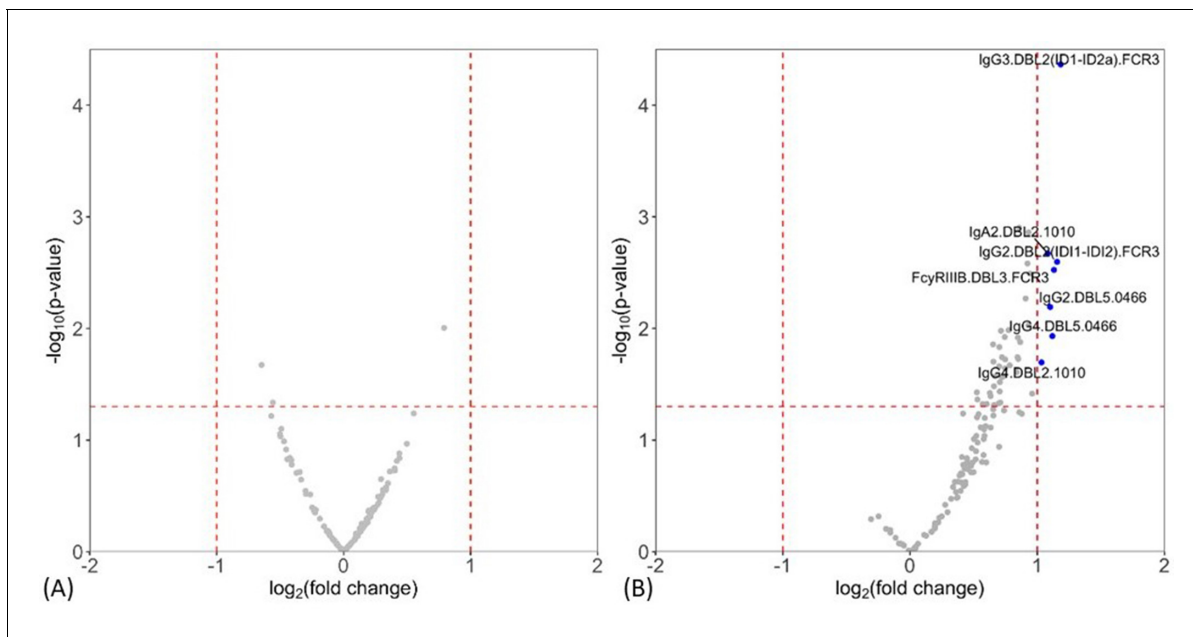


---

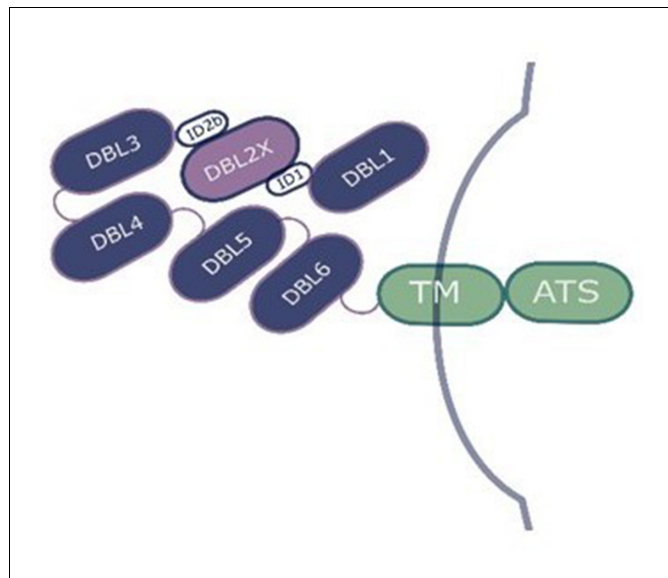
## Figures and figure supplements

Developing a multivariate prediction model of antibody features associated with protection of malaria-infected pregnant women from placental malaria

