Supplementary information

Systematic Serological Diversity of a Putative Malaria Vaccine Candidate and Broad Protection

Karamoko Niare, Timothy Chege, Kennedy Mwai, James Tuju, Lydia Nyamako, Micha Rosenkranz, Alfred Tiono, John Waitumbi, Bernhards Ogutu, Sodiomon Sirima, Gordon Awandare, Bourema Kouriba, Julian C. Rayner, Faith H.A. Osier.

Supplementary methods

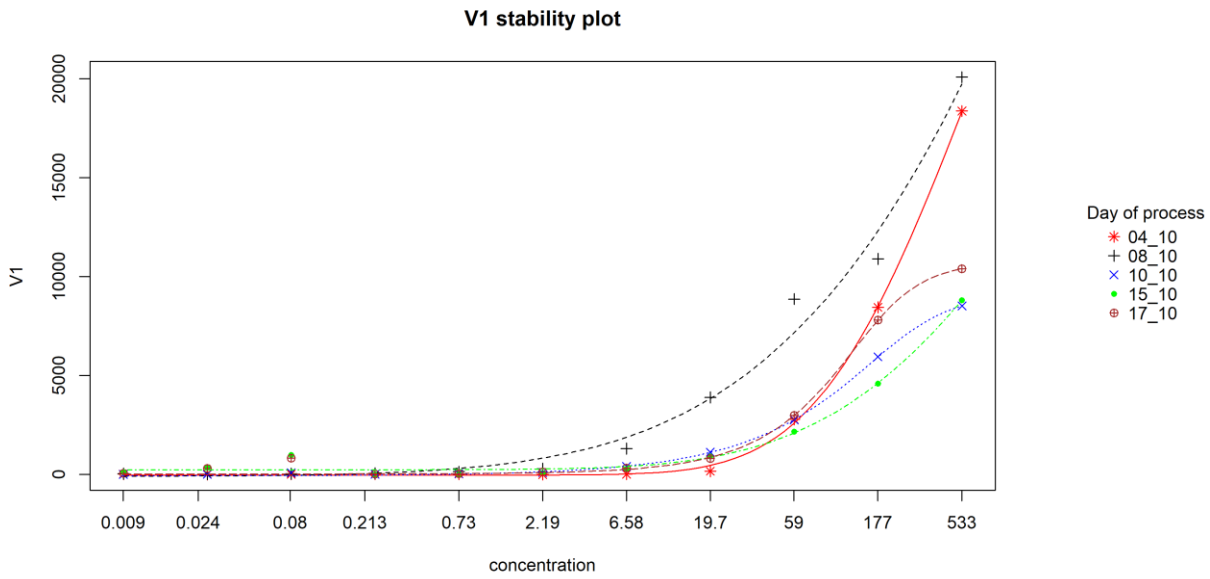
Quality control of protein microarray data

Although the normalization helps to correct for systematic variations in protein microarrays, producing quality raw data is key. To minimize technical and systematic errors, we took different quality control actions during the microarray design and processing including: i) the use of various types of control antibodies and sera, ii) the reduction of the number of sample batches, iii) random selection of samples from different sites to be used onto the same chips and, iv) limitation of the number of persons involved in sample processing and data extraction.

