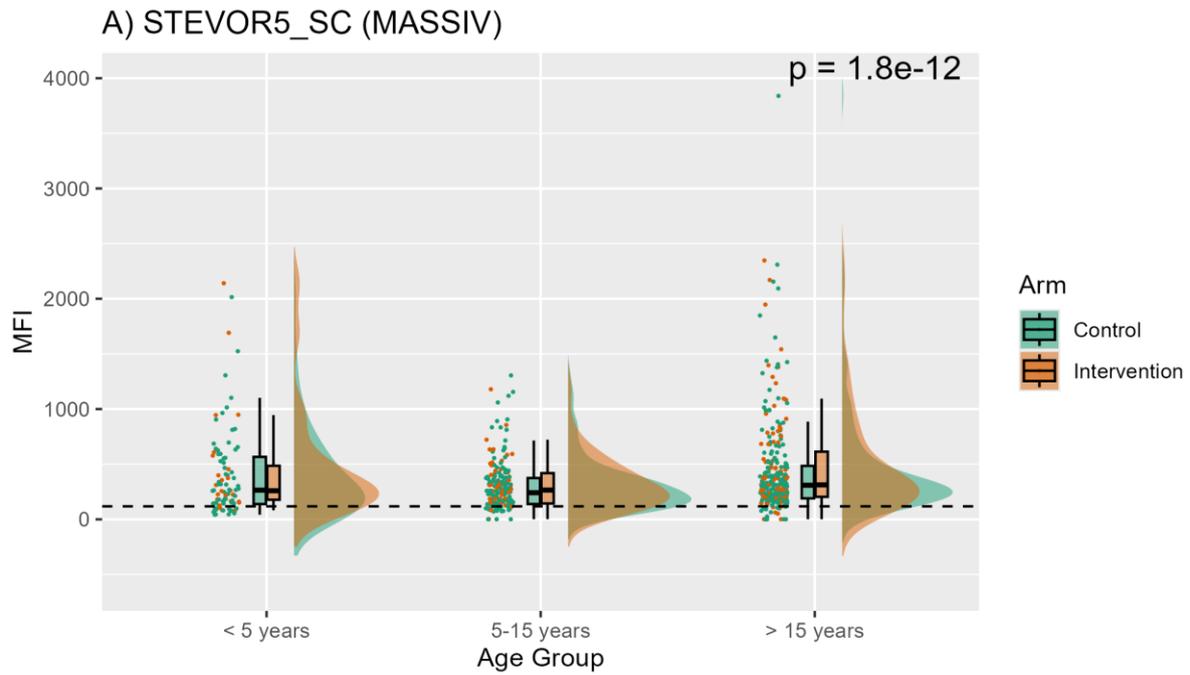
Supplementary Figure 5.5: *Pf*HSP40.Ag1 raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.



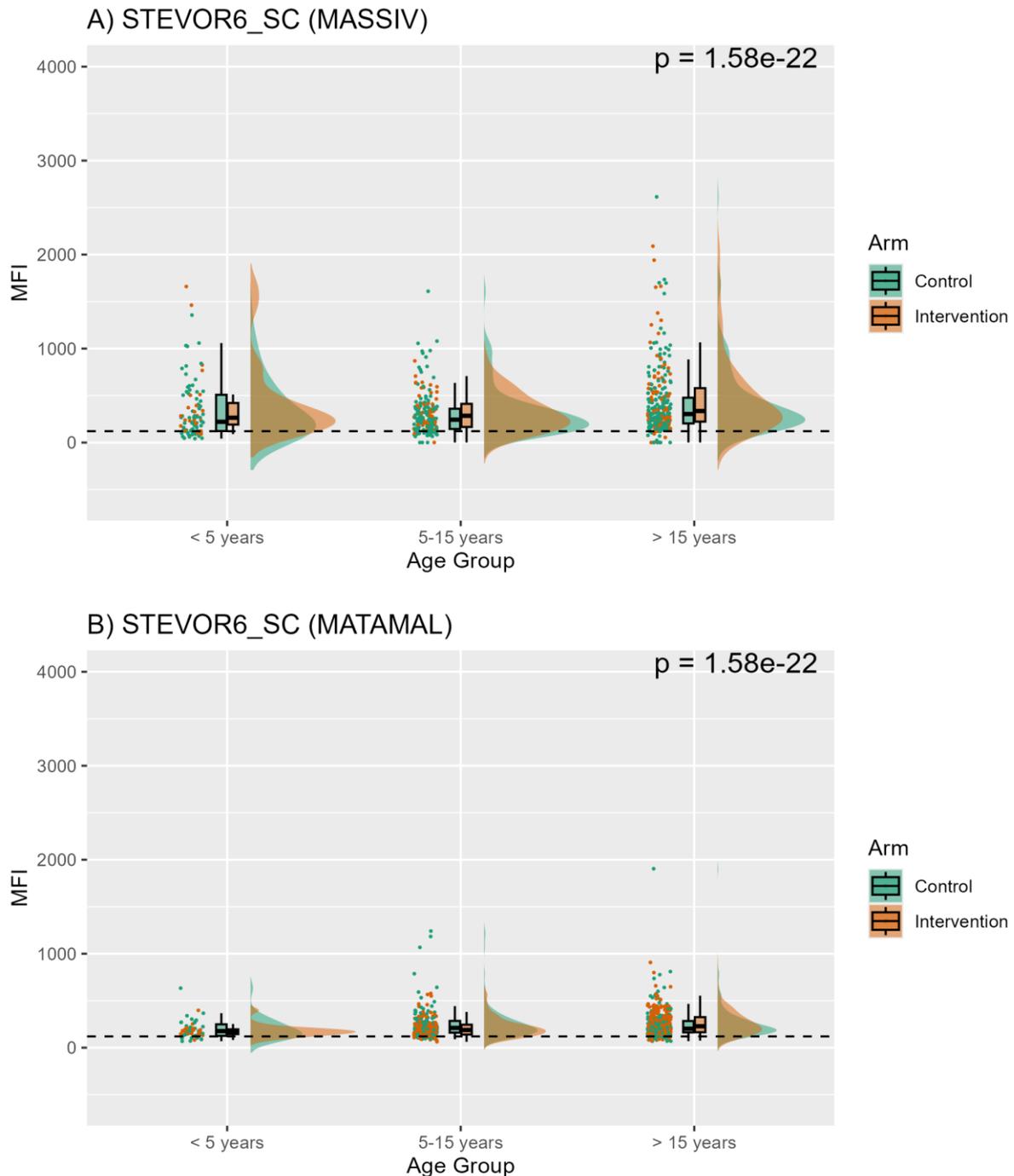
Supplementary Figure 5.6: *Pf*Rh5.1 raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.



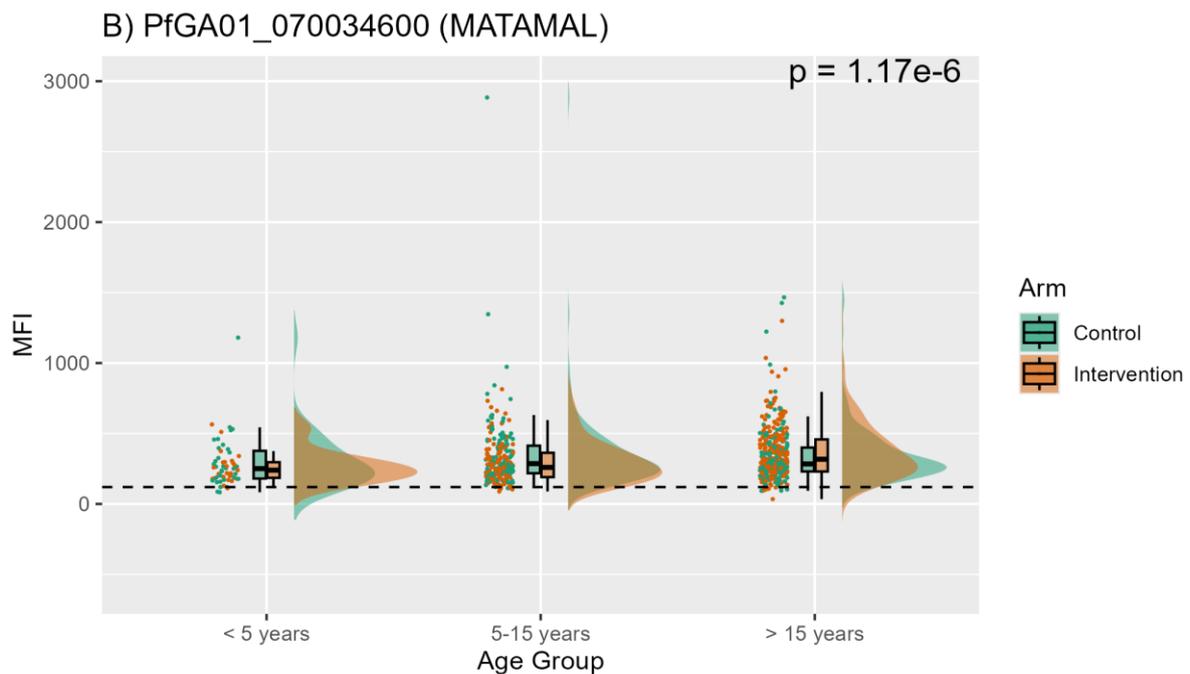
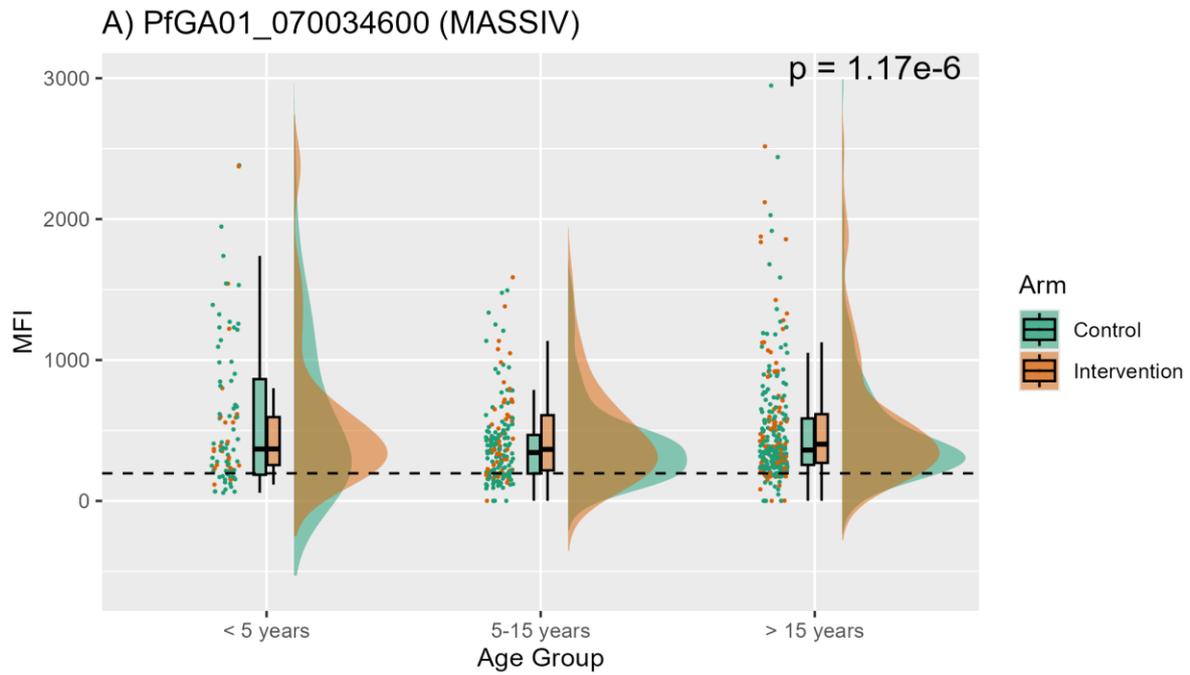
Supplementary Figure 5.7: STEVOR1_SC raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.