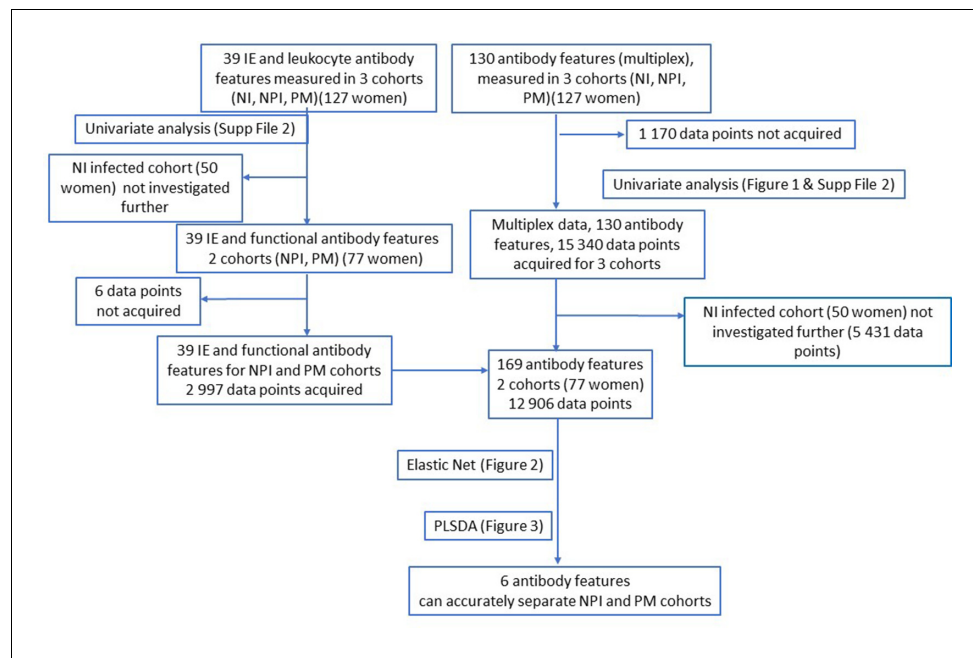
**Elizabeth H Aitken *et al***



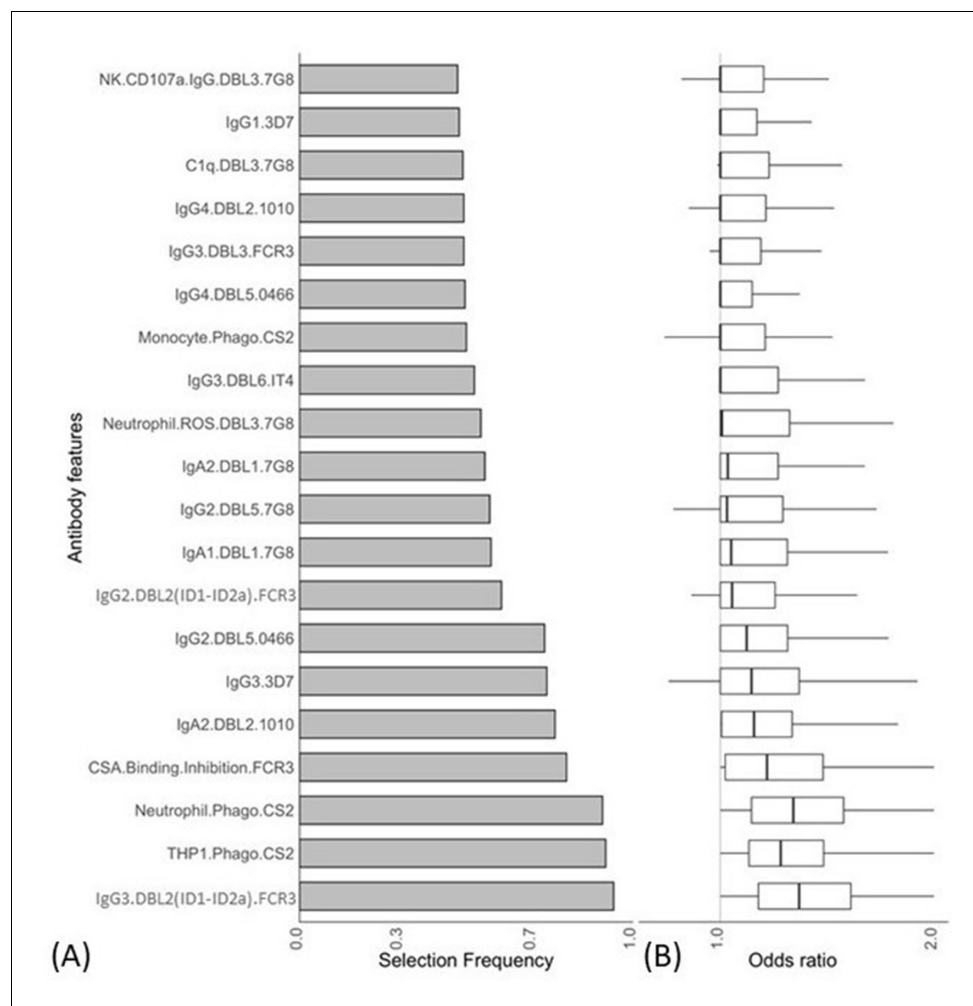
**Figure 1.** Individual antibody features to recombinant VAR2CSA Duffy binding-like (DBL) domain proteins measured by multiplex comparing women with placental malaria at delivery to (A) non-infected women and (B) women with non-placental infection. Fold-change ( $\log_2$  transformed), characterizing the magnitude of difference between the antibody levels of two groups (x-axis), is plotted against the  $-\log_{10}$  p-value, characterizing the statistical significance of the difference (y-axis). The vertical dotted lines ( $\log_2(2)$  and  $\log_2(0.5)$ ) mark a threshold for a twofold change, and the horizontal dotted lines ( $\log_{10}(0.05)$ ) mark the statistical significance threshold ( $p \leq 0.05$ , Welch's t-test). Antibody features beyond the thresholds are shown in blue and labeled.