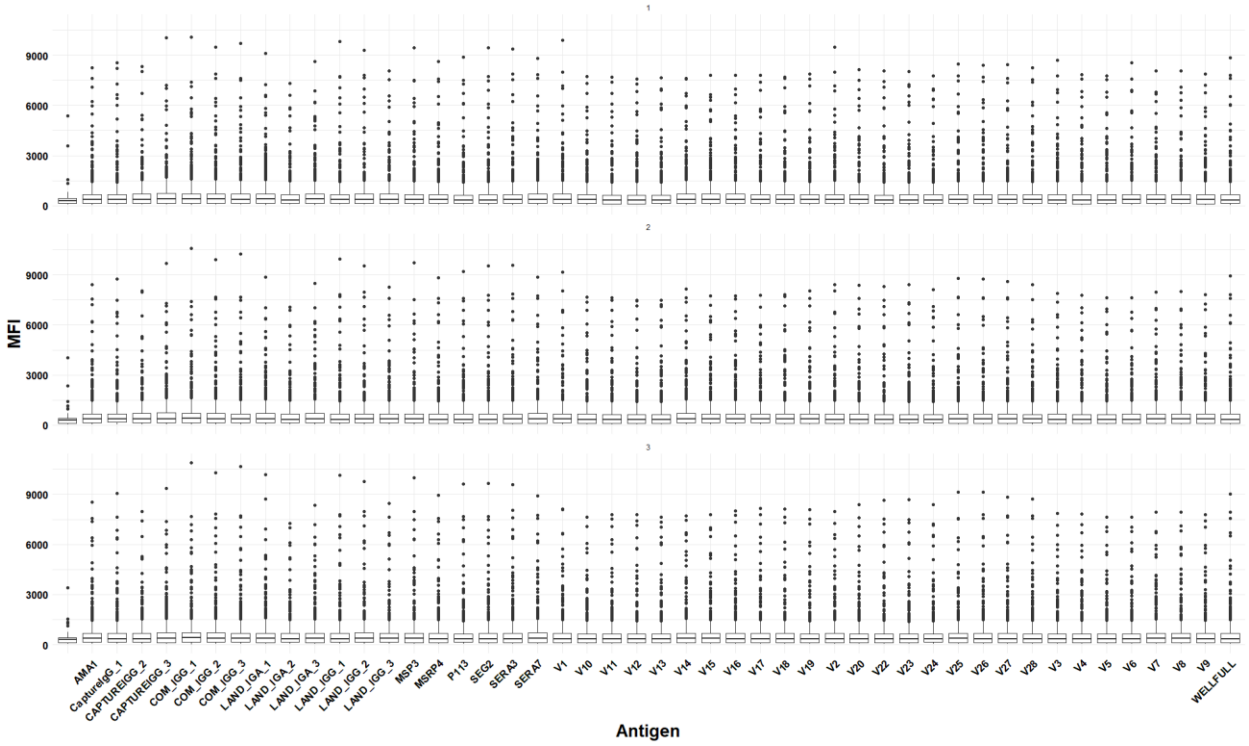
The median fluorescence intensity (MFI) values of the buffer spots were very low (median ≈ 0) with minimal chip-to-chip and day-to-day variations (Fig. S5A), suggesting low non-specific antibody binding and quality sample processing. The background signals were generally low despite some outliers (Fig. S5B) corresponding to 41 samples (5.7%) that looked degraded on the mini-array scans (Fig. S6). There was no significant difference in the background signal variations

between antigens (Fig. S5B). The buffer signals were low and consistent across days of sample processing (Fig. S5C). After removing these outliers, we remained with 675 samples for all the downstream analyses. We observed negligible intra-assay variations and MFI values from replicate spots were strongly correlated (P-value <0.001, R=0.9927, 95% CI [0.9926-0.9929] for IgG; and P-value < 0.001, R=0.9877, 95% CI [0.9875-0.9880] for IgM) (Fig. S5D) for both antigens and controls. A principal component analysis of the raw data of all the antigens showed clear separation of the naive controls and blanks from PHIS (Fig. S5E). As expected, the test samples that included seronegative subjects, low and high responders spread out into both PHIS and naive control clusters. Altogether, the data structure denoted minimal technical variations in general. Only variations in antibody reactivities among high responders were systematically higher compared to low responders (Fig. S7A). To deal with these common systematic variations in protein microarrays and batch effects while preserving biological differences between samples before statistical analyses, we normalized the data using combined ComBat and variance stabilizing normalization (vsn) methods (*1*–*3*). These normalization approaches could correct effectively for the systematic high variations in high responders (Fig. S7B). Overall, the quality of the IgG and IgM responses was highly comparable.