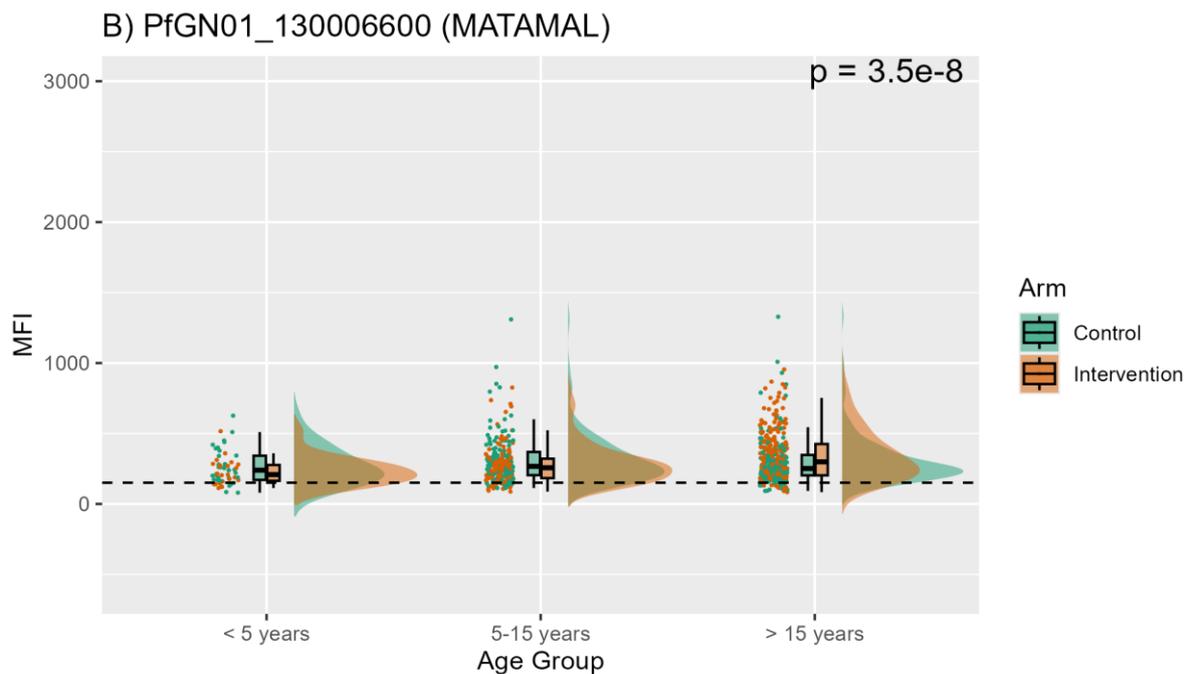
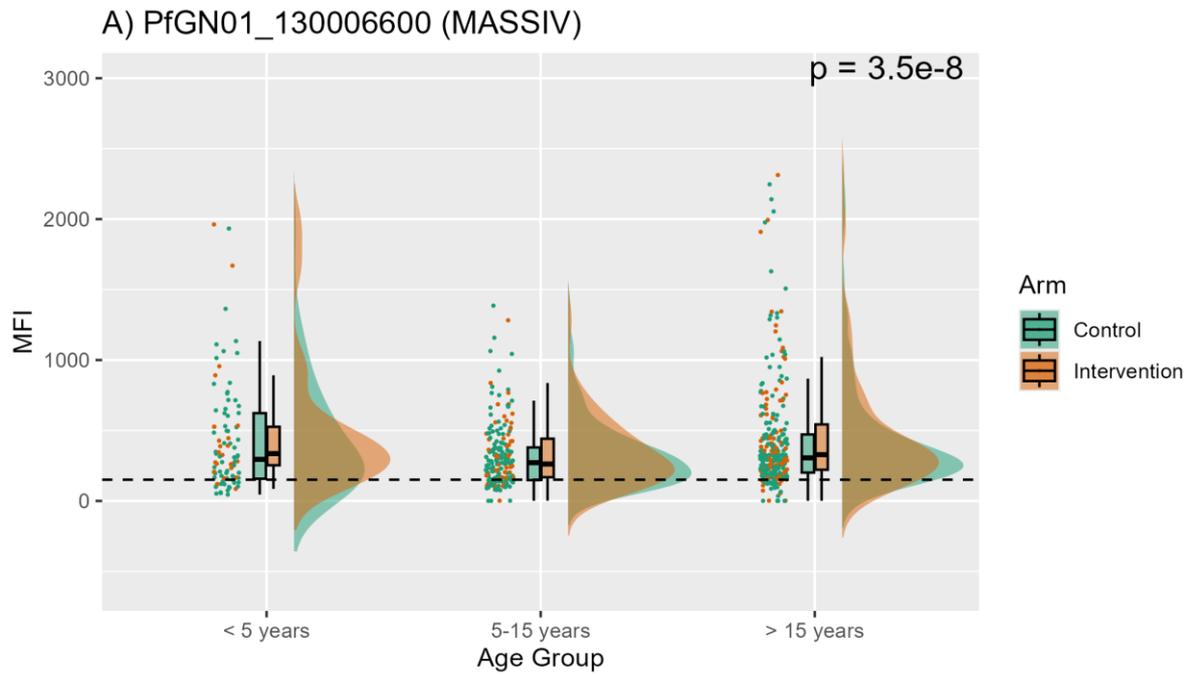
Supplementary Figure 5.8: STEVOR5_SC raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.



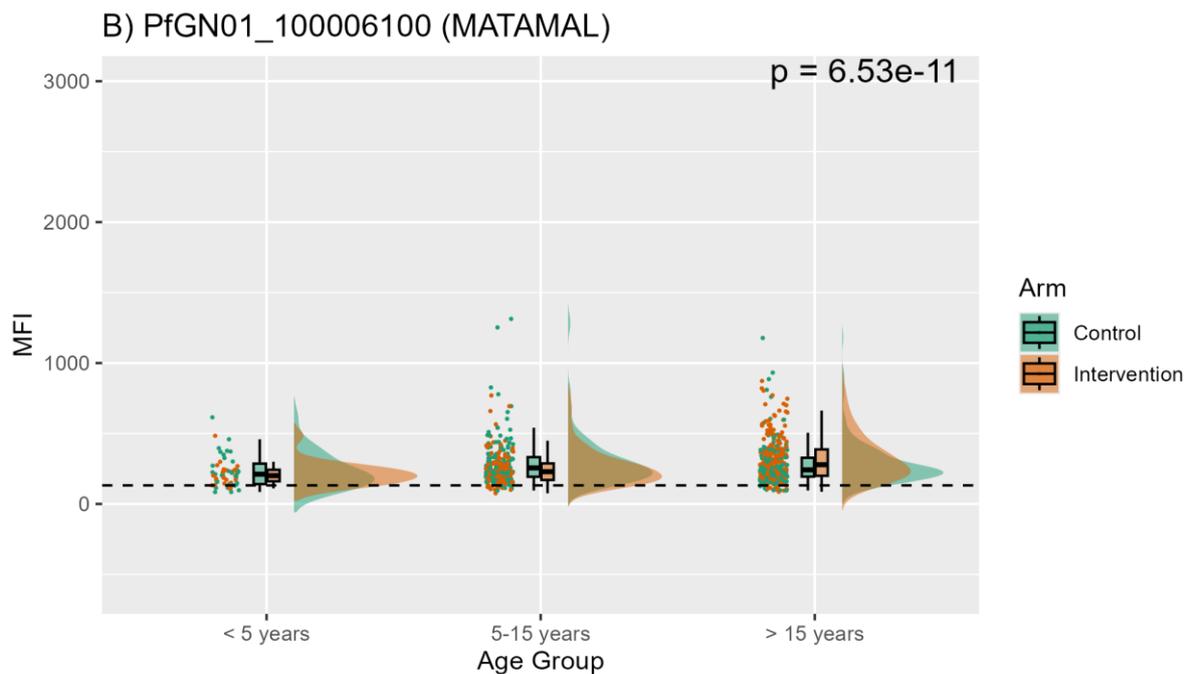
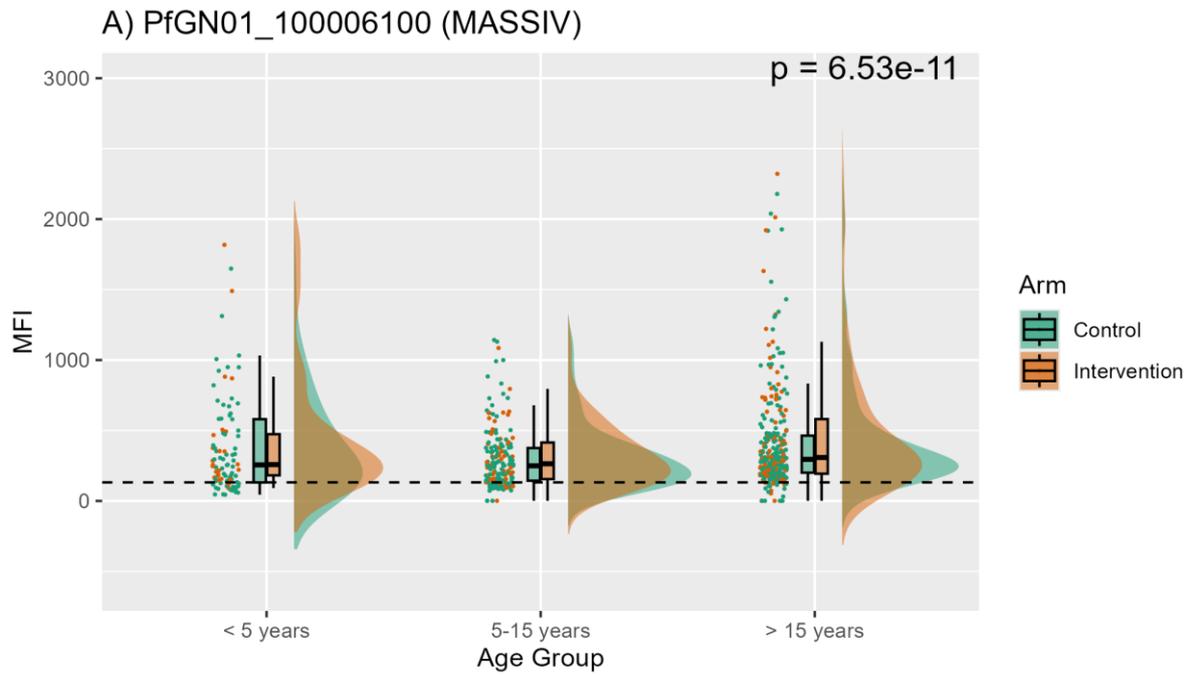
Supplementary Figure 5.9: STEVOR6_SC raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.



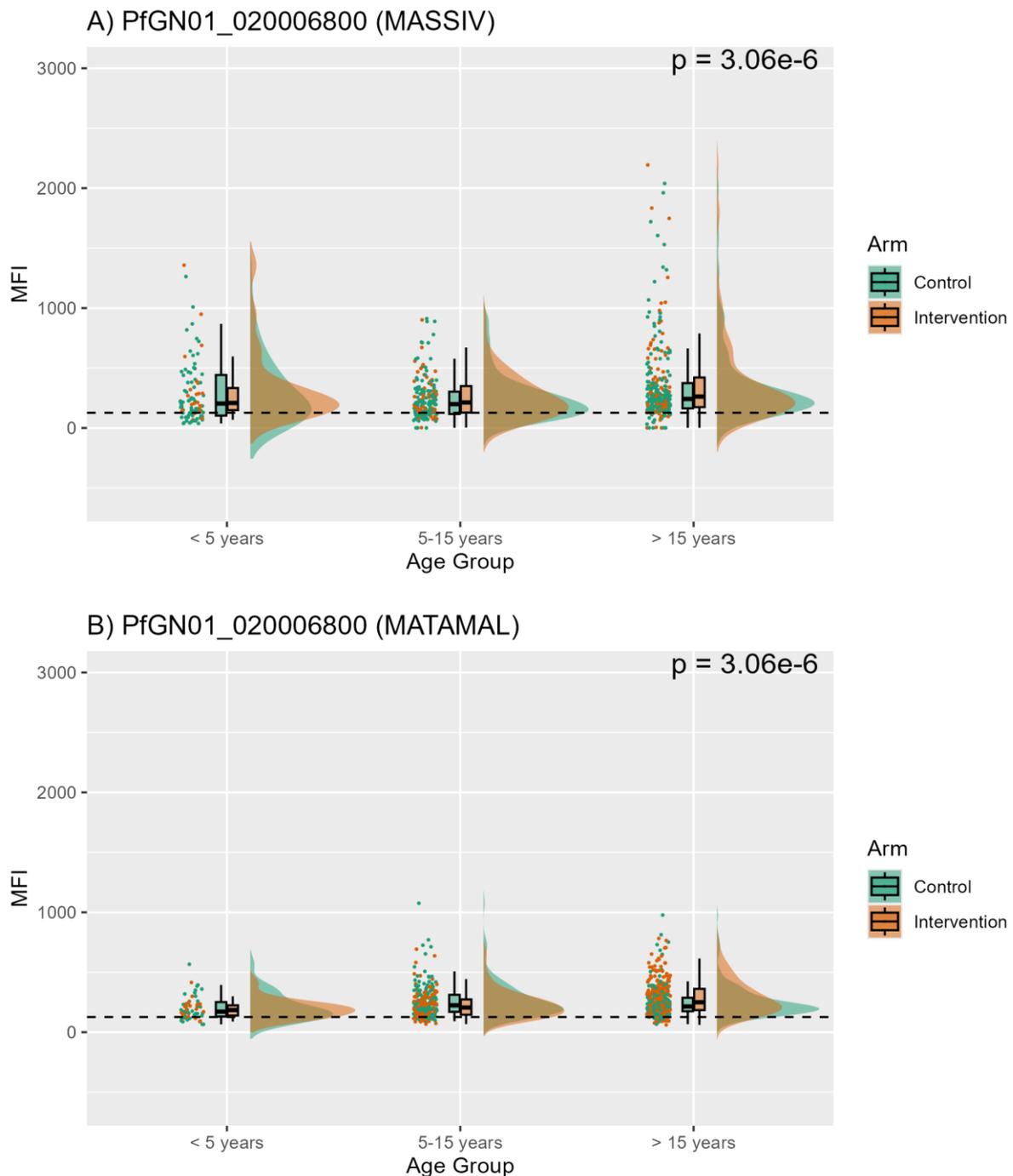
Supplementary Figure 5.10: *PfGA01_070034600* raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.