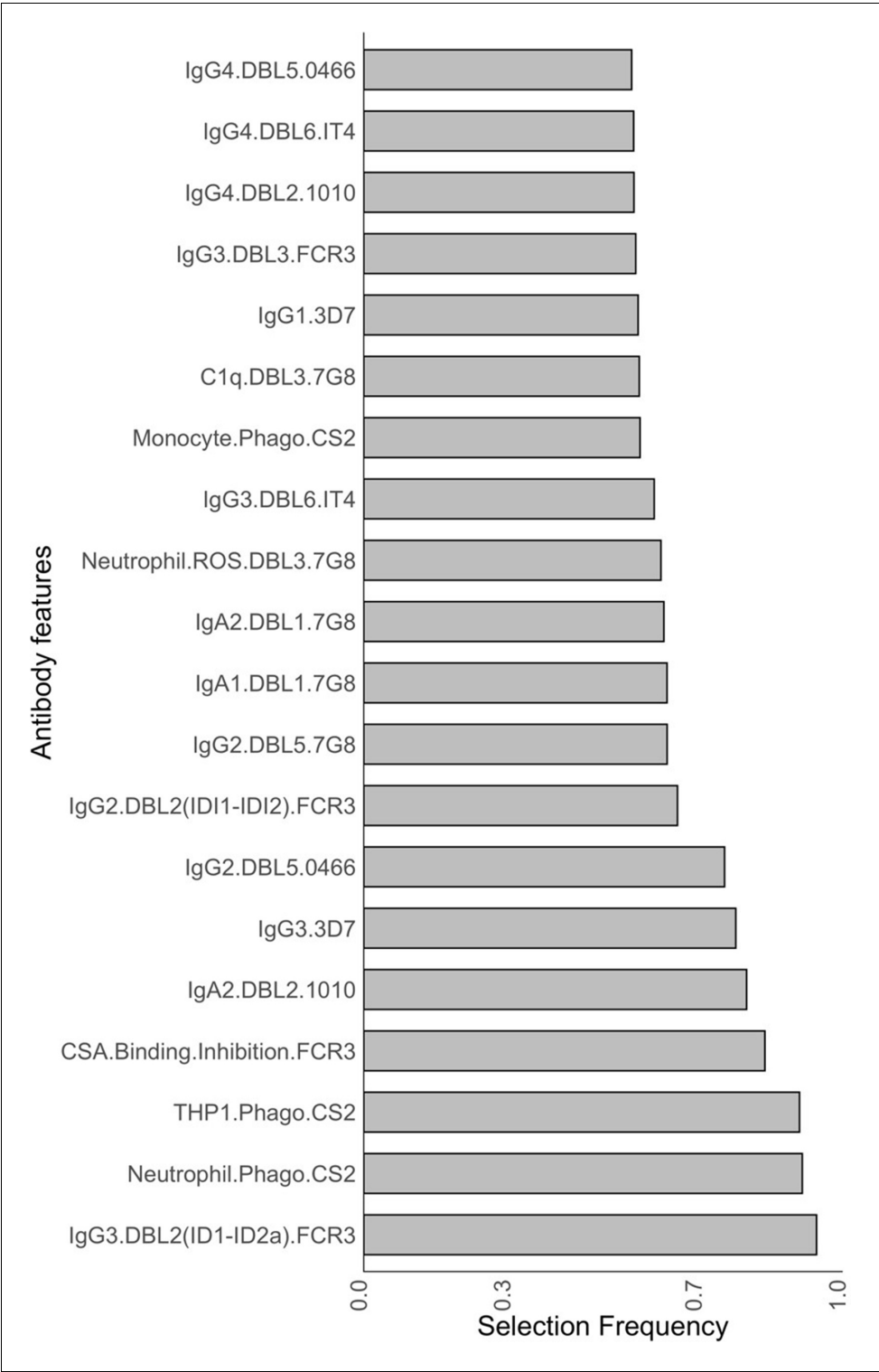
**Figure 1—figure supplement 1.** Diagram of the *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) VAR2CSA on the surface of the infected erythrocyte. DBL2, which is involved in binding to chondroitin sulfate A, is highlighted in light purple. DBL: Duffy binding-like domain; ID: interdomain region; TM: transmembrane segment; ATS: acid terminal segment.