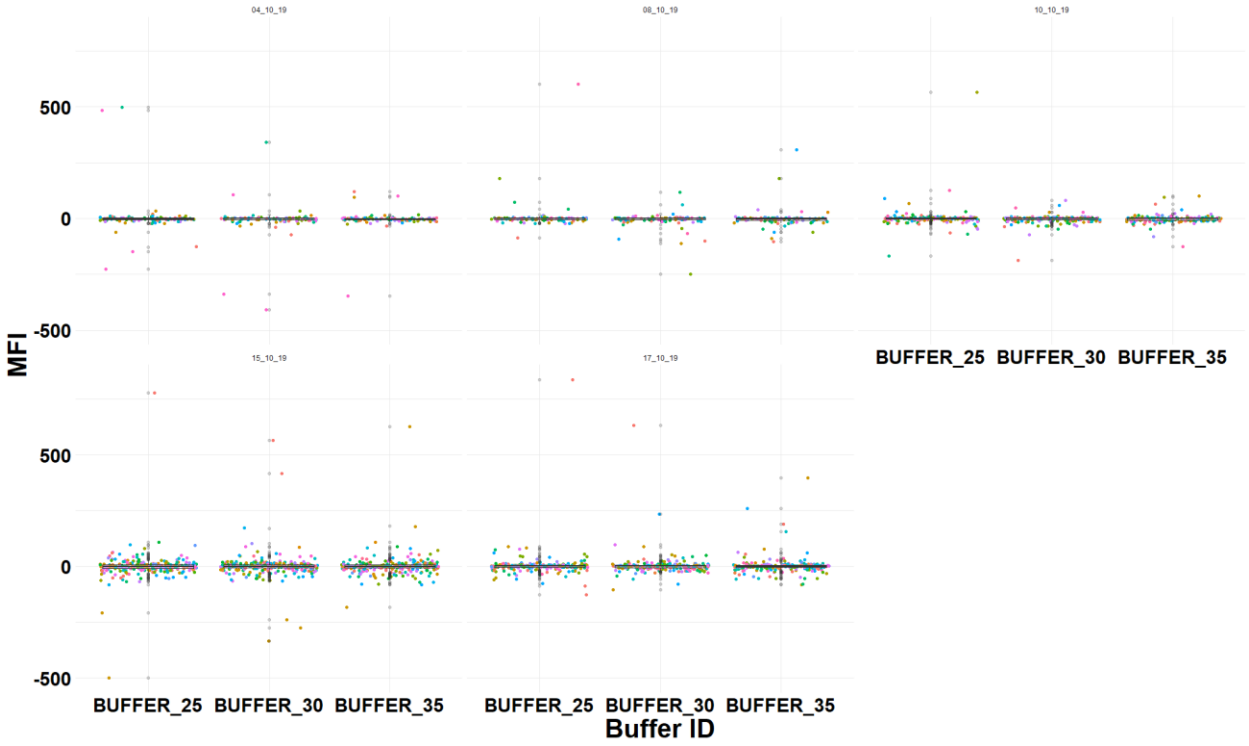
Supplementary figures



Supplementary Fig. 1 | Day-to-day variation using purified malaria immunoglobulin (MIG) dilutions against an illustrative antigen variant (V1). Concentrations ranged between 0.009 and 533 µg/mL (x-axis). Mean fluorescence intensity (MFI) of V1 represent the y-axis. The titration was performed during every run of microarray processing.