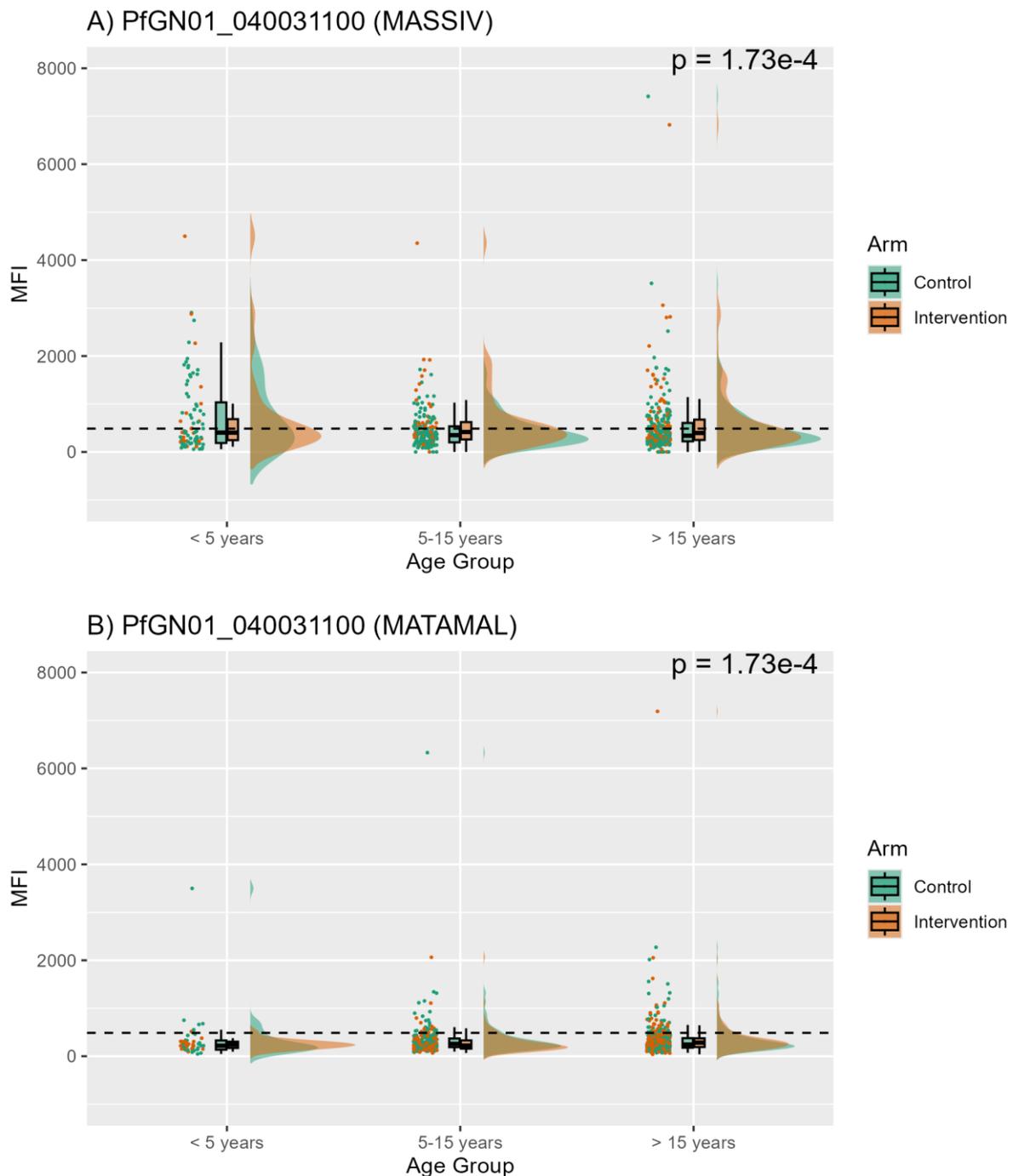
Supplementary Figure 5.11: *PfGN01_130006600* raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.



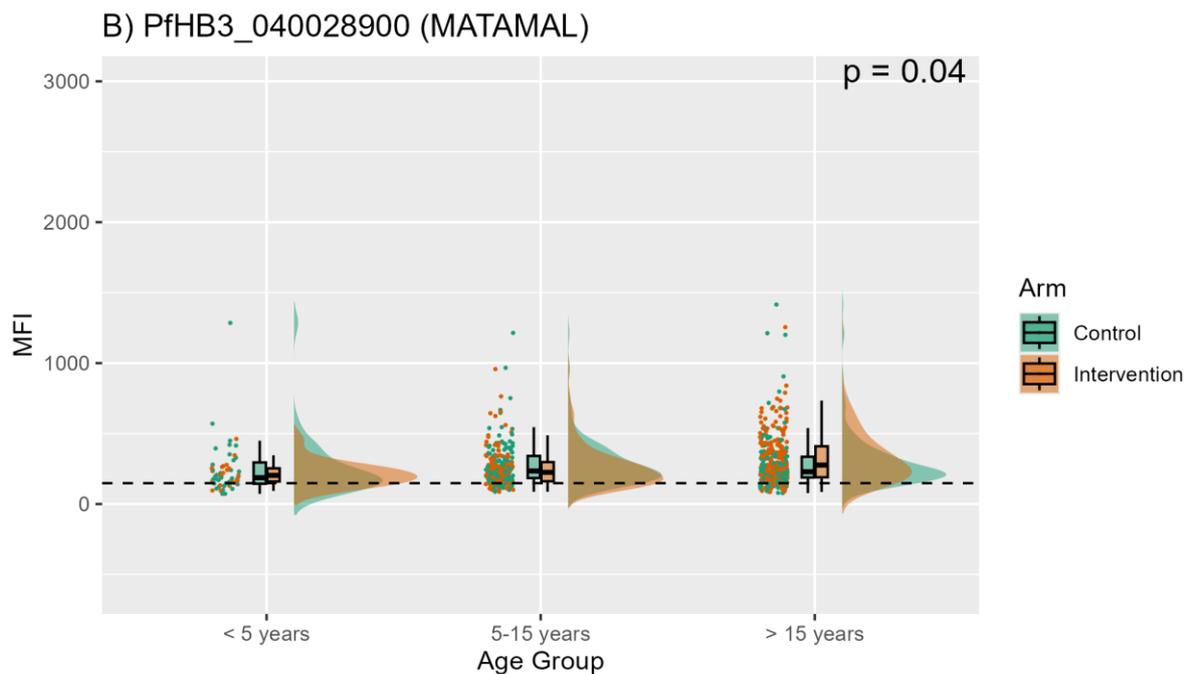
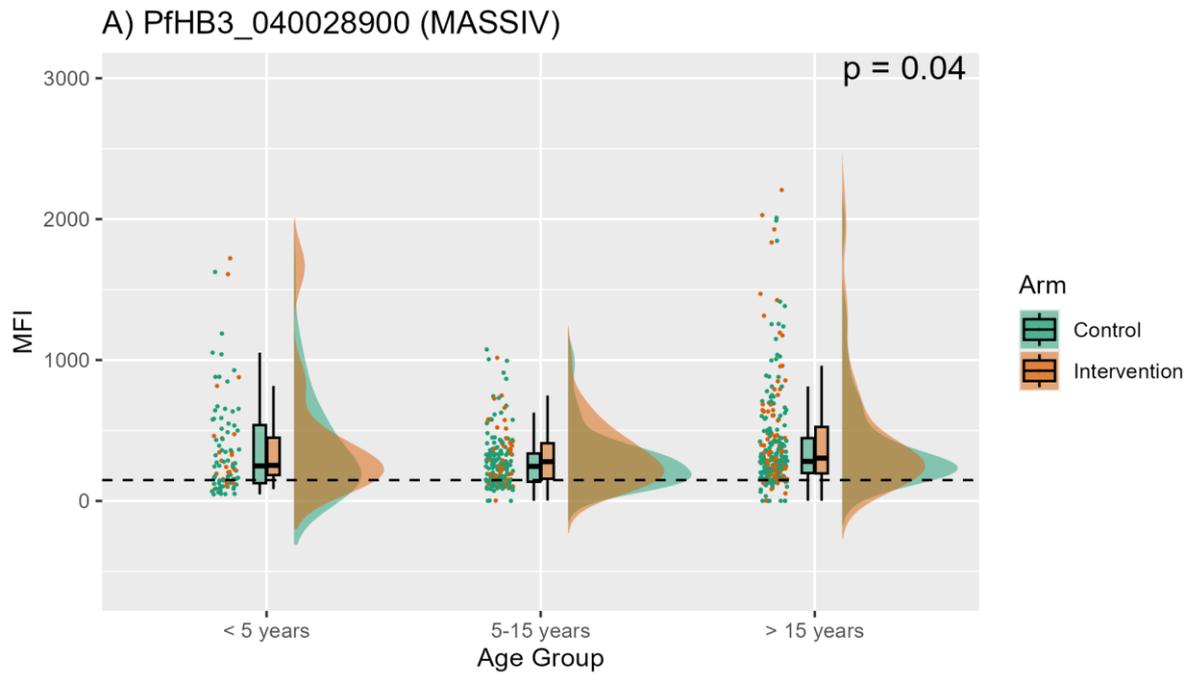
Supplementary Figure 5.12: *PfGN01_100006100* raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.



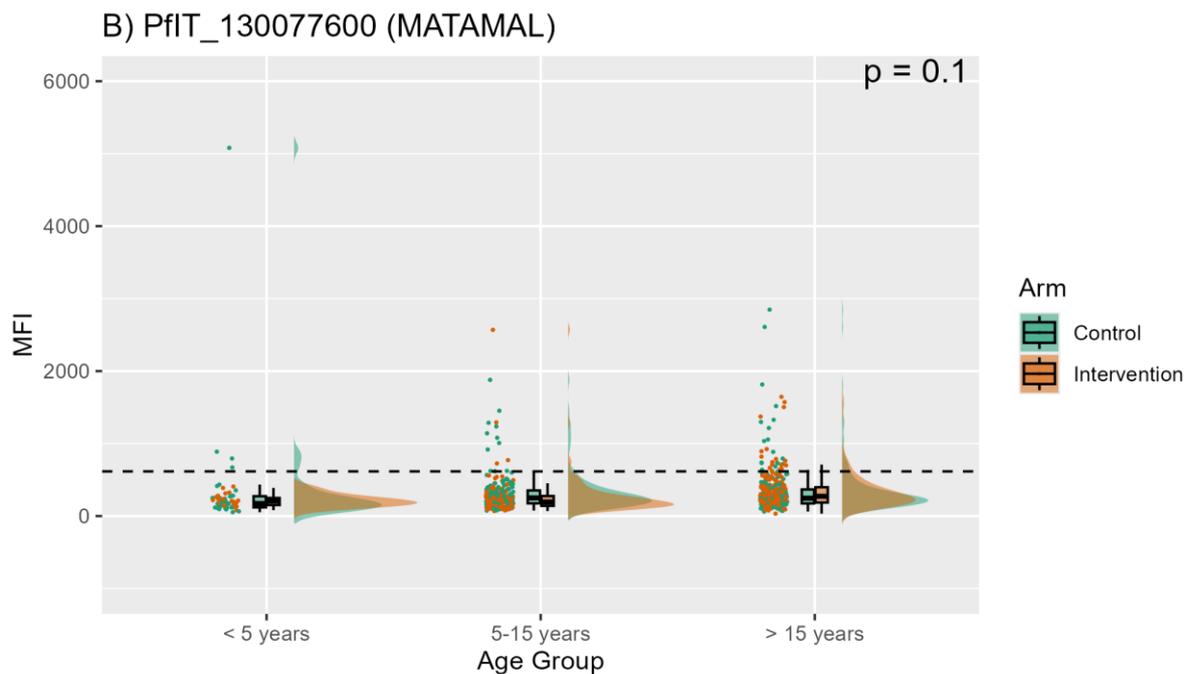
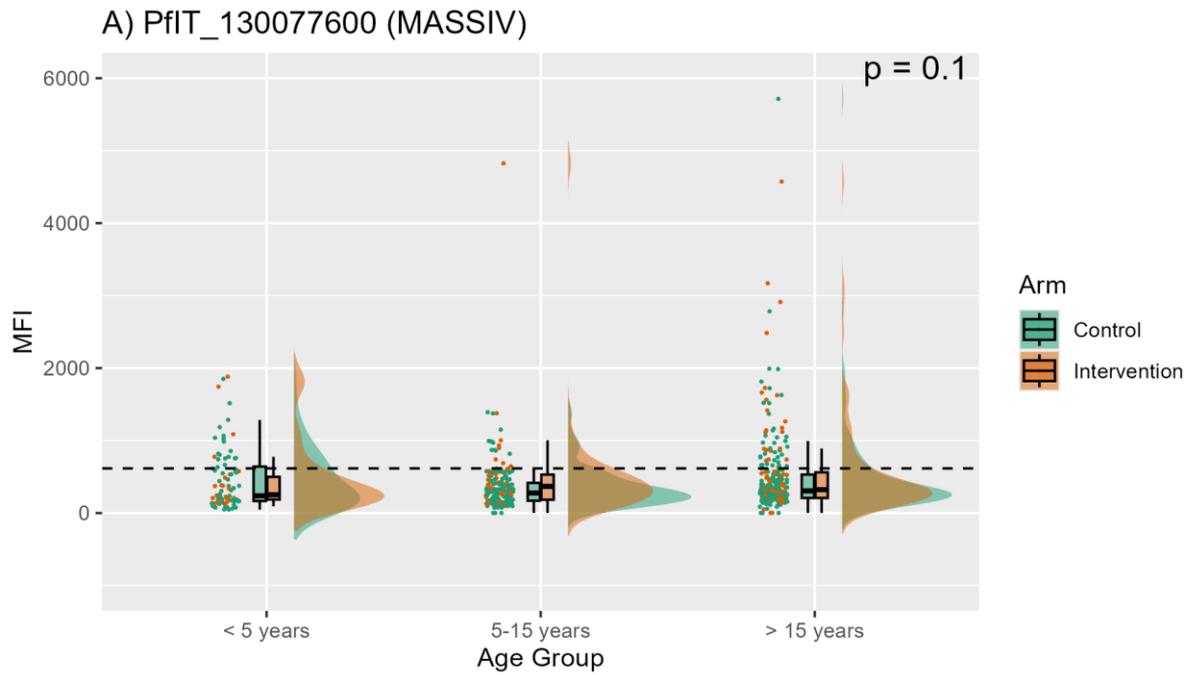
Supplementary Figure 5.13: PfGN01_020006800 raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.