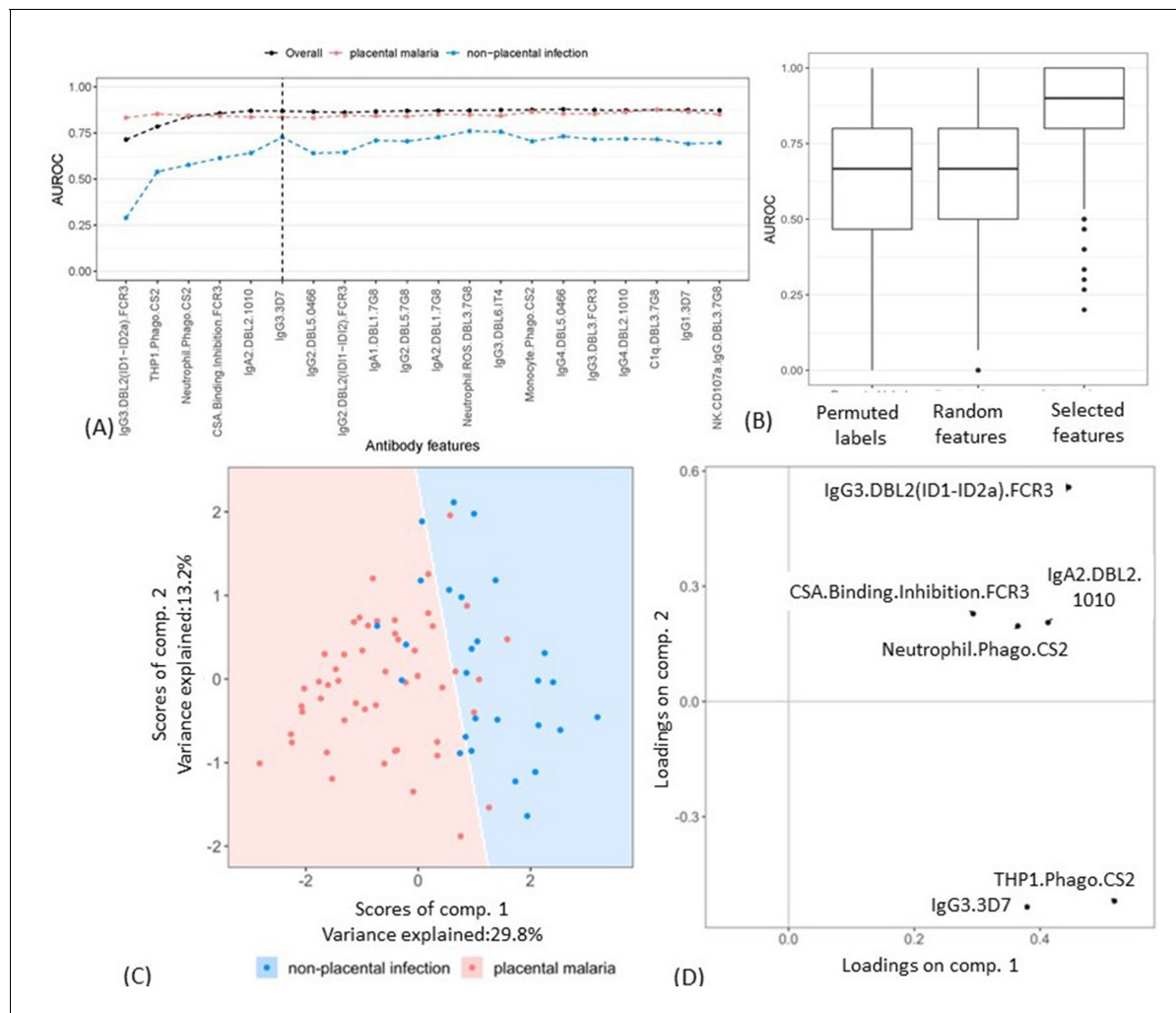
**Figure 1—figure supplement 2.** Flow chart of data acquisition and analysis. NPI: non-placental infection; PM: placental malaria; NI: non-infected; IE: infected erythrocyte; PLSDA: partial least squares discriminant analysis.