Supplementary Fig. 2| Distribution of microarray background signals. The median background MFI values were very low and consistent across antigens and control antibodies that were printed onto the chip. The outliers represent the degraded slide images which were removed before prior to data analysis

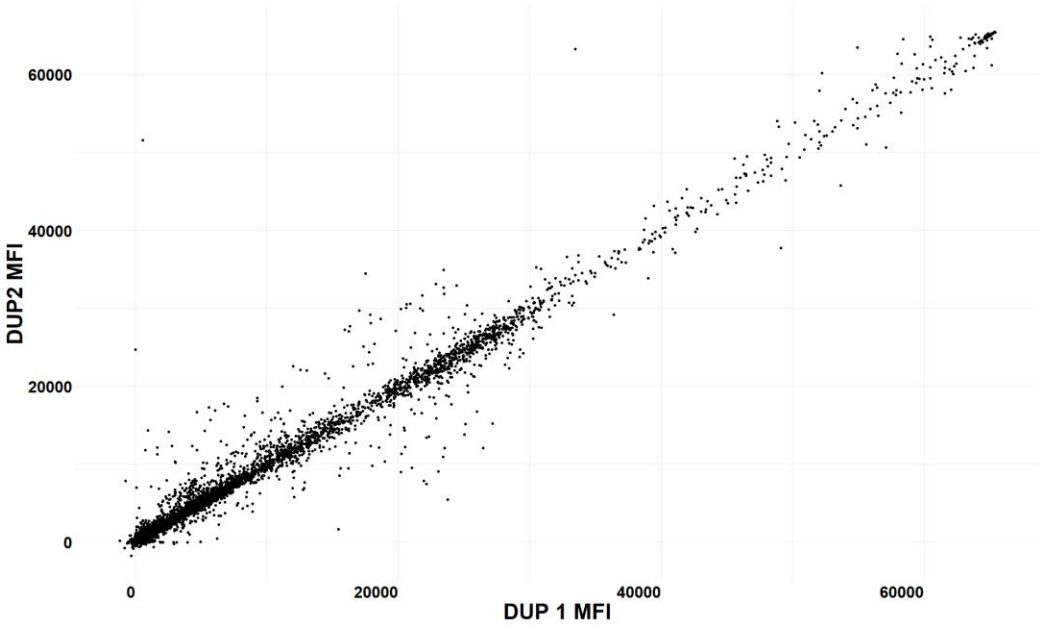


49

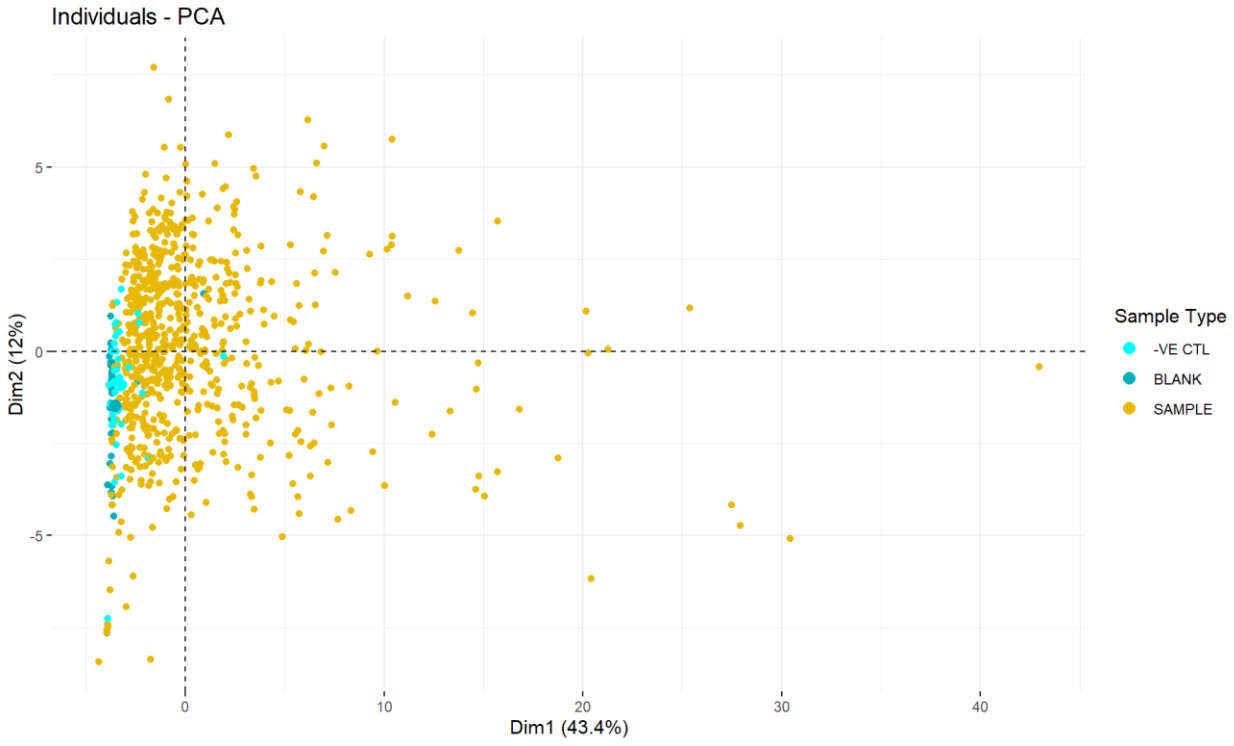
50 **Supplementary Fig. 3| Distribution of buffer spot mean fluorescence intensities by day of**
51 **sample processing**. The buffer signals were very low (close to zero) and consistent by day.
52 Three sets of buffer spots in triplicates (buffer spot set IDs: 25, 30 and 35) were used in each
53 run of sample processing.