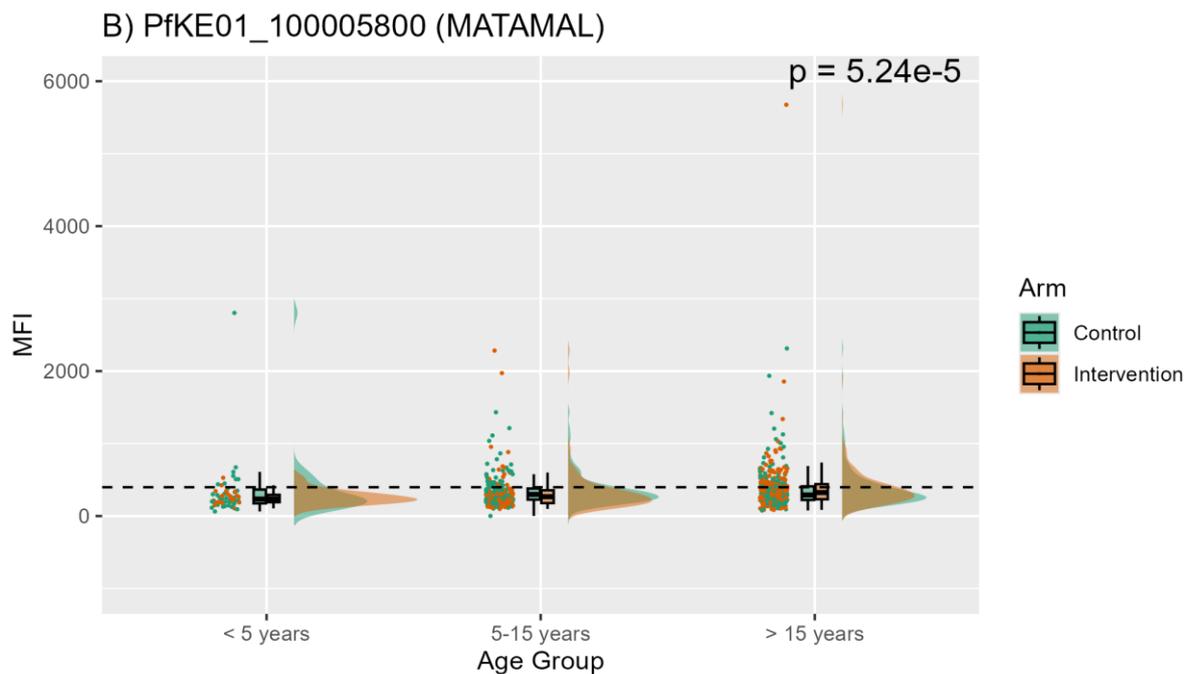
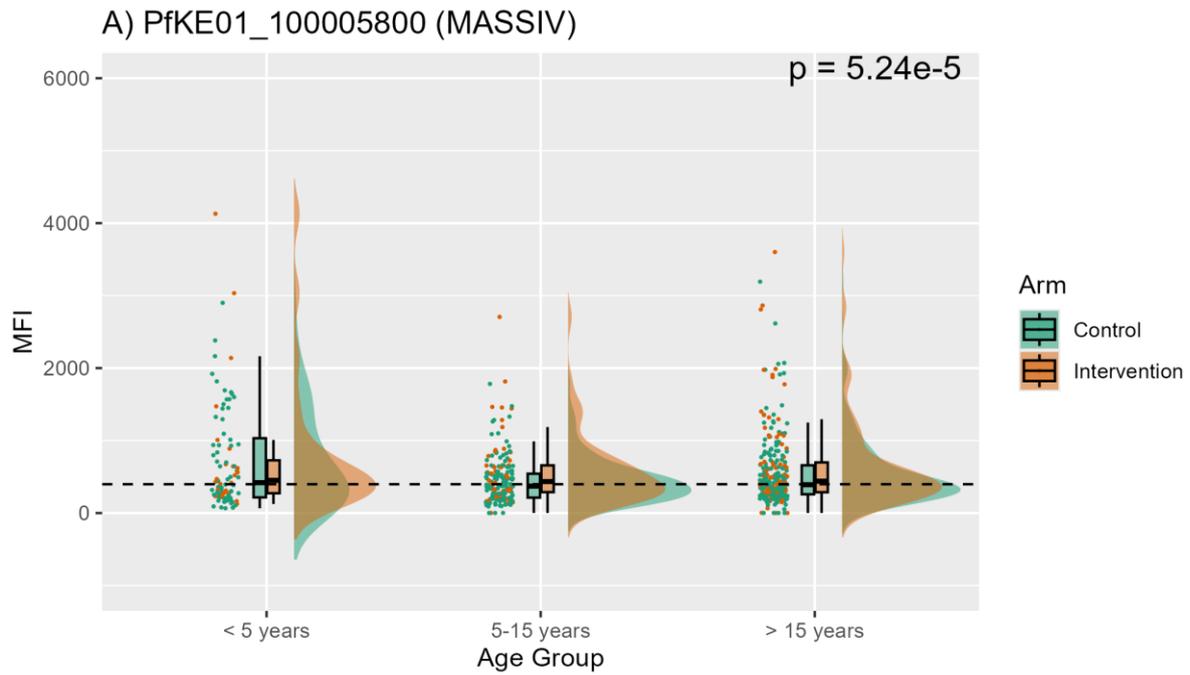
Supplementary Figure 5.14: *PfGN01_040031100* raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.



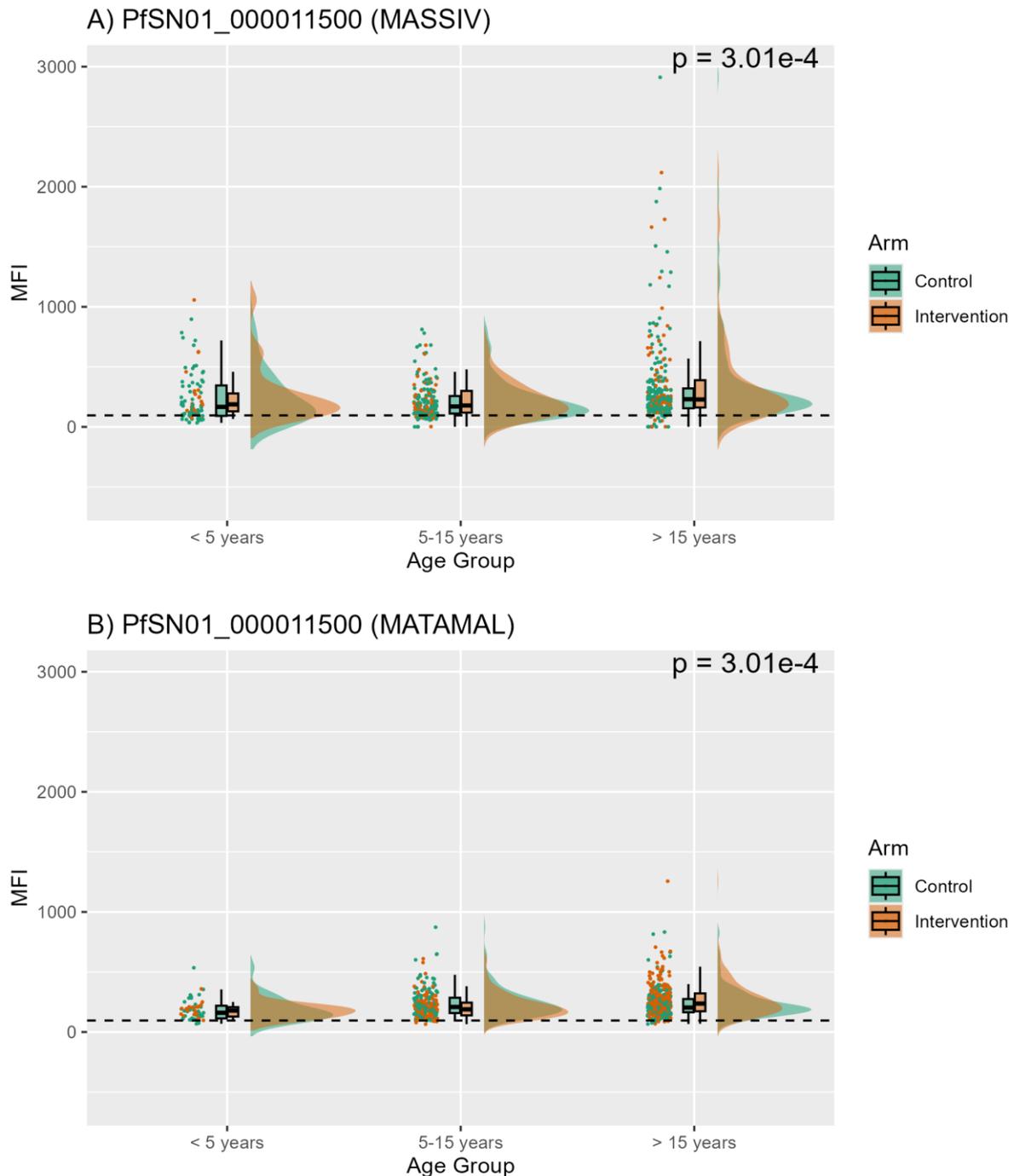
Supplementary Figure 5.15: PfHB3_040028900 raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.



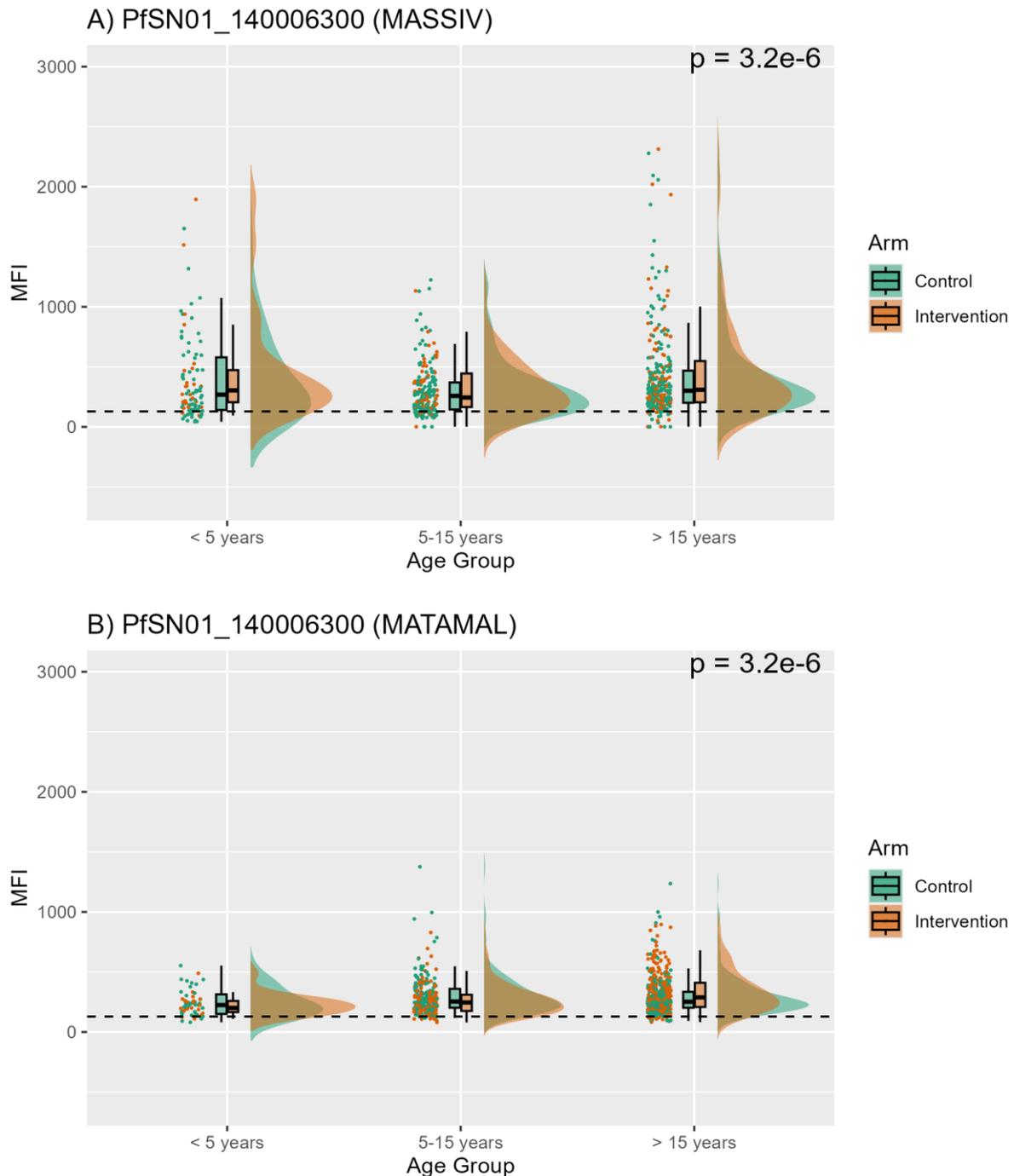
Supplementary Figure 5.16: *Pfit*_130077600 raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.