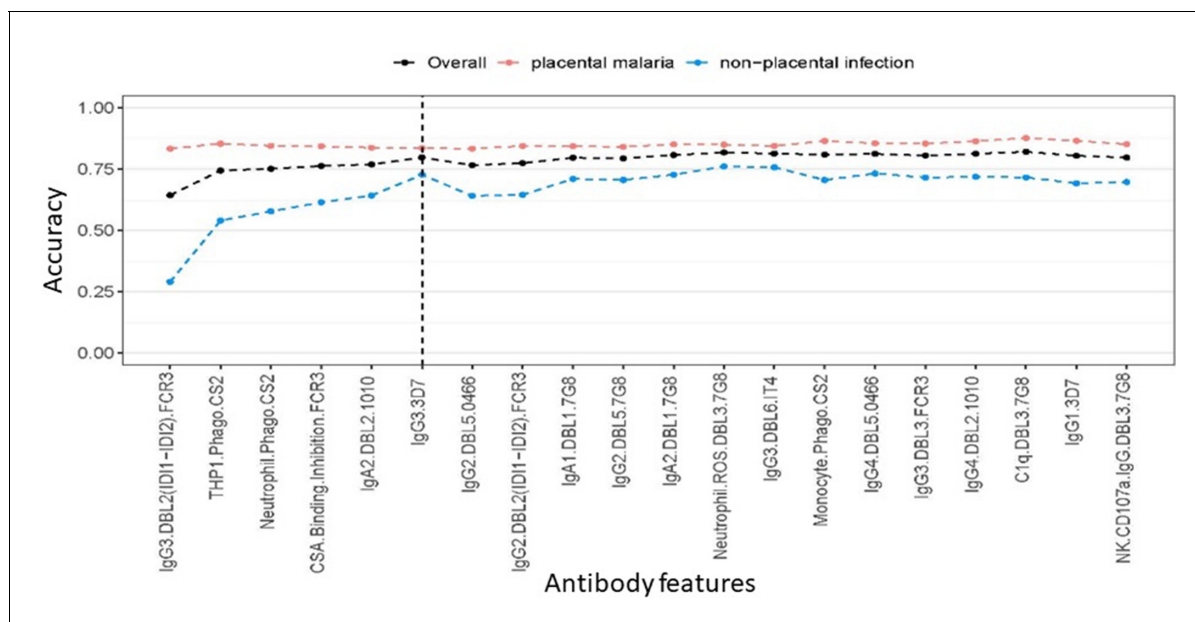
**Figure 2.** Antibody features that are influential in distinguishing between malaria-infected women with and without placental infection identified by the elastic net-regularized logistic regression model. Resampling (5000 repeats of 10-fold cross-validation) was used to obtain the selection frequencies and the odds ratios. **(A)** The top 20 antibody features are ranked in ascending order of selection frequency. **(B)** Boxplots of the estimated odds ratios, an odds ratio  $>1$  indicates the antibody feature is positively associated with non-placental infection at delivery. Boxplots are median, IQR, whiskers (the lowest data point that falls between Q1 and  $1.5 * Q1$  IQR, the highest data point that falls between Q3 and  $1.5 * Q3$  IQR). IQR: interquartile range.



**Figure 2—figure supplement 1.** Selection frequencies estimated across all  $\alpha$  values {0, 0.25, 0.5, 0.75 1} using the resampling of elastic net-regularized logistic regression. See section Identification of key antibody features.