54

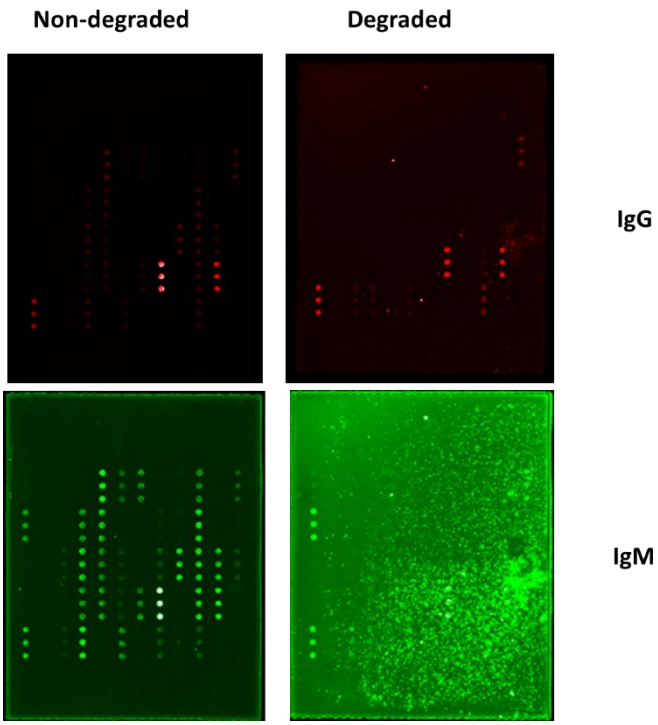
55



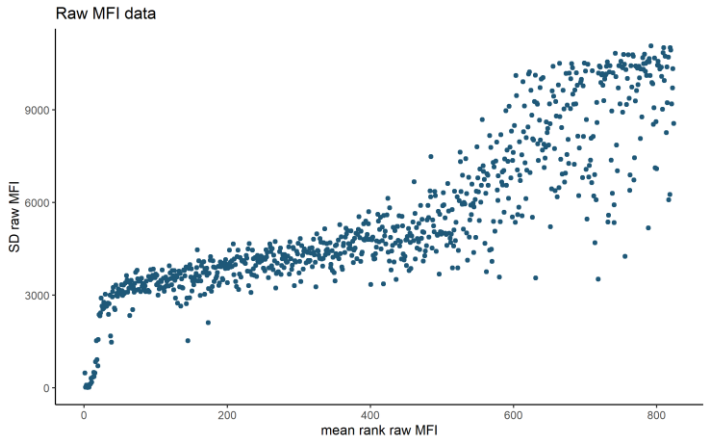
Supplementary Fig. 4 | Replicate analysis. MFI values of the first two spots (duplicate 1 [DUP1] and 2 [DUP2]) are compared. This plot represents IgG data which were highly comparable to those of IgM. The two spots were highly correlated (for IgG: R = 0.9927 [0.9926 – 0.9929], $P < 0.001$; for IgM: R = 0.9877 [0.9875 – 0.9880], $P < 0.001$).



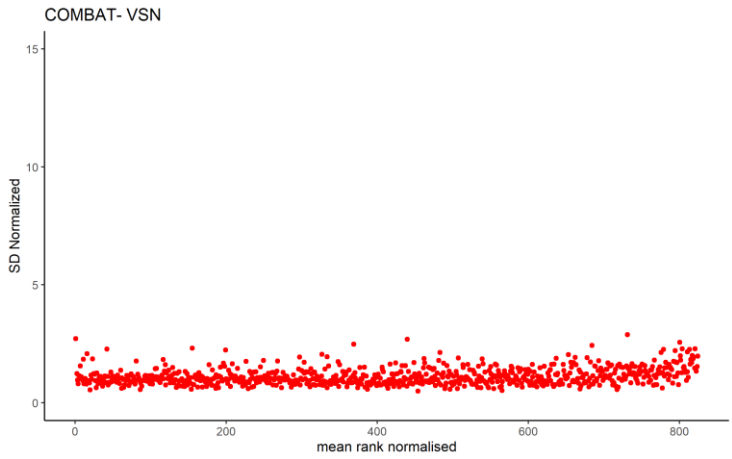
Supplementary Fig. 5| Principal component analysis of the raw antibody reactivity data showing the difference between samples and negative controls. Blank represents the spotted printing buffers. -VE CTL (negative controls) represents 22 malaria-naive sera from Europe.



Supplementary Fig. 6| Quality of illustrative miniarray images scanned after sample processing. Red and green images correspond to the same miniarrays scanned at 635 nm (IgG) and 532 nm (IgM), respectively.



Supplementary Fig. 7| Distribution of inter-individual MFI variations before data normalization. The standard deviations (SDs) were high in high responders.