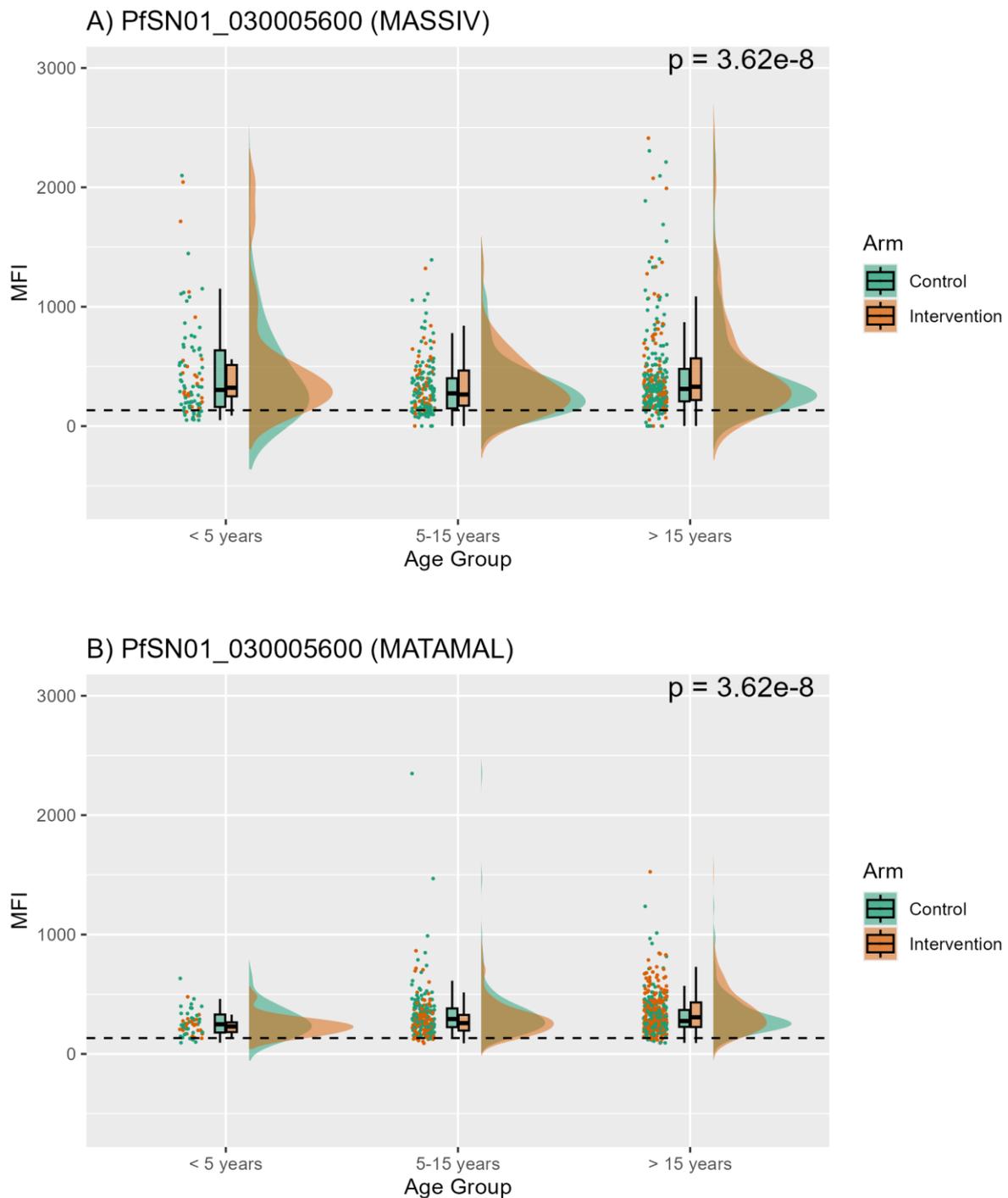
Supplementary Figure 5.17: *PfKE01_100005800* raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.



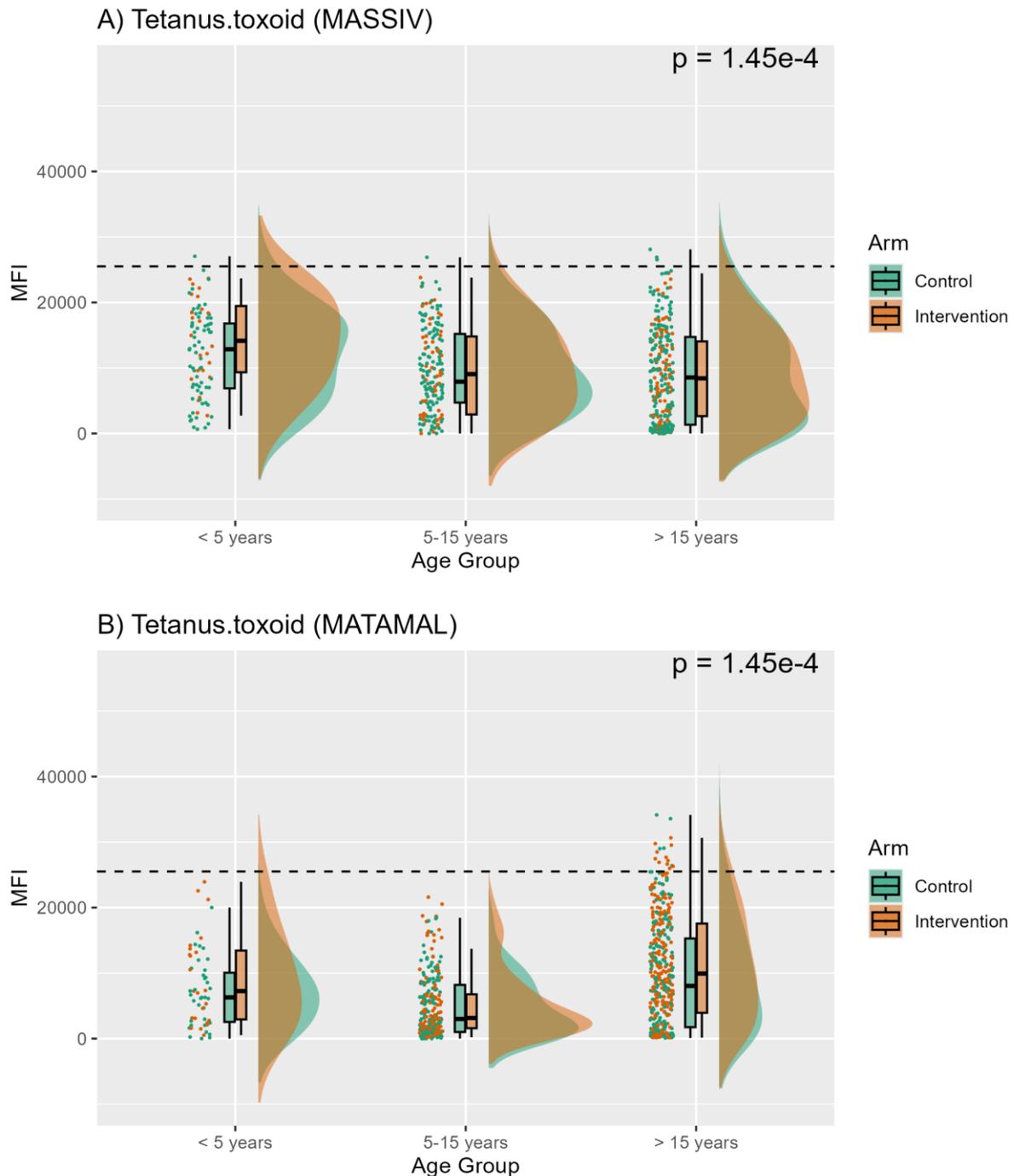
Supplementary Figure 5.18: PfSN01_000011500 raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.



Supplementary Figure 5.19: *PfSN01_140006300* raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.



Supplementary Figure 5.20: PfSN01_030005600 raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.



Supplementary Figure 5.21: Tetanus toxoid raincloud plots of cleaned MFI data for A) MASSIV samples and B) MATAMAL samples. Scatter plots represent each individual sample, boxplots represent mean plus standard deviation of the data and the cloud represents the background shape of the data. Data is colour coded per treatment arm. Dashed line represents the seropositivity threshold calculated using mean negative MFI plus three standard deviations. P-values represent the Man-Whitney non-parametric test for statistical differences.

Supplementary Table 1: Sample proportion seropositive to n number of STEVOR V2 recombinants.

STEVOR V2 (n)	FIGHTMAL, n (%)	MASSIV/MATAMAL, n (%)
1	34 (3.1)	12 (1.0)
2	25 (2.3)	18 (1.5)
3	14 (1.3)	12 (1.0)
4	18 (1.7)	20 (1.7)
5	22 (2.0)	17 (1.4)
6	18 (1.7)	40 (3.3)
7	54 (5.0)	49 (4.1)
8	206 (19.0)	560 (46.3)
9	136 (12.5)	145 (12.0)
10	161 (14.9)	98 (8.1)
11	396 (36.5)	236 (19.5)

Supplementary Table 2: Seropositivity threshold per antigen according to malaria naïve samples mean MFI plus three standard deviations.

Antigen Name	Threshold (MFI)
PfAMA1	317
PfMSP1.19	400
Etramp5.Ag1	147
CSP	747
HSP10.Ag1	1420
Rh5.1	205
STEVOR1_SC	120
STEVOR5_SC	118
STEVOR6_SC	121
PfGN01_130006600	151
PfGN01_100006100	132
PfGN01_020006800	127
PfGN01_040031100	487
PfSN01_000011500	96
PfSN01_140006300	129
PfSN01_030005600	133
PfKE01_100005800	398
PfIT_130077600	616
PfHB3_040028900	148
PfGA01_070034600	196
Tetanus.toxoid	25518

Supplementary Table 3: Antigen specific percentage seropositivity of FIGHTMAL samples stratified by age group and sampling time-point.

Antigen Name	Baseline			6 weeks post			16 weeks post		
	< 5 years	5 – 15 years	> 15 years	< 5 years	5 – 15 years	> 15 years	< 5 years	5 – 15 years	> 15 years
PfAMA1	98.97	100.00	100.00	98.99	100.00	100.00	95.48	98.78	NA
PfMSP1.19	98.97	100.00	100.00	98.99	97.59	100.00	90.95	89.02	NA
Etramp5.Ag1	96.92	98.83	98.50	93.47	94.58	94.35	84.42	84.15	NA
CSP	83.08	97.66	100.00	76.88	96.39	100.00	55.78	86.59	NA
HSP10.Ag1	74.36	70.18	79.70	61.81	43.98	71.77	41.21	37.80	NA
Rh5.1	97.44	98.83	100.00	92.46	93.37	100.00	88.44	89.02	NA
STEVR01_SC	90.77	95.32	96.24	84.92	87.35	86.29	82.41	89.02	NA
STEVR05_SC	84.10	93.57	96.99	75.38	83.73	84.68	78.39	87.80	NA
STEVR06_SC	83.08	92.98	95.49	73.37	85.54	77.42	79.90	91.46	NA
PfGN01_130006600	76.41	87.13	91.73	66.83	72.29	75.81	72.86	78.05	NA
PfGN01_100006100	77.44	88.30	92.48	67.84	72.89	75.00	76.38	81.71	NA
PfGN01_020006800	73.85	83.63	92.48	61.81	67.47	71.77	64.82	71.95	NA
PfGN01_040031100	50.77	54.39	46.62	40.70	34.94	37.90	68.84	68.29	NA
PfSN01_000011500	72.31	83.63	93.98	60.80	67.47	76.61	65.83	73.17	NA
PfSN01_140006300	77.95	87.72	92.48	67.84	74.10	79.03	71.86	79.27	NA
PfSN01_030005600	80.51	89.47	93.98	68.84	75.30	79.84	74.37	82.93	NA
PfKE01_100005800	55.90	62.57	62.41	47.74	45.78	49.19	72.86	78.05	NA
PfIT_130077600	31.79	35.67	36.84	22.61	19.88	27.42	54.77	52.44	NA
PfHB3_040028900	80.00	89.47	93.23	69.35	75.90	76.61	84.92	89.02	NA
PfGA01_070034600	73.33	88.89	83.46	67.34	68.67	66.94	83.92	90.24	NA
Tetanus.toxoid	1.03	0.00	13.53	0.50	0.00	12.90	1.01	1.22	NA

Supplementary Table 4: Antigen specific percentage seropositivity of MASSIV/MATAMAL samples stratified by age group and intervention arms.