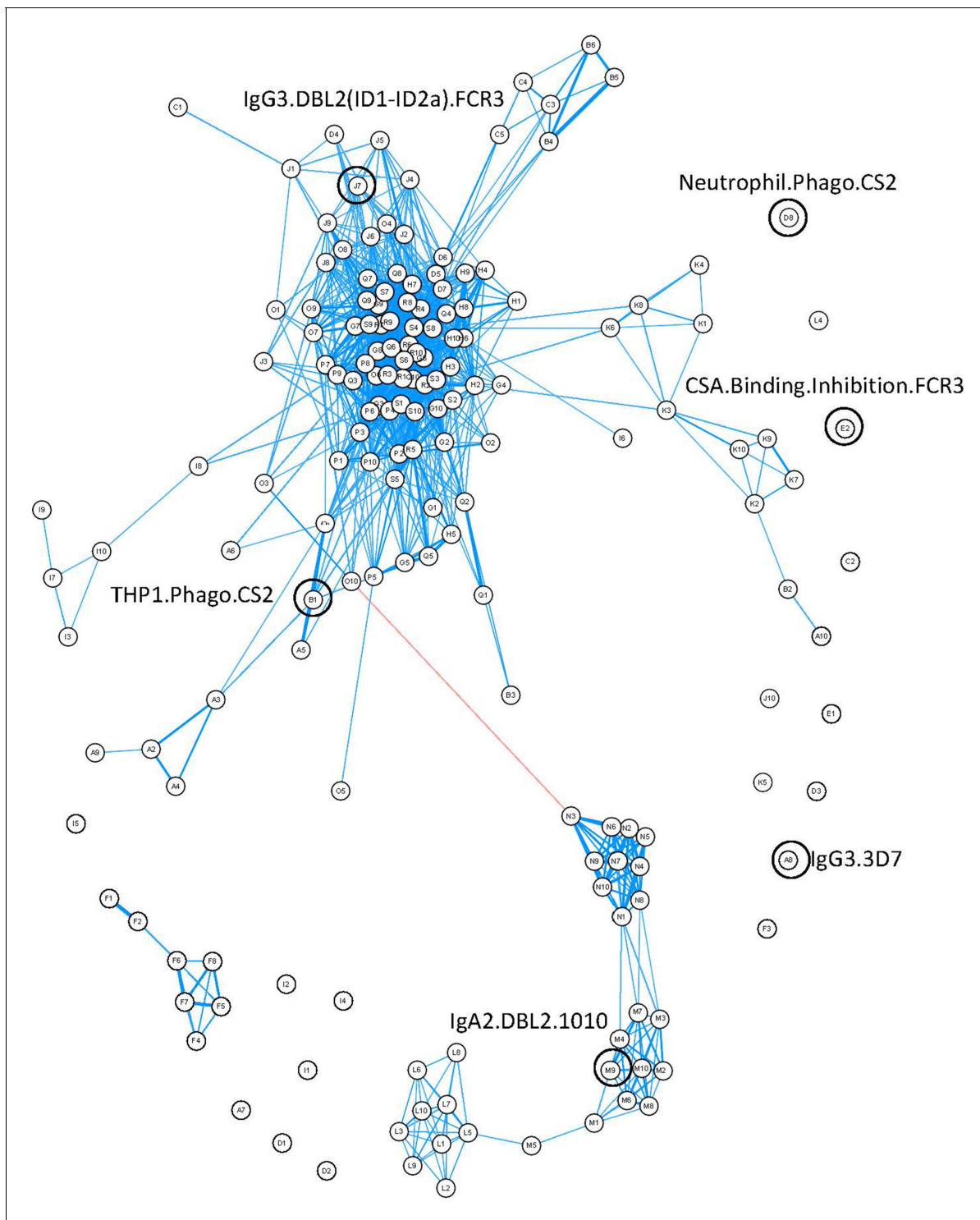


**Figure 3.** Selecting a minimal set of antibody features by partial least squares discriminant analysis (PLSDA). **(A)** Performance of PLSDA at classifying women as having placental malaria (PM) or non-placental infection (NPI) when the features (ranked by selection frequency using the elastic net; Figure 2) are added one by one to the model from highest to lowest rank (500 repeats of 10-fold cross-validation were performed to estimate accuracy for each model). The three lines represent the accuracy of classification of all women in the cohort (black), those with NPI (blue), and those with PM (red). The vertical dashed line denotes the cutoff beyond which the accuracy does not change significantly by adding more antibody features (used for selecting the minimal set of features). **(B)** Comparing the performance of the elastic net-regularized logistic regression (ENLR) + PLSDA applied on the original data of the top six variables (rightmost boxplot) with two permutation tests (null cases): (1) the PLSDA model was fitted to six randomly selected antibody features and the performance was computed for 500 repeats of 10-fold cross-validation resampling; (2) 100 datasets were generated by randomly permuting the group labels (PM and NPI) and the same analysis performed for the original dataset (i.e., building PLSDA models using the top six frequently selected antibody features found by resampling of elastic net) was repeated for each dataset. **(C)** Segregation of women with NPI (blue) and PM (red) using the scores of the two components of the PLSDA model with data from the selected six antibody features. The background colors show the predicted classification of the women for all the possible score values in the depicted range. **(D)** Feature loadings on the components of the PLSDA of the six selected antibody features (see Figure 3—source data 1 for more details about the factor loadings and group prediction using the PLSDA). Boxplots are median, IQR, whiskers (the lowest data point that falls between Q1 and 1.5 \* Q1 IQR, the highest data point that falls between Q3 and 1.5 \* Q3 IQR). IQR: interquartile range; AUROC: area under the receiver operating characteristic curve.

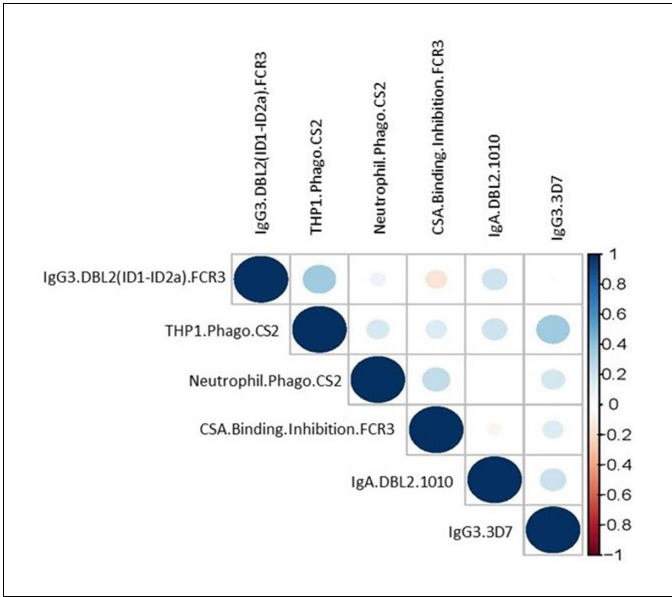


**Figure 3—figure supplement 1.** Accuracy of partial least squares discriminant analysis (PLSDA) at classifying women as having placental malaria or non-placental infection when the features (ranked by selection frequency using the elastic net; **Figure 2**) are added one by one to the model from highest to lowest rank (500 repeats of 10-fold cross-validation were performed to estimate accuracy for each model). The three lines represent the accuracy of classification of all women in the cohort (black), those with non-placental infection (blue), and those with placental malaria (red). The vertical dashed line denotes the cutoff beyond which the accuracy does not change significantly by adding more antibody features (used for selecting the minimal set of features).