Supplementary Fig. 8| Distribution of inter-individual MFI variations after data normalization. The high standard deviations (SDs) observed in high responders were drastically reduced by the double normalization approaches. COMBAT: a normalization approach that adjusts for batch effects using empirical Bayes framework. VSN: Variance stabilizing normalization.

References

1. W. Huber, A. Von Heydebreck, H. Sültmann, A. Poustka, M. Vingron, Variance

stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*. **18** (2002), doi:10.1093/bioinformatics/18.suppl_1.S96.

2. W. E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. **8**, 118–127 (2007).

3. G. Kamuyu, J. Tuju, R. Kimathi, K. Mwai, J. Mburu, N. Kibinge, M. Chong Kwan, S. Hawkings, R. Yaa, E. Chepsat, J. M. Njunge, T. Chege, F. Guleid, M. Rosenkranz, C. K. Kariuki, R. Frank, S. M. Kinyanjui, L. M. Murungi, P. Bejon, A. Färnert, K. K. A. Tetteh, J. G. Beeson, D. J. Conway, K. Marsh, J. C. Rayner, F. H. A. Osier, KILchip v1.0: A Novel Plasmodium falciparum Merozoite Protein Microarray to Facilitate Malaria Vaccine Candidate Prioritization. *Front. Immunol.* **9**, 2866 (2018).