Antigen Name	Control			Intervention		
	< 5 years	5 – 15 years	> 15 years	< 5 years	5 – 15 years	> 15 years
PfAMA1	47.06	78.95	97.86	54.17	68.52	96.84
PfMSP1.19	66.18	90.98	95.72	75.00	94.44	96.84
Etramp5.Ag1	95.59	100.00	100.00	100.00	100.00	100.00
CSP	29.41	43.61	90.91	20.83	59.26	88.42
HSP10.Ag1	23.53	40.60	51.87	25.00	29.63	49.47
Rh5.1	100.00	100.00	99.47	100.00	100.00	100.00
STEVOR1_SC	69.12	79.70	95.19	87.50	83.33	91.58
STEVOR5_SC	76.47	81.20	93.58	91.67	88.89	92.63
STEVOR6_SC	77.94	83.46	93.58	91.67	94.44	93.68
PfGN01_130006600	77.94	73.68	90.91	91.67	79.63	88.42
PfGN01_100006100	75.00	79.70	92.51	91.67	88.89	91.58
PfGN01_020006800	67.65	70.68	90.37	87.50	77.78	88.42
PfGN01_040031100	44.12	28.57	30.48	41.67	44.44	38.95
PfSN01_000011500	70.59	81.95	94.12	91.67	87.04	92.63
PfSN01_140006300	79.41	80.45	93.05	91.67	83.33	91.58
PfSN01_030005600	79.41	81.95	92.51	91.67	85.19	92.63
PfKE01_100005800	50.00	45.11	48.66	62.50	55.56	55.79
PfIT_130077600	26.47	8.27	20.86	16.67	14.81	24.21
PfHB3_040028900	67.65	71.43	89.30	83.33	79.63	87.37
PfGA01_070034600	73.53	72.93	89.84	91.67	79.63	86.32
Tetanus.toxoid	1.47	0.75	1.60	0.00	0.00	0.00

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	lsh1601985	Title	Miss
First Name(s)	HRISTINA CHAVDAROVA		
Surname/Family Name	VASILEVA		
Thesis Title	BIOCHEMICAL AND BIOINFORMATIC CHARACTERISATION OF UNDERSTUDIED ERYTHROCYTE SURFACE EXPRESSED HYPERVARIABLE PROTEIN FAMILIES IN PLASMODIUM FALCIPARUM.		
Primary Supervisor	Dr Kevin Tetteh		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Malaria Journal
---	-----------------

Please list the paper's authors in the intended authorship order:	Hristina Vasileva, Christian Kositz, Helena Brazal-Monzo, Harry Hutchins, Eunice Teixeira da Silva. Edgard Dabira, Mamadou Ousmane Ndiath, Chris Drakeley. Umberto D'Alessandro, Anna Last, Kevin K.A. Tetteh
Stage of publication	Submitted

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I have done all laboratory work, some sample collection, all data analysis and manuscript preparation.
--	--

SECTION E

Student Signature	HRISTINA VASILEVA
Date	30.05.2024

Supervisor Signature	ANNA LAST
Date	06.06.2024

Chapter 8: Discussion, Future Work and Limitations

8.1 Discussion

8.1.1 Importance of STEVORs as a multi-gene family

According to the findings of this thesis, most STEVOR variants are recognised rapidly, even after an initial infection. This immune recognition is not influenced by malaria exposure intensity or age, as high seroprevalence is observed even in children under five years of age. This raises the question: what is the significance of STEVOR being a multi-gene family?

Several hypotheses may provide answers to this question. Firstly, STEVORs likely have functional roles at various stages of the parasite's life cycle, such as mediating adhesion or red blood cell invasion. Previous studies have demonstrated that the STEVOR repertoire expressed on the apical ends of merozoites differs from that found on infected erythrocytes (1). Moreover, STEVORs are expressed in distinctive cascades, suggesting that different variants serve specific roles at various stages of the parasite's life cycle (2). This diversity equips the parasite with a versatile toolkit, enabling it to adapt to changing host environments and niches.

Secondly, this diversity may facilitate the parasite's ability to bind to different host cells. It is well-established that *Pf*EMP1 and RIFIN proteins bind to various receptors on human endothelial and immune cells, contributing to clinical outcomes such as cerebral malaria and severe malaria (3), (4), (5), (6). For STEVORs, glycophorin C has been identified as a receptor, but the possibility of additional receptors remains (7), (8). The ability to express different STEVOR variants might allow the parasite to adapt to bind alternative receptors during an infection, enhancing its survival and persistence.

Additionally, expressing a large repertoire of variants from the same protein family with similar functions could serve as a redundant survival mechanism. This redundancy might safeguard

the parasite against the host's immune system successfully targeting a single protein. Furthermore, STEVOR proteins may act as immunological decoys, diverting the host's immune responses away from critical targets like *PfEMP1* (9).

Finally, the broad recognition of STEVORs by the immune system may reflect ongoing low-level infections, resulting in chronic persistence that facilitates transmission. This persistence could help the parasite survive under varying conditions of immune pressure. For instance, STEVORs may contribute to maintaining a reservoir of asymptomatic carriers, particularly in low-endemicity settings, as observed in the MASSIV and MATAMAL samples.

8.1.2 STEVOR protein topology

The serological results presented in this thesis provide evidence that the semi-conserved region of STEVOR proteins is exposed to the immune system. Serorecognition rates for semi-conserved recombinant proteins were high, exceeding those for the hypervariable domain, which was previously hypothesised to be the only portion of the protein exposed to extracellular circulation and thus accessible to the immune system (10). These findings challenge the proposed two-transmembrane topology of STEVOR proteins, which would typically restrict external exposure to the loop regions flanked by the transmembrane domains (10). The observed external exposure of the semi-conserved region suggests that it may play a structural or adhesive role, indicating a functional necessity for its placement outside the cell.

Additionally, this study demonstrates that the semi-conserved region is itself quite variable, apart from two small, conserved segments. This variability aligns it more closely with the hypervariable domain in terms of suspected functional roles. Experimental evidence from studies on the RIFIN protein family, which shares significant similarities with STEVOR proteins in topology, localization, and function, supports the possibility of a single-transmembrane topology (11), (12). In RIFINs, a one-transmembrane topology exposes the

semi-conserved region, variable domain, and a hydrophobic segment to the extracellular environment (12). Given the evolutionary relationship between STEVOR and RIFIN proteins, it is highly plausible that STEVOR proteins also possess a one-transmembrane topology (11).

While computational tools such as TMHMM used in this study suggest the presence of one signal motif and two transmembrane domains, this prediction requires experimental validation. Protease protection assays could be employed to determine which parts of the STEVOR protein are intracellular and extracellular. Additionally, cryo-electron microscopy or X-ray crystallography could be used to resolve the precise 3D structure and membrane topology of STEVOR proteins. Testing antibodies specific to both the hypervariable and semi-conserved regions on intact versus permeabilized cells would further confirm the external exposure of these regions.

8.1.3 STEVOR function

If the semi-conserved regions of STEVOR proteins are consistently exposed to the immune system, it suggests that immune evasion might not be their primary role. Instead, these regions could play critical functional roles, such as binding host receptors or facilitating interactions necessary for parasite survival and adaptation (10). STEVOR proteins might contribute to maintaining chronic infections by enabling low-level parasitaemia to persist, ensuring the parasite's long-term survival within the host. They may also support parasite dissemination by promoting survival in distinct host tissues or enhancing interactions with specific receptors, such as glycophorin C.

Despite the presence of the PfEMP1 and RIFIN families, which are also involved in cytoadherence and immune evasion, STEVOR proteins likely provide functional redundancy, ensuring parasite survival under diverse immune or environmental pressures. They may act as specialized mediators in scenarios where PfEMP1 or RIFIN proteins are less effective, adapting

to specific host cell types or immune challenges. Alternatively, their diversification could represent an evolutionary advantage, equipping the parasite with a broader toolkit to navigate host heterogeneity. These hypotheses must be tested experimentally to unravel the importance of STEVORs and their role in parasite biology.

8.1.4 STEVOR as transmission marker

STEVOR proteins may have some potential as markers for tracking transmission through time, but there are significant limitations to their utility for this purpose according to the results from this study. The high and broad recognition of STEVORs, regardless of endemicity and exposure, suggests that they do not strongly differentiate between recent and past infections. This reduces their utility for pinpointing changes in transmission intensity over time. Furthermore, the potential cross-reactivity of antibodies to semi-conserved and hypervariable regions complicates the interpretation of serological data. Antibodies may reflect cross-reactive responses rather than specific exposures to distinct parasite variants.

Additionally, the study suggests that even a single exposure can induce broad responses to STEVORs. This rapid and extensive seroconversion makes it difficult to correlate antibody breadth to intensity of transmission over time. The lack of significant differences in STEVOR antibody levels between high- and low-endemicity settings undermines their value as markers for changing transmission intensity. If STEVOR responses reflect persistent low-level parasitaemia or immune decoy functions, they may not correlate well with actual transmission rates or reductions resulting from interventions.

8.2 Future work

One of the suggested reasons for the high breadth of responses to the STEVOR hypervariable domain library found, irrespective of the setting's *P. falciparum* endemicity level and levels of infection exposure, was that there was high degree of cross-reactivity of antibodies to individual library members. Performing antibody avidity testing, which is the functional antibody affinity, measured as the strength of interaction between antibodies and antigens, is proposed to be used as next steps to reveal if the high breadth of responses were due to true exposure to the members of the library, or due to antibodies cross-reactivity. Avidity testing can be performed with the use of dissociating agents, such as ammonium thiocyanate and guanidine hydrochloride (13). Antibody avidity testing can help differentiate true exposure from cross-reactivity by measuring the strength of antibody-antigen interactions; high avidity suggests specific and repeated exposure to individual STEVOR variants, while low avidity indicates potential cross-reactivity due to shared or conserved epitopes among the variants.

P. falciparum expresses STEVOR proteins at all parasitic stages, including but not limited to, gametocytes, trophozoites and merozoites (2),(14). Following the avidity testing, we could further identify highly cross-reactive antibodies for the design of monoclonal antibodies (mAbs) for the neutralisation of STEVOR variants expressed on the apical end of merozoites, shown that can inhibit merozoite infection of the red blood cells (1),(15).

Future research should focus on understanding the biological role of STEVOR proteins rather than solely pursuing serological profiling. A key area of exploration would be to determine whether STEVORs contribute directly to cytoadherence or rosetting, independently of *PfEMP1* and RIFINs, thereby elucidating the importance of this multi-gene family alongside the other two. Proposed experiments include generating *P. falciparum* lines with STEVOR gene deletions to assess their impact on cytoadherence and rosetting in vitro, using recombinant

STEVEOR proteins in adhesion assays to identify host receptors such as glycoporphin C or other erythrocyte membrane proteins, and testing whether anti-STEVEOR antibodies inhibit parasite adhesion to host cells, which would indicate a functional role in cytoadherence.

Additionally, clarifying the topology of STEVEOR proteins is crucial for understanding their interactions with host receptors. Techniques such as protease protection assays, cryo-electron microscopy, or X-ray crystallography could provide structural insights. Further investigations should explore the potential role of STEVEORs in immune modulation or as immunological decoys, using immune activation assays with peripheral blood mononuclear cells to assess cytokine production, T-cell activation, or immune suppression, as well as testing phagocytosis efficiency in the presence or absence of STEVEORs. Finally, identifying additional host receptors beyond glycoporphin C could involve mass spectrometry to pinpoint specific host proteins that bind STEVEORs, thereby expanding our understanding of their functional relevance.

8.3 Limitations

The selected STEVOR variants for creation of the hypervariable domain recombinant antigen library was based on the *in-silico* analysis using available data from PlasmoDB database. However, from the analysis it can be seen that each geographical isolate has its own STEVOR repertoire with some degree of variability. Therefore, selecting the variants for the library does not truly reflect the expressed STEVORs in parasites in the tested populations: Uganda, The Gambia and Bijagos Archipelago, but instead represents a geographically diverse array of sequence variants based on the *in-silico* interrogation of the STEVOR database protein sequences (n=493). An RNA seq approach could have been applied to samples from the tested populations to identify the exact variants expressed, and based on this data the creation of the library would have been more accurate. Although, this approach is not feasible to be applied for each studied population in terms of time and resources, as we know from literature that there is no apparent group of STEVORs expressed in high versus low endemic areas (16). Moreover, RNA seq does not provide information about STEVOR variants contribution to the infection immunity which was the focus of the study. Finally, the selection of the eleven variants for the library was justified due to time constrains around the expression and purification of the recombinant proteins whilst having field work commitments.

One of the main limitations of the study, also outlined as a limitation in the individual research papers, was the missing samples from the study populations. Unfortunately, this limitation could not be overcome since the study used residual samples from previous studies. A mitigation for this limitation would have been to design a specific cohort to address the question of this study, which was not feasible in terms of finances and time for completion of the thesis.