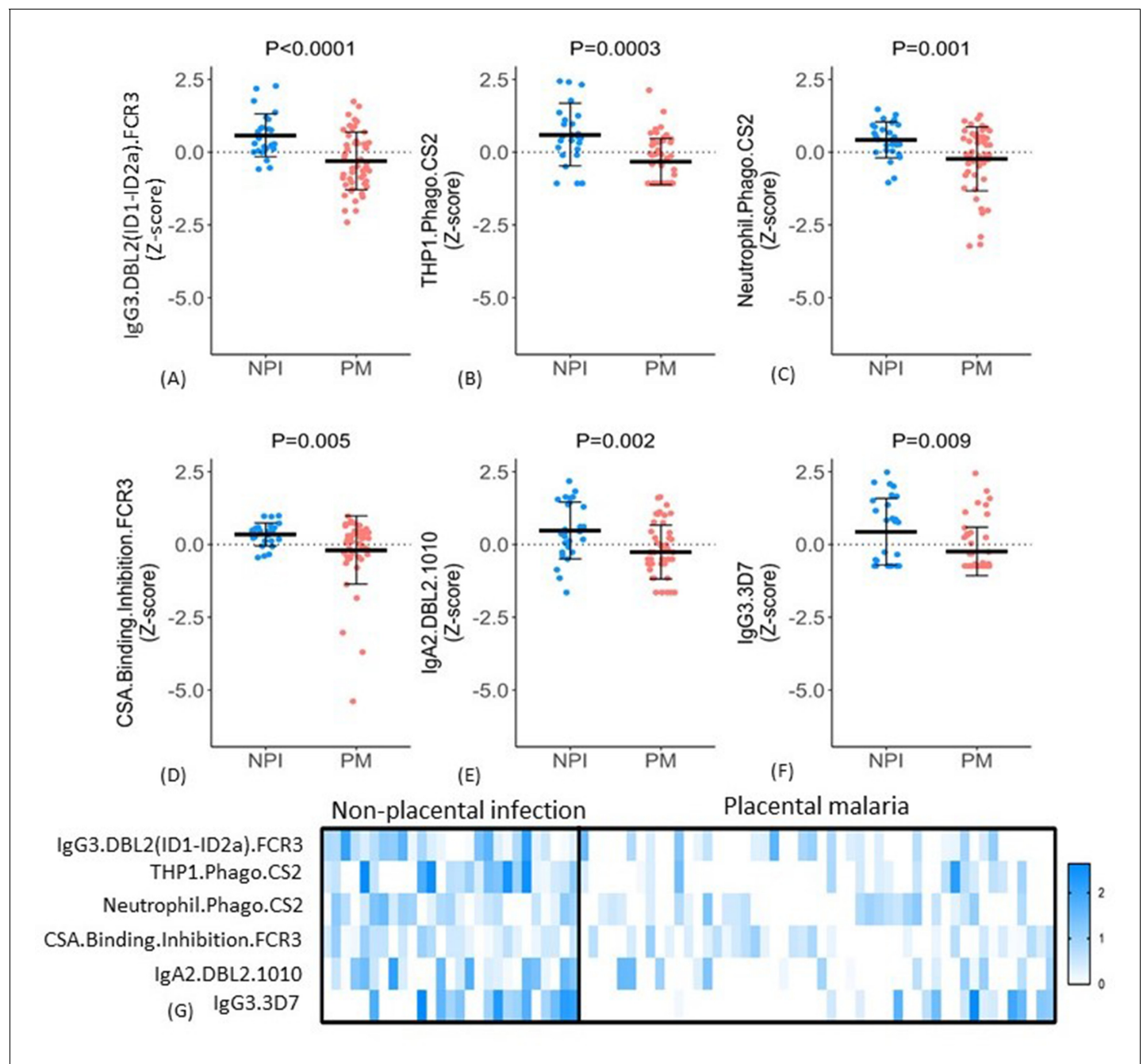




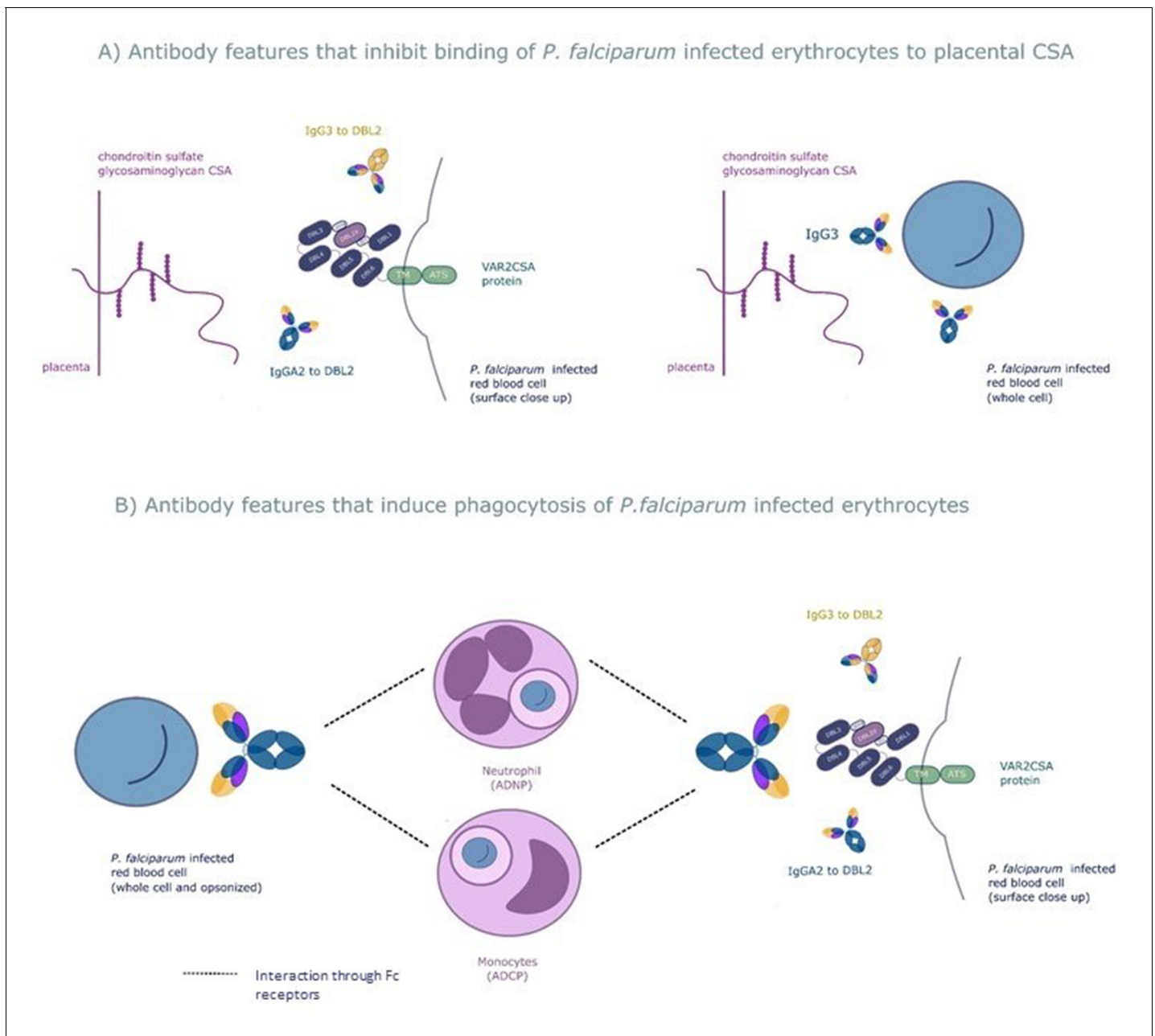
**Figure 4.** Correlation network of antibody features. Correlation network of all antibody features in both women with non-placental infection and those with placental malaria. The six selected antibody features do not cluster together. Antibody features with similar functions (denoted by the same letter) tend to correlate with each other. Blue: positive correlation; red: negative correlation; line width and closeness of variables increase with increasing correlation coefficients; only significant correlations (after Bonferroni correction for multiple comparisons) are shown. Selected antibody features identified by elastic net are highlighted and labeled. See **supplementary file 3** for a full list of feature names.



**Figure 4—figure supplement 1.** Correlation matrix of the six selected antibody features from women with non-placental infection at delivery and women with placental malaria. Color and size of dots represent estimates of the correlation coefficient ( $r$ ) (the further the correlation coefficient from 0 the larger the dot).



**Figure 5.