This study focuses on characterising STEVOR and RIFIN proteins and their involvement in infection immunity. To understand the nature of STEVORs and RIFINs as immunogens and their potential to stimulate a memory response, we could have measured the half-life of

antibody waning overtime, thus assessing the antigen (STEVEOR/RIFIN) kinetics, which was originally proposed as a component of Chapter 3 (Research Paper 1). This analysis was initially conducted on selected samples from individuals of all ages, samples taken three months before detected infection, and then followed for nine months after confirmed and treated *P. falciparum* infection. These samples came from the PRISM longitudinal cohort from Uganda, outlined in Chapter 3 (17). Preliminary analysis figures can be found in the Appendices of this thesis and were not included in the main body as the results did not provide any insight into the kinetics of STEVEOR and RIFIN recombinant antigens. This was also observed for the reference recombinants, the established markers of seroincidence (preliminary analysis also found in the Appendices). This data highlights the complexity and the individual uniqueness of the human immune system as one of the main challenges in human immunology research. To mitigate this limitation, the kinetics analysis could be conducted on samples from a controlled human malaria infection (CHMI) study on malaria naïve individuals. However, conducting an CHMI study is very time-consuming and costly. Additionally, results on antibody reactivity to variant surface antigens from CHMI using malaria naïve individuals show contradicting results when compared to CHMI studies conducted on individuals from African malaria-endemic settings (18),(19). Hence, conclusions made may not accurately reflect the immune response dynamics in people living in malaria endemic settings.

8.4 References

1. Blythe JE, Xue YY, Kuss C, Bozdech Z, Holder AA, Marsh K, et al. Plasmodium falciparum STEVOR proteins are highly expressed in patient isolates and located in the surface membranes of infected red blood cells and the apical tips of merozoites. *Infect Immun*. 2008 Jul;76(7):3329–36.
2. Singh H, Madnani K, Lim YB, Cao J, Preiser PR, Lim CT. Expression dynamics and physiologically relevant functional study of STEVOR in asexual stages of Plasmodium falciparum infection. *Cell Microbiol* [Internet]. 2017 Jun 1 [cited 2023 Nov 8];19(6). Available from: <https://pubmed.ncbi.nlm.nih.gov/28030753/>
3. Smith JD. The role of PfEMP1 adhesion domain classification in Plasmodium falciparum pathogenesis research. *Mol Biochem Parasitol* [Internet]. 2014 [cited 2023 Nov 7];195(2):82. Available from: </pmc/articles/PMC4159067/>
4. Urban BC, Ferguson DJP, Pain A, Willcox N, Plebanski M, Austyn JM, et al. Plasmodium falciparum-infected erythrocytes modulate the maturation of dendritic cells. *Nature* [Internet]. 1999 Jul 1 [cited 2024 Apr 26];400(6739):73–7. Available from: <https://pubmed.ncbi.nlm.nih.gov/10403251/>
5. Tan J, Pieper K, Piccoli L, Abdi A, Foglierini M, Geiger R, et al. A LAIR-1 insertion generates broadly reactive antibodies against malaria variant antigens. *Nature* [Internet]. 2015 Jan 7 [cited 2024 Nov 27];529(7584):105. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC4869849/>
6. Chen Y, Xu K, Piccoli L, Foglierini M, Tan J, Jin W, et al. Structural basis of malaria RIFIN binding by LILRB1-containing antibodies. *Nature* [Internet]. 2021 Apr 22 [cited 2024 Nov 27];592(7855):639. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8068667/>

7. Niang M, Bei AK, Madnani KG, Pelly S, Dankwa S, Kanjee U, et al. STEVOR Is a Plasmodium falciparum Erythrocyte Binding Protein that Mediates Merozoite Invasion and Rosetting. [cited 2023 Jun 10]; Available from: <http://dx.doi.org/10.1016/j.chom.2014.06.004>
8. Sanyal S, Egeé S, Bouyer G, Perrot S, Safeukui I, Bischoff E, et al. Plasmodium falciparum STEVOR proteins impact erythrocyte mechanical properties. Blood [Internet]. 2012 Jan 1 [cited 2024 Apr 28];119(2):e1. Available from: </pmc/articles/PMC3257022/>
9. Hou N, Jiang N, Ma Y, Zou Y, Piao X, Liu S, et al. Low-Complexity Repetitive Epitopes of Plasmodium falciparum Are Decoys for Humoural Immune Responses. Front Immunol. 2020 Apr 15;11:511575.
10. Zhou AE, Berry AA, Bailey JA, Pike A, Dara A, Agrawal S, et al. Antibodies to Peptides in Semiconserved Domains of RIFINs and STEVORs Correlate with Malaria Exposure. mSphere. 2019 Mar 20;4(2).
11. Blythe JE, Niang M, Marsh K, Holder AA, Langhorne J, Preiser PR. Characterization of the repertoire diversity of the Plasmodium falciparum stevor multigene family in laboratory and field isolates. Malar J [Internet]. 2009 [cited 2023 Apr 11];8(1):140. Available from: </pmc/articles/PMC2706845/>
12. Bachmann A, Scholz JAM, Janßen M, Klinkert MQ, Tannich E, Bruchhaus I, et al. A comparative study of the localization and membrane topology of members of the RIFIN, STEVOR and PfMC-2TM protein families in Plasmodium falciparum-infected erythrocytes. Malar J. 2015 Jul 16;14(1).
13. Brady AM, Unger ER, Panicker G. Description of a novel multiplex avidity assay for evaluating HPV antibodies. J Immunol Methods [Internet]. 2017 Aug 1 [cited 2024

- Jun 3];447:31. Available from: [/pmc/articles/PMC6519448/](#)
14. McRobert L, Preiser P, Sharp S, Jarra W, Kaviratne M, Taylor MC, et al. Distinct trafficking and localization of STEVOR proteins in three stages of the *Plasmodium falciparum* life cycle. *Infect Immun* [Internet]. 2004 Nov 1 [cited 2020 Jun 29];72(11):6597–602. Available from: <https://iai.asm.org/content/72/11/6597>
 15. Niang M, Bei AK, Madnani KG, Pelly S, Dankwa S, Kanjee U, et al. The variant STEVOR protein of *Plasmodium falciparum* is a red cell binding protein important for merozoite invasion and rosetting. *Cell Host Microbe* [Internet]. 2014 Jul 7 [cited 2023 Aug 29];16(1):81. Available from: [/pmc/articles/PMC4382205/](#)
 16. Thomson-Luque R, Votborg-Novél L, Ndovie W, Andrade CM, Niangaly M, Attipa C, et al. *Plasmodium falciparum* transcription in different clinical presentations of malaria associates with circulation time of infected erythrocytes. *Nat Commun* [Internet]. 2021 Dec 1 [cited 2024 May 17];12(1). Available from: [/pmc/articles/PMC8324851/](#)
 17. Kanya MR, Arinaitwe E, Wanzira H, Katureebe A, Barusya C, Kigozi SP, et al. Malaria Transmission, Infection, and Disease at Three Sites with Varied Transmission Intensity in Uganda: Implications for Malaria Control. *Am J Trop Med Hyg*. 2015;92(5):903–12.
 18. Turner L, Wang CW, Lavstsen T, Mwakalinga SB, Sauerwein RW, Hermsen CC, et al. Antibodies against PfEMP1, RIFIN, MSP3 and GLURP Are Acquired during Controlled *Plasmodium falciparum* Malaria Infections in Naïve Volunteers. *PLoS One* [Internet]. 2011 [cited 2024 May 2];6(12):29025. Available from: [/pmc/articles/PMC3236238/](#)
 19. Antibody recognition of heterologous variant surface antigens after a single *Plasmodium falciparum* infection in previously naive adults - PubMed [Internet].

[cited 2024 May 2]. Available from: <https://pubmed.ncbi.nlm.nih.gov/17488905/>

Appendices

Dot Blot Analysis

Aim: To compare dual His-Myc tag and single GST tag recombinant proteins in terms of background reactivity to negative serum samples.

Results:

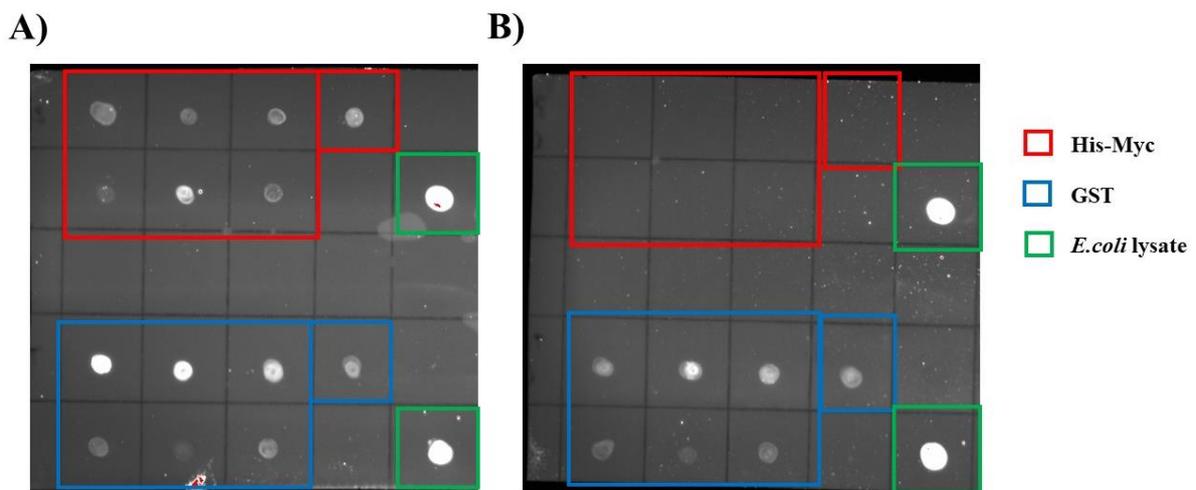


Figure 1: Dot blot of STEVOR and RIFIN recombinants blotted using A) PRISM1 control in 1/1000 dilution and B) PHE negative control in 1/1000 dilution. The secondary anti-human antibody was used in 1/10,000 dilution followed by TMB substrate for developing the images. The recombinants were dotted in the following order (STEVOR1_SC, STEVOR1_V2, STEVOR6_SC, STEVOR6_V2, STEVOR5_SC, RIFIN3_SC, RIFIN3_V2) for both His-Myc dual tag in red and GST single tag in blue. *E. coli* lysate in green is used as a positive control of the dot blot.

Discussion:

Recombinant proteins expressed with a single GST tag plasmid construct showed higher reactivity to negative serum samples compared to equivalent recombinants with dual His-Myc tags. Although this background reactivity is evident, it does not impact the seropositivity results presented in Research Paper 1 (Chapter 3) and Research Paper 3 (Chapter 7). The seropositivity

threshold per recombinant was determined by calculating the detected mean MFI plus three standard deviations from Luminex readings of negative samples, a method that accounts for the background reactivity associated with GST. **Kinetics Analysis**

Aim: To investigate STEVOR/RIFIN recombinant antigens kinetics using longitudinal serum samples from Uganda.

Results:

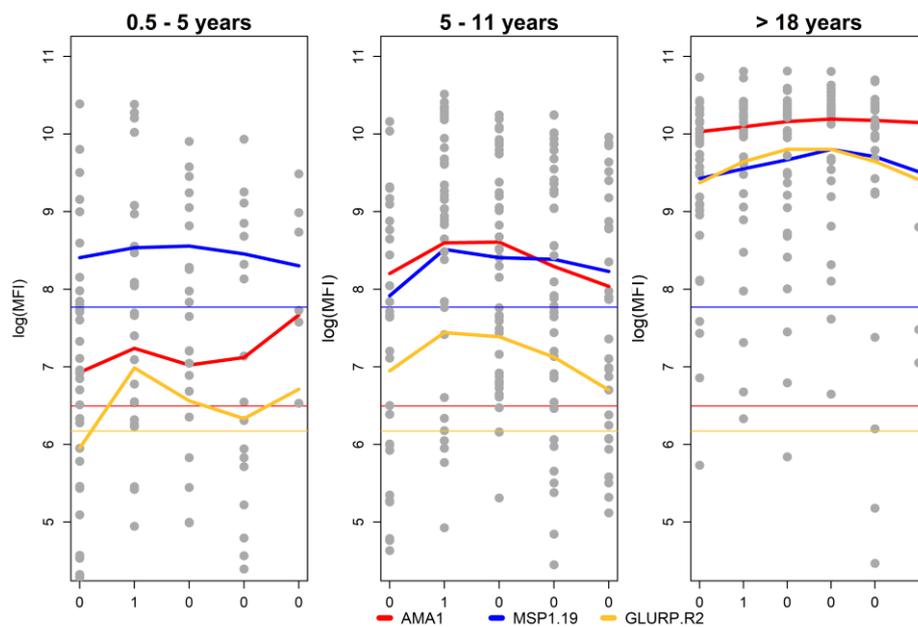


Figure 2: Kinetics analysis of long-term markers of *P. falciparum* seroincidence measured against individuals with no detected infection, followed by confirmed and treated infection, followed by three timepoints of sampling with no detected infection. Sampling was performed three months apart and samples were from the PRISM longitudinal cohort. Thick lines tracks the kinetics of antibodies overtime and thin lines indicate the seropositivity threshold of each target.