** Distribution of antibody features in women with non-placental infection (NPI) and placental malaria (PM). (A–F) Levels of each of the selected antibody features in individual women in the two groups (G) No single antibody feature was present in all individuals with NPI (or was absent in all those with PM). Errors bars are mean (SD), p-values derived from Welch's t-test. IE: infected erythrocyte; Z-score: distribution of the features was centered and scaled to have zero mean and a standard deviation of 1.



**Figure 6.** The six selected antibody features may protect women from placental malaria by (A) inhibiting infected red blood cells from binding chondroitin sulfate A (CSA) and sequestering in the placenta and/or (B) promoting phagocytosis of infected red blood cells by monocytes and/or neutrophils. Selected features that may inhibit placental sequestration include IgG3 to the whole infected red blood cell and IgG3 and IgA2 to VAR2CSA's CSA binding domain DBL2. Selected features that may promote parasite clearance by antibody dependent phagocytosis include IgG3 to the whole infected red blood cell and to DBL2, IgA2 to DBL2, and neutrophil and monocyte phagocytosis of whole infected red blood cells. ADNP: antibody-dependent neutrophil phagocytosis; ADCP: antibody-dependent cellular phagocytosis; VAR2CSA: a parasite protein expressed on the surface of the infected red blood cell, made up of Duffy binding-like domains (DBL).