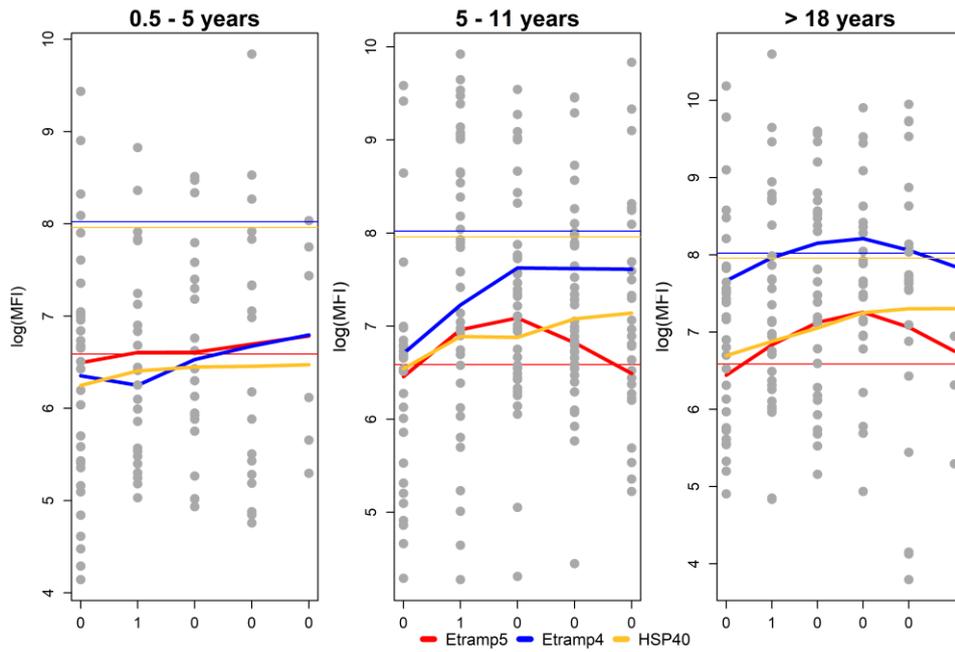


Figure 3: Kinetics analysis of short-term markers of *P. falciparum* seroincidence measured against individuals with no detected infection, followed by confirmed and treated infection, followed by three timepoints of sampling with no detected infection. Sampling was performed three months apart and samples were from the PRISM longitudinal cohort. Thick lines tracks the kinetics of antibodies overtime and thin lines indicate the seropositivity threshold of each target.

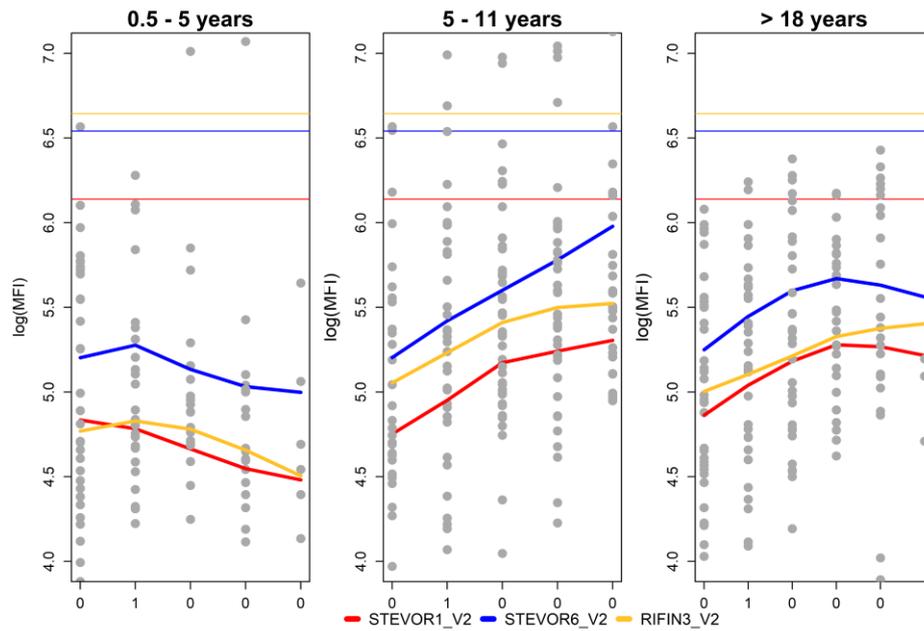


Figure 4: Kinetics analysis of hypervariable domain STEVOR recombinants measured against individuals with no detected infection, followed by confirmed and treated infection, followed by three timepoints of sampling with no detected infection. Sampling was performed three months apart and samples were from the PRISM longitudinal cohort. Thick lines tracks the kinetics of antibodies overtime and thin lines indicate the seropositivity threshold of each target.

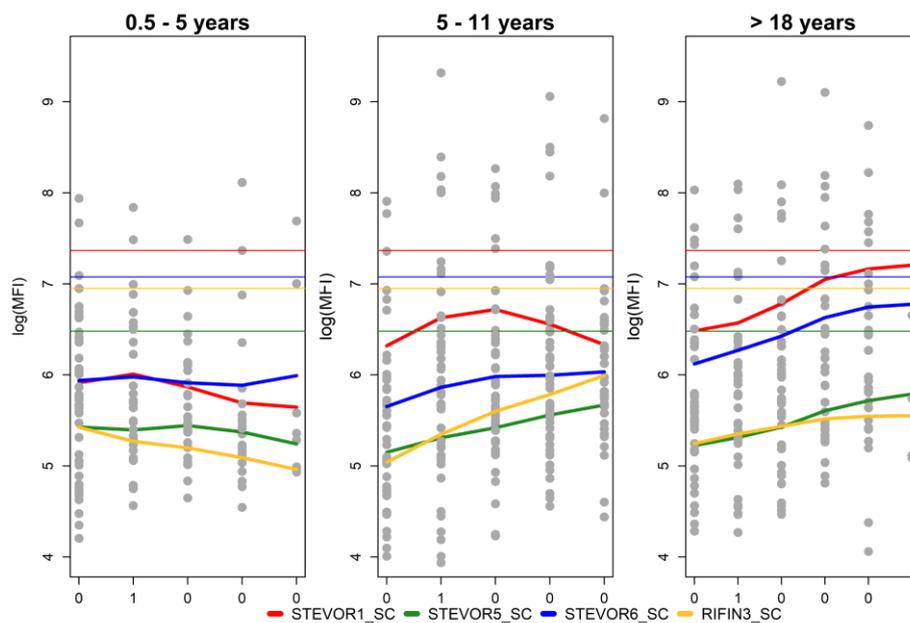


Figure 5: Kinetics analysis of semi-conserved domain STEVOR recombinants measured against individuals with no detected infection, followed by confirmed and treated infection, followed by three timepoints of sampling with no detected infection. Sampling was performed three months apart and samples were from the PRISM longitudinal cohort. Thick lines tracks the kinetics of antibodies overtime and thin lines indicate the seropositivity threshold of each target.

Discussion:

The samples used in this analysis differ from those in Research Paper 1. These samples are from individuals initially diagnosed as malaria-negative by LAMP and/or microscopy, then testing positive for malaria three months later, followed by three time points of three months where they tested negative. Importantly, the recruitment timing relative to intervention phases is unknown, so this sample group is distinct from that in Research Paper 1.

For each Luminex assay, seropositivity thresholds were determined by readouts from negative samples on the same assay plates, as Luminex fluorescence technology is highly sensitive. Readouts can vary for the same samples tested against the same antigens due to factors like antigen or reagent batches, room temperature, and operator differences. Therefore, the threshold plotted on the kinetics plots in Supplementary Figures 2 to 5 differs from those calculated in Research Paper 1 (Chapter 3) and Research Paper 3 (Chapter 7). Additionally, the STEVOR recombinants used in Research Paper 3 are different from those used in the kinetics analysis, making direct comparisons unfeasible.

This analysis aimed to investigate protein kinetics, regardless of negative sample readouts, to estimate antibody decay half-life following infection, without reinfection. However, the sample data were insufficient for this analysis. A key limitation is the unknown malaria status of participants before entering the longitudinal study, which impacts antibody dynamics against different antigens over time. Analyzing data at the individual level may yield a subset of individuals suitable for modeling antibody half-life decay over time.

***In-silico* model for protein sequence digitalization and variability visualization using Principal Component Analysis on R 4.3 computational language**

```
# Load necessary libraries
```

```
install.packages("seqinr")
```

```
library(seqinr)
```

```
# Read the CSV file containing sequence alignment data
```

```
dat3d7 <- read.csv("Additional_file_1_STEVOR_Variable_domain_allignmnet.csv", header = TRUE)
```

```
write.table(dat3d7, "STEVOR_V2_alligned_new.txt", sep = "\t", row.names = FALSE)
```

```
##### R scripts for calculations
```

```
# Reading the aligned sequence data
```

```
# The data should be formatted in tab-separated text with two columns:
```

```
# (name of sequence) \t (aligned sequence)
```

```
sites <- read.table(file = "STEVOR_V2_alligned_new.txt", header = TRUE, sep = "\t")
```

```
sites <- as.matrix(sites)
```

```
dim(sites)
```

```

# Finding the size of data

n_sample <- dim(sites)[1]

n_seq <- nchar(sites[2, 2])

# Translation of the sequence to boolean vectors

bool <- array(0, dim = c(n_sample, 21 * n_seq))

# Set column and row names for boolean matrix

colnames(bool) <- c(paste("a_", 1:n_seq, sep = ""), paste("r_", 1:n_seq, sep = ""),
  paste("n_", 1:n_seq, sep = ""), paste("d_", 1:n_seq, sep = ""),
  paste("c_", 1:n_seq, sep = ""), paste("q_", 1:n_seq, sep = ""),
  paste("e_", 1:n_seq, sep = ""), paste("g_", 1:n_seq, sep = ""),
  paste("i_", 1:n_seq, sep = ""), paste("l_", 1:n_seq, sep = ""),
  paste("k_", 1:n_seq, sep = ""), paste("m_", 1:n_seq, sep = ""),
  paste("f_", 1:n_seq, sep = ""), paste("p_", 1:n_seq, sep = ""),
  paste("s_", 1:n_seq, sep = ""), paste("t_", 1:n_seq, sep = ""),
  paste("y_", 1:n_seq, sep = ""), paste("v_", 1:n_seq, sep = ""),
  paste("h_", 1:n_seq, sep = ""), paste("w_", 1:n_seq, sep = ""),
  paste("-", 1:n_seq, sep = ""))

rownames(bool) <- sites[, 1]

```

```

# Fill the boolean matrix based on sequence characters

for (s in 1:n_sample) {

  se <- tolower(sites[s, 2])

  for (le in 1:n_seq) {

    base <- substr(se, le, le)

    bool[s, le + (match(base, c("a", "r", "n", "d", "c", "q", "e", "g", "i", "l", "k", "m", "f", "p", "s", "t", "y",
"v", "h", "w", "-")) - 1) * n_seq] <- 1

  }

}

apply(bool, 1, sum)

##### PCA

# Centering the data

center <- apply(bool, 2, mean)

diffs <- sweep(bool, 2, center)

diffs <- diffs / sqrt(2) # Compensating for doubled counts in Euclidean distance metrics

# Checking distribution of the distances

dist <- sqrt(apply(diffs^2, 1, sum))

```

```
qqnorm(dist)
```

```
hist(dist)
```

```
### PCA core
```

```
res_svd <- svd(diffs)
```

```
Left <- res_svd$u      # Left singular vector
```

```
Right <- res_svd$v     # Right singular vector
```

```
sqL <- diag(res_svd$d) # Diagonal matrix of the singular values
```

```
### Calculation of principal components
```

```
sPC_nuc <- Right %*% sqL / sqrt(n_sample)
```

```
sPC_sample <- Left %*% sqL / sqrt(n_seq)
```

```
rownames(sPC_nuc) <- colnames(bool)
```

```
rownames(sPC_sample) <- rownames(bool)
```

```
#### Output to text files
```

```
write.table(sPC_sample, file = "sPC_sample_STEVOR_V2_new.txt", sep = "\t")
```

```
write.table(sPC_nuc, file = "sPC_AAs_STEVOR_V2_new.txt", sep = "\t")
```

```
write.csv(sPC_sample, file = "sPC_sample_STEVOR_V2_new.csv", row.names = TRUE)
```

```

write.csv(sPC_nuc, file = "sPC_AAs_STEVOR_V2_new.csv", row.names = TRUE)

# Convert PCA results to a data frame for plotting

PCA <- as.data.frame(sPC_sample)

# Define colors for groups

colors <- c(rgb(10, 100, 255, maxColorValue = 255), rgb(140, 255, 100, maxColorValue = 255),
            rgb(255, 50, 10, maxColorValue = 255), rgb(100, 100, 100, maxColorValue = 255))

#### Output to PNG images

# Plotting the PCA results

png(width = 2100, height = 2300, pointsize = 60, file = "sPC_stevor_V2_new.png")

par(lwd = 4, mex = 0.6, mai = c(4, 4, 3, 0.2))

plot(PCA$V1, PCA$V2, col = colors, pch = 16, main = "STEVOR (V2)", xlab = "PC1", ylab = "PC2",
     axes = TRUE, cex = 1.2, cex.lab = 1.6, cex.main = 2)

legend("topright", legend = levels(grouped$colors), col = colors, pch = 16, bty = 'n', cex = 0.8)

dev.off()

```